Pacific Journal of Mathematics

Volume 193 No. 2 April 2000

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. 193, No. 2, 2000

ON THE DIOPHANTINE EQUATION $\frac{x^n-1}{x-1}=y^q$

Yann Bugeaud, Maurice Mignotte, and Yves Roy

We prove that if $(x,y,n,q)\neq(18,7,3,3)$ is a solution of the Diophantine equation $(x^n-1)/(x-1)=y^q$ with q prime, then there exists a prime number p such that p divides x and q divides $p-1$. This allows us to solve completely this Diophantine equation for infinitely many values of x. The proofs require several different methods in diophantine approximation together with some heavy computer calculations.

1. Introduction.

A longstanding conjecture claims that the Diophantine equation

$$\frac{x^n-1}{x-1}=y^q \quad \text{in integers } x>1, y>1, n>2, q\geq 2 \tag{1}$$

has finitely many solutions, and, maybe, only those given by

$$\frac{3^5-1}{3-1}=11^2,\quad \frac{7^4-1}{7-1}=20^2 \text{ and } \frac{18^3-1}{18-1}=7^3.$$

Among the known results, let us mention that Ljunggren [**11**] solved (1) completely when $q=2$ and Ljunggren [**11**] and Nagell [**12**] when $3|n$ and $4|n$: they proved that in these cases there is no solution, except the previous ones. For more information and in particular for finiteness type results under some extra hypotheses, we refer the reader to Shorey & Tijdeman [**17**], [**18**] and to the survey of Shorey [**16**].

Very recently, some new results have been obtained by Bennett, Bugeaud, Mignotte, Roy, Saradha and Shorey, and, now, Equation (1) is completely solved when x is a square (there is no solution in this case [**15**], [**6**], [**1**]), when x is a power of any integer in the interval $\{2,\dots,10\}$ (the only two solutions are listed above [**5**]) or, under hypothesis (H) below, when x is a power of a prime number [**5**]. In [**5**] and [**6**], the proofs require several different methods in diophantine approximation together with some heavy computer calculations, one of the main tools being a new lower bound for linear forms in two p-adic logarithms (see [**3**]), which applies very well to (1) and allows us to considerably reduce the time of computation.

In the present work, using the same methods, we show that (1) has no solution (x,y,n,q) with $y\equiv 1 \pmod{x}$. As a corollary, we answer a question left open by Edgar [**7**]. Further, we give an irrationnality statement for Mahler's numbers, which completes results of [**15**] and [**3**].

2. Statement of the results.

First, we warn the reader that the proofs of the results obtained in Le [**10**] and in Yu & Le [**19**] are incorrect. Indeed, they all depend on Lemma 3 of [**10**], which is false, see the comment of Yuan Ping-Zhi [**20**]. The purpose of the present article is to supply a correct proof of all their claims. However, we notice that our method is essentially different from theirs.

Theorem 1. *Equation* (1) *has no solution* (x, y, n, q) *where* x *and* y *satisfy the following hypothesis*

(H) *Every prime divisor of* x *also divides* $y - 1$,

except $(18, 7, 3, 3)$. *Consequently, for all other solutions* (x, y, n, q) *of* (1) *with* q *prime, there exists a prime number* p *such that* p *divides* x *and* q *divides* $p - 1$.

The last assertion easily follows from the first one. Indeed, let (x, y, n, q) be a solution of (1), not satisfying (H). Let p be a prime factor of x such that p does not divide $y - 1$. Regarding (1) modulo p, we have $y^q \equiv 1 \pmod{p}$. However, $y \not\equiv 1 \pmod{p}$ and, since q is prime, q must divide $p - 1$. Consequently, Equation (1) with $q \geq 3$ prime implies that $x > 2q$.

Remark. Saradha and Shorey [**15**] showed that Equation (1) with (H) implies that $\max\{x, y, n, q\}$ is bounded by an effectively computable absolute constant. Further, Theorem 1 was proved for $x > 6 \times 10^{19}$ in [**3**] and for $x = z^t$ with $t \geq 1$ and $z = 6$, 10 or z prime in [**5**].

The main tool of the proof of Theorem 1 is a sharp lower bound for the p-adic distance between two rational numbers, obtained by Bugeaud [**3**]. Although this improvement of the estimates of Bugeaud & Laurent [**4**] seems to be very slight, it is of great interest for the problem investigated here. Indeed, it leads to considerably smaller numerical upper bounds, and allows us to do the numerical computation in a reasonable time. We should also mention that this tool can be used only when there is a prime factor of x dividing $y - 1$.

For fixed coprime rational positive integers a and b, we denote by $\operatorname{ord}_b a$ the least positive integer value of t for which $a^t \equiv 1 \pmod{b}$. It is known (cf. [**7**]) that every solution (x, y, n, q) of (1) with y an odd prime satisfies $n = \operatorname{ord}_y x$, and Edgar [**7**] asked whether we also have $q = \operatorname{ord}_x y$. Using Theorem 1, we are able to answer positively Edgar's question.

Theorem 2. *Every solution* (x, y, n, q) *of* (1) *satisfies* $q = \operatorname{ord}_x y$.

A problem arising in the theory of finite groups and strongly connected to Equation (1) is to find prime numbers P and Q and rational integers $n \geq 3$ and $a \geq 1$ such that $(Q^n - 1)/(Q - 1) = P^a$, see e.g., [**8**]. Our Theorem 1 allows us to prove that the latter equation with $a \geq 2$ is not solvable for $Q \in \{2, 3, 5, 7, 13, 17, 19, 37, 73, 97, \dots\}$. In order to give a precise statement, we need first to introduce some notations. Let S_1 be the set of all positive integers greater than 1 and composed only by 2 and by the primes of the form $2^a + 1$, for $a \geq 1$. We remark that S_1 is much bigger than the set also denoted by S_1 in [**15**]. Let p be a prime number of the form $2^a\,3^b + 1$, with integers $a \geq 0$, $b > 0$ and $p \not\equiv 55 \pmod{63}$. Thus $p \equiv 1, 4, 7 \pmod 9$, $p \not\equiv 1 \pmod 7$, and when $p \equiv 1 \pmod 9$, we have $p \not\equiv 6 \pmod 7$. Let f be any nonnegative integer satisfying $f \not\equiv 1, 4 \pmod 6$ if either $p \equiv 1 \pmod 9$, $p \equiv 3, 4 \pmod 7$ or $p \equiv 4 \pmod 9$, and $f \not\equiv 2, 5 \pmod 6$ if either $p \equiv 1 \pmod 9$, $p \equiv 2, 5 \pmod 7$ or $p \equiv 7 \pmod 9$. Let S_2 be the set of all numbers of the form $2^f\,p$. Put $S_3 = S_1 \cup S_2$ and notice that S_3 is an infinite set. The next statement directly follows from Theorem 1. It completes Corollary 2 of [**15**] and Corollary 1 of [**3**]. For its proof, the reader is directed to [**15**] and [**3**].

Corollary 1. *Equation* (1) *has no solution* (x, y, n, q), *where* $x = h^t$, *with* $h \in S_3$ *and* $t \geq 1$, *other than* $(h, t, y, n, q) = (3, 1, 11, 5, 2)$, $(7, 1, 20, 4, 2)$ *and* $(18, 1, 7, 3, 3)$.

As already mentioned in details in [**15**], Theorem 1 can be applied to obtain irrationality statements. Let $g \geq 2$ and $h \geq 2$ be integers. For any integer $m \geq 1$, we define $(m)_h = a_1 \cdots a_r$ to be the sequence of digits of m written in basis h, i.e., $m = a_1\,h^{r-1} + \cdots + a_r$, with $a_1 > 0$ and $0 \leq a_i < h$ for $1 \leq i \leq r$. For a sequence $(n_i)_{i\geq 1}$ of nonnegative integers, we put

$$a_h(g) = 0.(g^{n_1})_h\,(g^{n_2})_h \cdots$$

and we call Mahler's numbers the real numbers obtained in this way. It is known that $a_h(g)$ is irrational for any unbounded sequence $(n_i)_{i\geq 1}$; see the work of Sander [**14**] for an account of earlier results in this direction. Sander also considered the case when $(n_i)_{i\geq 1}$ is bounded with exactly two elements occurring infinitely many times, which are called limit points. As mentioned in [**15**], his paper contained an incorrect application of a result of Shorey & Tijdeman [**17**], hence his Theorem 3 remains unproved. Here, we extend Corollary 3 of [**15**] and Theorem 5 of [**3**] as follows.

Theorem 3. *Let* $(n_i)_{i\geq 1}$ *be a bounded sequence of nonnegative integers which is not ultimately periodic and has exactly two limit points* $N_1 < N_2$. *Let* $g \geq 2$ *and* $h \geq 2$ *be integers such that* $g \neq 1 + h + \ldots + h^{L-1}$ *for every integer* $L \geq 2$ *if* $(N_1, N_2) = (0, 1)$. *Assume also that* (N_1, N_2, g, h) *is not equal to* $(0, 2, 11, 3)$, $(0, 2, 20, 7)$, $(0, 3, 7, 18)$ *or to* $(1, 4, 7, 18)$ *and that* $g^{N_2 - N_1}$ *is* *not equal to* $1+h$ *whenever* $g^{N_1} < h$. *If* $h \in S_3$, *then* $a_h(g)$ *is irrational.*

Remark. We point out that all the assumptions imposed in Theorem 3 are necessary. Indeed, if (N_1, N_2, g, h) is equal to $(0, 2, 11, 3)$, $(0, 2, 20, 7)$, $(0, 3, 7, 18)$ or to $(1, 4, 7, 18)$, we see that all the digits of g^{N_1} and g^{N_2} with respect to base h are identically equal to 1 in the first three cases and to 7 in the last case. Further, if $g^{N_1} < h$ and $g^{N_2-N_1} = 1+h$, then we write $g^{N_2} = g^{N_1} + g^{N_1} h$ to see that all the digits of g^{N_1} and g^{N_2} are identically equal to g^{N_1} with respect to base h. Observe that there are instances, for example $(N_1, N_2, g, h) = (0, 2, 2, 3), (0, 2, 3, 8), (0, 3, 2, 7), (0, 4, 2, 15), (0, 2, 3, 8)$, when the relation $g^{N_2-N_1} = h+1$ with $g^{N_1} < h$ and $h \in S_3$ is satisfied. Finally, if $N_0 = 0$, $N_1 = 1$ and $g = 1 + h + \cdots + h^{L-1}$ for an integer $L \geq 2$, then the digits of g^{N_1} and g^{N_2} are identically equal to 1. Thus $a_h(g)$ is rational in each of these cases.

3. Auxiliary results.

Our main auxiliary result is a lower bound for the p-adic distance between two powers of algebraic numbers. Before stating it, we have to introduce some notation.

Let p be a prime number and denote by v_p the p-adic valuation normalized by $v_p(p) = 1$. Let x_1/y_1 and x_2/y_2 be two nonzero rational numbers and denote by g the smallest positive integer such that

$$v_p\big((x_1/y_1)^g - 1\big) > 0 \quad \text{and} \quad v_p\big((x_2/y_2)^g - 1\big) > 0.$$

Assume that there exists a real number E such that $v_p\big((x_1/y_1)^g - 1\big) \geq E > 1/(p-1)$. Theorem BU below provides explicit upper bounds for the p-adic valuation of

$$\Lambda = \left(\frac{x_1}{y_1}\right)^{b_1} - \left(\frac{x_2}{y_2}\right)^{b_2},$$

where b_1 and b_2 are positive integers. As in [**3**], we let $A_1 > 1, A_2 > 1$ be real numbers such that

$$\log A_i \geq \max\{\log|x_i|, \log|y_i|, E\,\log p\},\ (i = 1, 2)$$

and we put

$$b' = \frac{b_1}{\log A_2} + \frac{b_2}{\log A_1}.$$

Theorem BU. *With the above notation, let* x_1/y_1 *and* x_2/y_2 *be multiplicatively independent and assume that either* p *is odd or* $v_2(x_2/y_2 - 1) \geq 2$.

Then we have the upper estimates

$$v_p(\Lambda) \le \frac{36.1\, g}{E^3\,(\log p)^4}\big(\max\{\log b' + \log(E\,\log p) + 0.4, 6\,E\,\log p, 5\}\big)^2 \log A_1 \log A_2$$

and

$$v_p(\Lambda) \le \frac{53.8\, g}{E^3\,(\log p)^4}\big(\max\{\log b' + \log(E\,\log p) + 0.4, 4\,E\,\log p, 5\}\big)^2 \log A_1 \log A_2.$$

Proof. This is Theorem 2 of [**3**]. □

The following lemma is due to Saradha & Shorey [**15**] and originate in a work of Le [**9**]. Its proof uses Skolem's method.

Lemma 1. *Let* (x, y, n, q) *be a solution of Equation* (1) *satisfying the hypothesis* (H). *Then we have*

$$x^{n+1-2\beta} \le \left(\frac{n+3}{4}\right)^2 \left(2 + \frac{4}{x}\right)^{n-1} q^{\frac{\alpha\, q}{q-1}},$$

where $\alpha = n+1$ *if* q *does not divide* x, $\alpha = 2n$ *if* q *divides* x, *and* $\beta = \max\{1, n/q\}$.

Proof. This is Lemma 18 of [**15**]. □

Lemma 2. *Let* (x, y, n, q) *be a solution of Equation* (1) *satisfying the hypothesis* (H) *and such that* q *does not divide* x. *Then we have* $x \le 2000$ *if* $q = 3$ *and*

$$x \le \max\{961, 2.1382\, q\},$$

if $q \ge 5$.

Proof. The case $q = 3$ follows from Lemma 1. Further, we deduce from the hypothesis (H) that $y \equiv 1 \pmod{x}$. Arguing as in the proof of Lemma 19 of [**15**], we get $n \ge q+2$ and, for $q \ge 5$, we conclude exactly as in that lemma, and obtain the claimed upper bound. □

Lemma 3. *Equation* (1) *has no solution* (x, y, n, q) *with* $y \le 2n$.

Proof. Let (x, y, n, q) be a solution of (1). Recall that a primitive prime divisor of $x^n - 1$ is congruent to 1 modulo n and that there exists a primitive prime divisor for every odd n (see [**13**], page 20), which, consequently, is greater or equal to $2n+1$. If n is even, then $n = 2m$ with m odd, and we observe that $(x^m - 1)/(x-1)$ and $x^m + 1$ are relatively prime, each having a primitive prime factor (see [**13**], page 20). Hence, the lemma is proved. □

An important tool of our proof is a corollary to the following very deep result of Bennett [**1**], which completes an earlier work of Bennett & de Weger [**2**].

Theorem BE. *If a, b and q are integers with $b > a \geq 1$ and $q \geq 3$, then the equation*

$$|ax^q - by^q| = 1$$

has at most one solution in positive integers (x, y).

Proof. This is Theorem 1.1 of [**1**]. □

Corollary BE. *Equation* (1) *has no solution (x, y, n, q) with $n \equiv 1 \pmod{q}$ and $q \geq 3$.*

Proof. This is Corollary 1.2 of [**1**], and this follows easily from Theorem BE. Indeed, let (x, y, n, q) be a solution of (1), and assume that for a rational integer ℓ we have $n = q\,\ell + 1$. Then we get $x\,(x^l)^q - 1 = (x-1)\,y^q$. In view of Theorem BE, this is impossible, since $(1, 1)$ is a solution of the equation $x\,X^q - (x-1)\,Y^q = 1$. □

4. Proof of Theorem 1.

Let (x, y, n, q) be a solution of (1) satisfying the following assumption

(H) Every prime divisor of x also divides $y - 1$.

In view of the results of [**12**], [**11**] and [**5**] stated in the beginning of Section 1, we can suppose that $n \geq 5$ and $x \geq 11$. Moreover, we can restrict ourselves to the case when q is an odd prime number. Indeed, if ℓ is a prime divisor of q, we observe that $(x, y^{q/\ell}, n, \ell)$ is a solution of (1) such that every prime divisor of x also divides $y - 1$, hence divides $y^{q/\ell} - 1$.

- Sharp absolute upper bound for x and q under the assumption (H).

Our first goal is to obtain an absolute bound for q and for x, which improves Theorem 4 of [**3**]. More precisely, we distinguish the cases q divides x and q does not divide x, and we compute an upper estimate in each case. We proceed as follows: Firstly, we apply Theorem BU in order to bound q by a polynomial in $\log x$ and, secondly, we deduce from Lemmas 1 and 2 that x is smaller than a polynomial in q.

Application of Theorem BU.

We put

$$\Lambda = (1 - x) - \left(\frac{1}{y}\right)^q = -x^n\,y^{-q},$$

and we note that $1 - x$ and $1/y$ are multiplicatively independent (for a proof, see [**15**], below inequality (51)). Let p be a prime factor of x and let $\alpha \geq 1$ be such that p^α divides x but $p^{\alpha+1}$ does not. We assume that $p^\alpha \neq 2$, and notice that, if 4 divides x, then $y \equiv 1 \bmod 4$ and Theorem BU applies with the prime $p=2$.

Since p^α divides $(y-1)\cdot \frac{y^q-1}{y-1}$ and, in view of (H), p divides $y-1$, we get that p^α divides $y-1$ when $q \neq p$. If $q=p$, we infer from $\frac{y^p-1}{y-1} \equiv p \bmod p^2$ that $\max\{p, p^{\alpha-1}\}$ divides $y-1$. Thus, we deduce that, if $\alpha \geq 2$, we have $y \geq x^{(\alpha-1)/\alpha} \geq x^{1/2}$, whence, by $x^n > y^q$, we obtain the inequality

$$n \geq (q+1)/2, \tag{2}$$

which appears to be very useful. Applying Theorem BU and Lemma 3, we get $y \geq 11$ and

$$v_p(\Lambda) \leq \delta\, \frac{36.1}{\alpha^3\,(\log p)^4}\left(\max\left\{\log\left(\frac{q}{\log x}+0.42\right)+0.4,\; 6\,\alpha\,\log p, 5\right\}\right)^2 \log y\,\log(x-1)$$

and

$$v_p(\Lambda) \leq \delta\, \frac{53.8}{\alpha^3\,(\log p)^4}\left(\max\left\{\log\left(\frac{q}{\log x}+0.42\right)+0.4,\; 4\,\alpha\,\log p, 5\right\}\right)^2 \log y\,\log(x-1),$$

with $\delta = 1$ if $p \neq q$ or $\alpha = 1$ and $\delta = \alpha/(\alpha-1)$ if $p=q$ and $\alpha \geq 2$.

Further, $v_p(\Lambda) = n\alpha$ and $n\log x \geq q\,\log y$, thus we get

$$q \leq \delta\,\max\left\{\frac{36.1}{\alpha^2\,(\log p)^2}\left(\log\left(\frac{q}{\log x}+0.42\right)+0.4\right)^2,\; 36\cdot 36.1, \frac{25\cdot 36.1}{\alpha^2\,(\log p)^2}\right\}\frac{\log^2 x}{\alpha^2\,(\log p)^2} \tag{3}$$

and

$$q \leq \delta\,\max\left\{\frac{53.8}{\alpha^2\,(\log p)^2}\left(\log\left(\frac{q}{\log x}+0.42\right)+0.4\right)^2,\; 16\cdot 53.8, \frac{25\cdot 53.8}{\alpha^2\,(\log p)^2}\right\}\frac{\log^2 x}{\alpha^2\,(\log p)^2}. \tag{4}$$

Application of Lemma 1.

If q does not divide x, Lemma 2 provides the upper bound $x \leq 2000$ if $q=3$ and, else,

$$x \leq \max\{961, 2.1382q\}. \tag{5}$$

Assume now that q divides x. If $q = 3$, an easy calculation leads to $x \le 160000$. If $q \ge 5$ and $n \ge q$, it follows from Lemma 1 that

$$x \le \left(\frac{n+3}{4}\right)^{\frac{2q}{n(q-2)+q}} \left(2+\frac{4}{x}\right)^{\frac{q(n-1)}{n(q-2)+q}} q^{\frac{q^2}{q-1}\,\frac{2n}{n(q-2)+q}},$$

whence

$$x \le \left(\frac{q+3}{4}\right)^{\frac{2}{q-1}} \left(2+\frac{4}{x}\right)^{\frac{q}{q-2}} q^{\frac{2q^2}{(q-1)(q-2)}}, \tag{6}$$

and, assuming $x \ge 10^6$ and considering separately the cases $q \ge 800$ and $q < 800$, we get

$$x \le \max\bigl\{1.4 \times 10^6, 2.05\, q^{2q^2/(q-1)(q-2)}\bigr\}. \tag{7}$$

If $q \ge 5$ and $n < q$, it follows from Lemma 1 that

$$x \le \left(\frac{n+3}{4}\right)^{2/(n-1)} \left(2+\frac{4}{x}\right) q^{\frac{q}{q-1}\,\frac{2n}{n-1}}. \tag{8}$$

Since the map $q \mapsto q^{q/(q-1)}$ is increasing, we infer from (2) that

$$x \le \left(\frac{n+3}{4}\right)^{2/(n-1)} \left(2+\frac{4}{x}\right) (2n-1)^{\frac{n(2n-1)}{(n-1)^2}},$$

and, assuming that $n \le 400$, we get $x \le 1.4 \times 10^6$. For $n > 400$, we deduce from (8) under the assumption $x \ge 10^6$ the bound

$$x \le 2.05\, q^{\frac{2.005\, q}{(q-1)}}. \tag{9}$$

By (5), (7) and (9), we see that in all cases we have

$$x \le \max\bigl\{1.4 \times 10^6, 2.1382\, q, 2.05\, q^{2.005\, q/(q-1)}, 2.05\, q^{2q^2/(q-1)(q-2)}\bigr\},$$

which implies that $x \le 1.4 \times 10^6$ whenever $q \le 802$ and, since $2.005 \ge 2q/(q-2)$ as soon as $q \ge 802$, we always have

$$x \le \max\bigl\{1.4 \times 10^6, 2.05\, q^{2.005\, q/(q-1)}\bigr\}. \tag{10}$$

Absolute upper bound for q and for x.

Suppose first that q does not divide x and recall that we have assumed $x \ge 11$. If $x \ne 12$, then the primary part of x, i.e., the greatest prime power dividing x, is at least equal to 5. Using (3) with p^α replaced by 5 and noticing that $\delta = 1$, we get

$$q \le \max\left\{5.39 \left(\log\left(\frac{q}{\log x} + 0.42\right) + 0.4\right)^2, 501.8\right\} \log^2 x. \tag{11}$$

We make a direct computation in the case $x = 12$ and combine (11) with (5) to obtain that, for all x, we have

$$q \leq 72000$$

and, consequently,

$$x \leq 154000. \tag{12}$$

Suppose now that q divides x, whence q^2 divides x, as easily seen. We then use (4) with the prime number $p = q$ and with $\alpha \geq 2$. Since $\delta \leq 2$ and $4\,\alpha\,\log 3 \geq 5$, we get

$$q \leq 2\,\max\left\{\frac{53.8}{\alpha^2\,(\log q)^2}\left(\log\left(\frac{q}{\log x}+0.42\right)+0.4\right)^2, 860.8\right\}\frac{\log^2 x}{\alpha^2\,(\log q)^2}. \tag{13}$$

Combining (13) with (10) and replacing α by 2 yields

$$q \leq 1901$$

and

$$x \leq 7.76 \times 10^6. \tag{14}$$

- Strategy of the computational part of the proof.

In view of the above discussion, we are left to consider a finite number of pairs (x, q). It follows from (12) and (14) and from the results of [**12**], [**11**] and [**5**] that we have to prove that, for given integers $11 \leq x_0 \leq 7.76\times10^6$ and $q_0 \geq 3$, there is no solution (x_0, y, n, q_0) of (1) satisfying hypothesis (H), with $n \not\equiv 0 \pmod 3$ and $n \not\equiv 0 \pmod 4$. Furthermore, if such solution exists, known results on Catalan's equation imply that n must be odd. Indeed, if $n = 2m$, then $x_0^m - y_1^{q_0} = -1$ for some positive integer y_1 dividing y, and this is impossible since $x_0 < 10^{11}$, by a result of Hyyrö (see [**13**], pages 261 and 263).

We now describe our algorithm.

Firstly, using (3) or (4), we compute the bound on q obtained for x_0. If this bound is smaller than q_0 we have of course nothing to do (in this case, there is no solution), otherwise we have to work. And to work means:

- to consider the first prime numbers p with $p \equiv 1 \pmod{q_0}$ and to work modulo p,
- for each p, it leads to some conditions on the exponent n of Equation (1), more precisely it implies that n belongs to some set modulo $p - 1$,
- it appears that combining these conditions for several values of p and using that $n \not\equiv 0 \pmod 3$ and $n \not\equiv 0 \pmod 2$ yield $n \equiv 1 \pmod{q_0}$, which is excluded by Corollary BE, due to Bennett.

It remains to describe how we proceed to treat all the pairs (x, q).

We first prove that there is no solution satisfying (H), $q \leq 97$, $n \not\equiv 0 \pmod 3$ and $n \not\equiv 0 \pmod 2$. Indeed, we prove a little more, namely that there is no solution with $q \leq 97$ and x less than the bound given by (6).

Secondly, assuming that $q \geq 101$ and q divides x, we treat the pairs (x, q) with

$$101 \leq q \leq 1901, \quad x \leq 7.76 \times 10^6 \quad \text{and} \quad q^2 \text{ divides } x.$$

This can be done very quickly, since there are not many pairs to consider.

Lastly, we are left with pairs (x, q) such that q does not divide x. After some hundreds of hours of CPU with very fast computers, we could treat the full range $101 \leq q \leq 100000$ and x bounded by (5). It appeared that there is no solution. □

5. Proof of Theorems 2 and 3.

Proof of Theorem 2. Let (x, y, n, q) be a solution of (1) and set $k = \mathrm{ord}_x y$. We have $y^q \equiv 1 \pmod x$ and k divides q. Letting $q' = q/k$ and $y' = y^k$ and assuming that $q' > 1$, we see that (x, y', n, q') is also a solution of (1), which satisfies $y' \equiv 1 \pmod x$. In view of Theorem 1, this is impossible. Thus $q' = 1$ and $k = q$, as asserted. □

Proof of Theorem 3. Sander ([**14**], Theorem 2) proved that $a_h(g)$ is irrational if and only if $g^{N_2 - N_1} \neq (h^{tL} - 1)/(h^t - 1)$ for every integer $L \geq 1$, where t is given by the inequalities $h^{t-1} \leq g^{N_1} < h^t$. As noticed in [**15**], we have $(N_1, N_2) = (0, 1)$ or $N_2 - N_1 \geq 2$. To the first case corresponds the first condition in the statement of Theorem 3. Now we assume that $N_2 - N_1 \geq 2$ and $L = 2$, i.e., $g^{N_2 - N_1} = h^t + 1$. We have $t \geq 2$ by an assumption of Theorem 3 and we observe that $(g, h, t) = (3, 2, 3)$ is excluded by the definition of t. Let p be a prime divisor of $N_2 - N_1$. In view of the results on Catalan equation, $N_2 - N_1$ and t are odd and we have $p \geq 5$. We rewrite the equation as $h^t = G^p - 1$ with $G = g^{(N_2 - N_1)/p}$ to observe that h is divisible by a prime number of the form $1 + sp$. Thus $1 + sp = 2^a + 1$ or $1 + sp = 2^a 3^b + 1$ since $h \in S_3$. This is not possible for $p \geq 5$.

Consequently, we have $N_2 - N_1 \geq 2$ and $L \geq 3$, whence we deduce from Corollary 1 that $(h, t, g, L, N_2 - N_1)$ belongs to $\{(3, 1, 11, 5, 2), (7, 1, 20, 4, 2), (18, 1, 7, 3, 3)\}$, and we conclude by the definition of t. □

Remark. Shorey has pointed out to us that an assumption $g^{N_2 - N_1} \neq h + 1$ whenever $g^{N_1} < h$ should be added in Corollary 3 of [**15**]. Indeed, observe that $a_h(g)$ is rational if $g^{N_2 - N_1} = h + 1$ with $g^{N_1} < h$.

Acknowledgement. The authors are very grateful to Tarlok Shorey. His numerous remarks helped us to considerably improve the presentation of the paper.

References

[1] M. Bennett, Rational approximation to algebraic number of small height: The diophantine equation $|ax^n - by^n| = 1$, J. Reine Angew Math., to appear.

[2] M. Bennett and B.M.M. de Weger, *On the diophantine equation* $|ax^n - by^n| = 1$, Math. Comp., **67** (1998), 413-438.

[3] Y. Bugeaud, *Linear forms in p-adic logarithms and the diophantine equation* $(x^n - 1)/(x-1) = y^q$, Math. Proc. Cambridge Philos. Soc., **127** (1999), 373-381.

[4] Y. Bugeaud and M. Laurent, *Minoration effective de la distance p-adique entre puissances de nombres algébriques*, J. Number Th., **61** (1996), 311-342.

[5] Y. Bugeaud and M. Mignotte, *On integers with identical digits*, Mathematika, to appear.

[6] Y. Bugeaud, M. Mignotte, Y. Roy and T.N. Shorey, *On the diophantine equation* $(x^n - 1)/(x-1) = y^q$, Math. Proc. Cambridge Philos. Soc., **127** (1999), 353-372.

[7] H. Edgar, *Problems and some results concerning the diophantine equation* $1 + A + A^2 + \cdots + A^{x-1} = P^y$, Rocky Mountain J. Math., **15** (1985), 327-329.

[8] R. M. Guralnick, *Subgroups of prime power index in a simple group*, J. Algebra, **81** (1983), 304-311.

[9] M. Le, *A note on the diophantine equation* $(x^m - 1)/(x-1) = y^n$, Acta Arith., **64** (1993), 19-28.

[10] ———, *A note on perfect powers of the form* $x^{m-1} + \cdots + x + 1$, Acta Arith., **69** (1995), 91-98.

[11] W. Ljunggren, *Noen Setninger om ubestemte likninger av formen* $(x^n - 1)/(x-1) = y^q$, Norsk. Mat. Tidsskr., **25** (1943), 17-20.

[12] T. Nagell, *Note sur l'équation indéterminée* $(x^n - 1)/(x-1) = y^q$, Norsk. Mat. Tidsskr., **2** (1920), 75-78.

[13] P. Ribenboim, *Catalan's Conjecture*, Academic Press, Boston, 1994.

[14] J.W. Sander, *Irrationality criteria for Mahler's numbers*, J. Number Theory, **52** (1995), 145-156.

[15] N. Saradha and T.N. Shorey, *The equation* $(x^n - 1)/(x-1) = y^q$ *with x square*, Math. Proc. Cambridge Philos. Soc., **125** (1999), 1-19.

[16] T.N. Shorey, *Exponential diophantine equations involving product of consecutive integers and related equations*, to appear.

[17] T.N. Shorey and R. Tijdeman, *New applications of diophantine approximations to Diophantine equations*, Math. Scand., **39** (1976), 5-18.

[18] ———, *Exponential Diophantine Equations*, Cambridge Tracts in Mathematics, **87** (1986), Cambridge University Press, Cambridge.

[19] L. Yu and M. Le, *On the diophantine equation* $(x^m - 1)/(x-1) = y^n$, Acta Arith., **83** (1995), 363-366.

[20] P.-Z. Yuan, *Comment*: "A note on perfect powers of the form $x^{m-1} + \cdots + x + 1$", Acta Arith., **69** (1995), 91-98, by Maohua Le and "On the diophantine equation $(x^m - 1)/(x - 1) = y^n$", Acta Arith., **83** (1995), 363-366, by Li Yu & Maohua Le, Acta Arith., **83** (1998), 199.

Received September 25, 1998.

Université Louis Pasteur
67084 Strasbourg
France
E-mail address: bugeaud@math.u-strasbg.fr

Université Louis Pasteur
67084 Strasbourg
France
E-mail address: mignotte@math.u-strasbg.fr

Université Louis Pasteur
67084 Strasbourg
France
E-mail address: yr@dpt-info.u-strasbg.fr

PACIFIC JOURNAL OF MATHEMATICS
Vol. 193, No. 2, 2000

COMMUTATORS WITH POWER CENTRAL VALUES ON A LIE IDEAL

Luisa Carini and Vincenzo De Filippis

Let R be a prime ring of characteristic $\neq 2$ with a derivation $d \neq 0$, L a noncentral Lie ideal of R such that $[d(u), u]^n$ is central, for all $u \in L$. We prove that R must satisfy s_4 the standard identity in 4 variables. We also examine the case R is a 2-torsion free semiprime ring and $[d([x, y]), [x, y]]^n$ is central, for all $x, y \in R$.

Let R be a prime ring and d a nonzero derivation of R. A well known result of Posner [**14**] states that if the commutator $[d(x), x] \in Z(R)$, the center of R, for any $x \in R$, then R is commutative.

In [**11**] C. Lanski generalizes the result of Posner to a Lie ideal. To be more specific, the statement of Lanski's theorem is the following:

Theorem ([**11**, Theorem 2, page 282])**.** *Let R be a prime ring, L a noncommutative Lie ideal of R and $d \neq 0$ a derivation of R. If $[d(x), x] \in Z(R)$, for all $x \in L$, then either R is commutative, or* $\mathrm{char}(R) = 2$ *and R satisfies s_4, the standard identity in* 4 *variables.*

Here we will examine what happens in case $[d(x), x]^n \in Z(R)$, for any $x \in L$, a noncommutative Lie ideal of R and $n \geq 1$ a fixed integer.

One cannot expect the same conclusion of Lanski's theorem as the following example shows:

Example 1. Let $R = M_2(F)$, the 2×2 matrices over a field F, and take $L = R$ as a noncommutative Lie ideal of R. Since $[x, y]^2 \in Z(R)$, for all $x, y \in R$, then also $[d(x), x]^2 \in Z(R)$, for all $x \in R$.

We will prove that:

Theorem 1.1. *Let R be a prime ring of characteristic different from* 2*, L a noncentral Lie ideal of R, d a nonzero derivation of R such that $[d(u), u]^n \in Z(R)$, for any $u \in L$. Then R satisfies s_4.*

We will proceed by first proving that:

Lemma 1.1. *Let R be a prime ring of characteristic different from* 2*, L a noncentral Lie ideal of R, d a nonzero derivation of R, $n \geq 1$. If d satisfies $[d(u), u]^n = 0$, for any $u \in L$, then R is commutative.*

We then examine the case R is a 2-torsion free semiprime ring. The results we obtain are:

Theorem 2.1. *Let R be a 2-torsion free semiprime ring, d a nonzero derivation of R, n a fixed positive integer, U the left Utumi quotient ring of R and $[d([x,y]),[x,y]]^n = 0$, for any $x,y \in R$. Then there exists a central idempotent element e of U such that on the direct sum decomposition $eU \oplus (1-e)U$, d vanishes identically on eU and the ring $(1-e)U$ is commutative.*

Theorem 2.2. *Let R be a 2-torsion free semiprime ring, d a nonzero derivation of R, n a fixed positive integer, U the left Utumi quotient ring of R and $[d([x,y]),[x,y]]^n \in Z(R)$, for any $x,y \in R$. Then there exists a central idempotent e of U such that, on the direct sum decomposition $U = eU \oplus (1-e)U$, the derivation d vanishes identically on eU and the ring $(1-e)U$ satisfies s_4.*

1. The case: R prime ring.

In all that follows, unless stated otherwise, R will be a prime ring of characteristic $\neq 2$, L a Lie ideal of R, $d \neq 0$ a derivation of R and $n \geq 1$ a fixed integer such that $[d(x),x]^n \in Z(R)$, for all $x \in L$.

For any ring S, $Z(S)$ will denote its center, and $[a,b] = ab - ba$, $[a,b]_2 = [[a,b],b]$, $a,b \in S$. In addition s_4 will denote the standard identity in 4 variables.

We will also make frequent use of the following result due to Kharchenko [**8**] (see also [**12**]):

Let R be a prime ring, d a nonzero derivation of R and I a nonzero two-sided ideal of R. Let $f(x_1,\dots,x_n,d(x_1,\dots,x_n))$ a differential identity in I, that is

$$f(r_1,\dots,r_n,d(r_1),\dots,d(r_n)) = 0 \quad \forall r_1,\dots,r_n \in I.$$

One of the following holds:

1) Either d is an inner derivation in Q, the Martindale quotient ring of R, in the sense that there exists $q \in Q$ such that $d = ad(q)$ and $d(x) = ad(q)\,(x) = [q,x]$, for all $x \in R$, and I satisfies the generalized polynomial identity

$$f(r_1,\dots,r_n,[q,r_1],\dots,[q,r_n]) = 0;$$

2) or I satisfies the generalized polynomial identity

$$f(x_1,\dots,x_n,y_1,\dots,y_n) = 0.$$

Lemma 1.1. *Let R be a prime ring of characteristic different from 2, U a noncentral Lie ideal of R, d a nonzero derivation of R and $n \geq 1$. If $([d(u),u])^n = 0$, for any $u \in L$, then R is commutative.*

Proof. Since we assume that char $(R) \neq 2$, by a result of Herstein [**6**], $L \supseteq [I, R]$, for some $I \neq 0$, an ideal of R, and also L is not commutative. Therefore we will assume throughout that $L \supseteq [I, R]$. Without loss of generality we can assume $L = [I, I]$.

Hence $[d([x, y]), [x, y]]^n = 0$, for any $x, y \in I$, then I satisfies the differential identity

$$f(x, y, d(x), d(y)) = [[d(x), y] + [x, d(y)], [x, y]]^n = 0.$$

If the derivation d is not inner, by Kharchenko's theorem [**8**], I satisfies the polynomial identity

$$f(x, y, t, z) = [[z, y] + [x, t], [x, y]]^n = 0$$

and in particular, for $z = 0$,

$$[[x, t], [x, y]]^n = 0.$$

Since the latter is a polynomial identity for I, and so for R too, it is well known that there exists a field F such that R and F_m satisfy the same polynomial identities (see [**7**, page 57, page 89]). Let e_{ij} the matrix unit with 1 in (i,j)-entry and zero elsewhere. Suppose $m \geq 2$. If we choose $x = e_{11}$, $y = e_{21}$, $t = e_{12}$, then we get the contradiction

$$0 = [[e_{11}, e_{12}], [e_{11}, e_{21}]]^n = [e_{12}, -e_{21}]^n = (-1)^n e_{11} + e_{22} \neq 0.$$

Therefore $m = 1$ and so R is commutative.

Let now d be an inner derivation induced by an element $A \in Q$, the Martindale quotient ring of R. Then, for any $x, y \in I$, $([A, [x, y]]_2)^n = 0$. Since by [**2**] I and Q satisfy the same generalized polynomial identities, we have $([A, [x, y]]_2)^n = 0$, for any $x, y \in Q$. Moreover, since Q remains prime by the primeness of R, replacing R by Q we may assume that $A \in R$ and C is just the center of R. Note that R is a centrally closed prime C-algebra in the present situation [**4**], i.e., $RC = R$. By Martindale's theorem in [**13**], RC (and so R) is a primitive ring which is isomorphic to a dense ring of linear transformations of a vector space V over a division ring D. Since R is primitive then there exist a vector space V and the division ring D such that R is dense of D-linear transformation over V.

Assume first that $\dim_D V \geq 3$.

Step 1.

We want to show that, for any $v \in V$, v and Av are linearly D-dependent.

Since if $Av = 0$ then $\{v, Av\}$ is D-dependent, suppose that $Av \neq 0$. If v and Av are D-independent, since $\dim_D V \geq 3$, then there exists $w \in V$ such that v, Av, w are also linearly independent. By the density of I, there exist $x, y \in I$ such that

$$xv = 0,\ xAv = w,\ xw = v$$

$$yv = 0,\ yAv = 0,\ yw = w.$$

These imply that

$$[A,[x,y]]_2 v = -v \text{ and } 0 = ([A,[x,y]]_2)^n v = (-1)^n v,$$

which is a contradiction.

So we can conclude that v are Av are linearly D-dependent, for all $v \in V$.

Step 2.

We show here that there exists $b \in D$ such that $Av = vb$, for any $v \in V$. Now choose $v, w \in V$ linearly independent. Since $\dim_D V \geq 3$, there exists $u \in V$ such that v, w, u are linearly independent. By Step 1, there exist $a_v, a_w, a_u \in D$ such that

$$Av = va_v,\ Aw = wa_w,\ Au = ua_u \text{ that is } A(v+w+u) = va_v + wa_w + ua_u.$$

Moreover $A(v+w+u) = (v+w+u)a_{v+w+u}$, for a suitable $a_{v+w+u} \in D$. Then $0 = v(a_{v+w+u} - a_v) + w(a_{v+w+u} - a_w) + u(a_{v+w+u} - a_u)$ and, because v, w, u are linearly independent, $a_u = a_w = a_v = a_{v+w+u}$. This completes the proof of Step 2.

Let now $r \in R$ and $v \in V$. By Step 2, $Av = vb, r(Av) = r(vb)$, and also $A(rv) = (rv)b$. Thus $0 = [A,r]v$, for any $v \in V$, that is $[A,r]V = 0$. Since V is a left faithful irreducible R-module, $[A,r] = 0$, for all $r \in R$, i.e., $A \in Z(R)$ and $d = 0$, which contradicts our hypothesis.

Therefore $\dim_D V$ must be ≤ 2. In this case R is a simple GPI ring with 1, and so it is a central simple algebra finite dimensional over its center. From Lemma 2 in [**10**] it follows that there exists a suitable field F such that $R \subseteq M_k(F)$, the ring of all $k \times k$ matrices over F, and moreover $M_k(F)$ satisfies the same generalized polynomial identity of R.

If we assume $k \geq 3$, by the same argument as in Steps 1 and 2, we get a contradiction.

Obviously if $k = 1$ then R is commutative. Thus we may assume $R \subseteq M_2(F)$, where $M_2(F)$ satisfies $([A,[x,y]]_2)^n = 0$.

Since for any $a, b \in M_2(F)$, $[a,b]^2 \in Z(R)$ then it follows easily that $([A,[x,y]]_2)^2 = 0$, for any $x, y \in M_2(F)$. Let $A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$. If we choose $x = e_{12}, y = e_{21}$ then we get:

$$[A, e_{11} - e_{22}]_2 = \begin{bmatrix} 0 & 4a_{12} \\ 4a_{21} & 0 \end{bmatrix}$$

$$0 = ([A, e_{11} - e_{22}]_2)^2 = \begin{bmatrix} 16(a_{12}a_{21}) & 0 \\ 0 & 16(a_{12}a_{21}) \end{bmatrix}.$$

Therefore either $a_{12} = 0$ or $a_{21} = 0$. Without loss of generality we can pick $a_{12} = 0$.

Now let $[x, y] = [e_{11}, e_{12} + e_{21}] = e_{12} - e_{21}$. In this case we have:

$$[A, e_{12} - e_{21}]_2 = \begin{bmatrix} 2(a_{22} - a_{11}) & -2a_{21} \\ -2a_{21} & 2(a_{11} - a_{22}) \end{bmatrix}$$

$$\left(\begin{bmatrix} 2(a_{22} - a_{11}) & -2a_{21} \\ -2a_{21} & 2(a_{11} - a_{22}) \end{bmatrix} \right)^2 = 0$$

that is

$$4(a_{21})^2 + 4(a_{11} - a_{22})^2 = 0$$

$$(a_{21})^2 = -(a_{22} - a_{11})^2 \quad \mathbf{(1)}.$$

On the other hand if $[x, y] = [e_{11}, e_{12} - e_{21}] = e_{12} + e_{21}$ then

$$([A, e_{12} + e_{21}]_2)^2 = \begin{bmatrix} 2(a_{11} - a_{22}) & -2a_{21} \\ 2a_{21} & 2(a_{22} - a_{11}) \end{bmatrix}$$

$$\left(\begin{bmatrix} 2(a_{11} - a_{22}) & -2a_{21} \\ 2a_{21} & 2(a_{22} - a_{11}) \end{bmatrix} \right)^2 = 0$$

that is

$$4(a_{22} - a_{11})^2 - 4(a_{21})^2 = 0$$

$$(a_{21})^2 = (a_{22} - a_{11})^2 \quad \mathbf{(2)}.$$

(1) and **(2)** imply that $a_{21} = 0$ and $a_{11} = a_{22}$ which means that A is a central matrix in $M_2(F)$, $A \in F$ and $d = 0$, a contradiction. Therefore $k = 1$, i.e., R is commutative. □

Lemma 1.2. *Let $R = M_k(F)$, the ring of $k \times k$ matrices over a field F of characteristic $\neq 2$. If $q \neq 0$ is a noncentral element of R such that $([q, [x, y]]_2)^n \in F$, for any $x, y \in R$, then $k \leq 2$.*

Proof. Suppose $k \geq 3$. Let i, j, r be distinct indices and $q = \sum a_{mn} e_{mn}$, with $a_{mn} \in F$. For simplicity we assume that $i = 1, j = 2, r = 3$. If we choose $[x, y] = [e_{12}, e_{23} - e_{31}] = e_{13} + e_{32}$, then

$$[q, [x, y]]_2 = a_{21} e_{11} + a_{21} e_{22} - 2a_{21} e_{33} + \sum_{n \neq 1} \gamma_n e_{1n} + \sum_{m \neq 2} \delta_m e_{m2}$$

with $\gamma_n, \delta_m \in F$, and

$$([q, [x, y]]_2)^n = (a_{21})^n e_{11} + (a_{21})^n e_{22} + (-2a_{21})^n e_{33} + \sum_{n \neq 1} \alpha_n e_{1n} + \sum_{m \neq 2} \beta_m e_{m2}$$

with $\alpha_n, \beta_m \in F$. Since by assumption $([q, [x, y]]_2)^n \in F$, then $\alpha_n = \beta_m = 0$, for all m, n, and $(a_{21})^n = (-2a_{21})^n = 0$, i.e., $a_{21} = 0$. In a similar way we may conclude that $a_{ij} = 0$, for any $i \neq j$. Therefore if $k \geq 3$, q is a diagonal matrix, $q = \sum_t a_{tt} e_{tt}$, with $a_t \in F$.

If we show that q is a central matrix, then we get a contradiction to our assumption and so k must be less or equal than 2.

Let $[x,y] = [e_{ij} - e_{ji}, e_{jj}] = e_{ij} + e_{ji}$. Therefore

$$[q,[x,y]]_2 = 2(a_{ii} - a_{jj})e_{ii} + 2(a_{jj} - a_{ii})e_{jj}$$

and

$$([q,[x,y]]_2)^n = 2^n(a_{ii} - a_{jj})^n e_{ii} + 2^n(a_{jj} - a_{ii})^n e_{jj}.$$

Since $([q,[x,y]]_2)^n \in F$ and $k \geq 3$, it follows that $a_{ii} = a_{jj}$. Thus q is a central matrix.

Notice that if $n = 1$ then by using the same argument and choosing $[x,y] = e_{12}$, we get $N = [q,[x,y]]_2 = -2e_{12}qe_{12}$, which has rank 1 and so it cannot be central in $M_k(F)$, with $k \geq 2$. This implies that if $n = 1$ then $k = 1$, and R must be a commutative field. The proof of Lemma 1.2 is now complete. □

Theorem 1.1. *Let R be a prime ring of characteristic different from 2, L a noncentral Lie ideal of R, d a nonzero derivation of R such that $[d(u),u]^n \in Z(R)$, for any $u \in L$. Then R satisfies s_4.*

Proof. Let I be the nonzero two-sided ideal of R such that $0 \neq [I,R] \subseteq L$ and J be any nonzero two-sided ideal of R. Then $V = [I,J^2] \subseteq L$ is a Lie ideal of R. If, for every $v \in V$, $[d(v),v]^n = 0$, by Lemma 1.1, R is commutative. Otherwise, by our assumptions, $J \cap Z(R) \neq 0$. Let now K be a nonzero two-sided ideal of R_Z, the ring of the central quotients of R. Since $K \cap R$ is an ideal of R then $K \cap R \cap Z(R) \neq 0$, that is K contains an invertible element in R_Z, and so R_Z is simple with 1.

Moreover we may assume $L = [I,I]$. For any $x,y \in I$, $[d([x,y]),[x,y]]^n \in Z(R)$, i.e.,

$$[[d([x,y]),[x,y]]^n, r] = 0 \quad \text{for any } x \in R.$$

Thus I satisfies the differential identity

$$f(x,y,r,d(x),d(y)) = [[[d(x),y] + [x,d(y)],[x,y]]^n, r] = 0.$$

If the derivation is not inner, by [**8**], I satisfies the polynomial identity

$$f(x,y,r,z,t) = [[[t,y] + [x,z],[x,y]]^n, r] = 0$$

and in particular, for $z = 0$,

$$[[[t,y],[x,y]]^n, r] = 0.$$

In this case we know that there exists a field F such that R and F_m satisfy the same polynomial identities. Thus $[[t,y],[x,y]]^n$ is central in F_m. Suppose $m \geq 3$ and choose $x = e_{32}, y = e_{33}, t = e_{23}$.

$$[t,y] = e_{23}, \;\; [x,y] = -e_{32}$$

$$[[t,y],[x,y]] = -e_{22} + e_{33}$$

$$[[t,y],[x,y]]^n = (-1)^n e_{22} + e_{33} \notin Z(R)$$

contrary to our assumptions. This forces $m \leq 2$, i.e., R satisfies s_4.

Notice that in the case $n = 1$, $[[t, y], [x, y]]$ must be central in F_m. But if $m \geq 2$ and $t = e_{11}$, $y = e_{12}$, $x = e_{21}$, we get the contradiction $[[t, y], [x, y]] = 2e_{12} \notin Z(R)$. Therefore m must be equal to 1 and R is commutative.

Now let d be an inner derivation induced by an element $A \in Q$. By localizing R at $Z(R)$ it follows that $([A, [x, y]]_2)^n \in Z(R_Z)$, for all $x, y \in R_Z$.

Since R and R_Z satisfy the same polynomial identities, in order to prove that R satisfies $S_4(x_1, x_2, x_3, x_4)$, we may assume that R is simple with 1 and $[R, R] \subseteq L$.

In this case, $([A, [x, y]]_2)^n \in Z(R)$, for all $x, y \in R$. Therefore R satisfies a generalized polynomial identity and it is simple with 1, which implies that $Q = RC = R$ and R has a minimal right ideal. Thus $A \in R = Q$ and R is simple artinian that is $R = D_k$, where D is a division ring finite dimensional over $Z(R)$ [**13**]. From Lemma 2 in [**10**] it follows that there exists a suitable field F such that $R \subseteq M_k(F)$, the ring of all $k \times k$ matrices over F, and moreover $M_k(F)$ satisfies the generalized polynomial identity $[([A, [x, y]]_2)^n, z] = 0$. By Lemma 1.2, if $n \geq 2$ then $k \leq 2$ and R satisfies s_4, also if $n = 1$ then $k = 1$ and R must be commutative. □

2. The case: R semiprime ring.

In all that follows R will be a 2-torsion free semiprime ring. We cannot expect the same conclusion of previous section to hold, as the following example shows:

Example 2. Let R_1 be any prime ring not satisfying s_4 and $R_2 = M_2(F)$, the ring of 2×2 matrices over the field F. Let $R = R_1 \oplus R_2$, d a nonzero derivation of R such that $d = 0$ in R_1. Consider $L = [R, R]$. It is a non-central Lie ideal of R. Let $r_1, s_1 \in R_1$, $r_2, s_2 \in R_2$, $u = [(r_1, r_2), (s_1, s_2)]$. Therefore $d(u) = (0, d([r_2, s_2]))$ and $[d(u), u] = (0, [d([r_2, s_2]), [r_2, s_2]])$. Since $[d([r_2, s_2]), [r_2, s_2]]^2 \in Z(R_2)$, then

$$[d(u), u]^2 = (0, [d([r_2, s_2]), [r_2, s_2]])^2 = (0, [d([r_2, s_2]), [r_2, s_2]]^2) \in Z(R)$$

but R does not satisfy s_4.

The related object we need to mention is the left Utumi quotient ring U of R. For basic definitions and preliminary results we refer the reader to [**1**], [**5**], [**9**].

In order to prove the main result of this section we will make use of the following facts:

Claim 1 ([**1**, Proposition 2.5.1])**.** Any derivation of a semiprime ring R can be uniquely extended to a derivation of its left Utumi quotient ring U, and so any derivation of R can be defined on the whole U.

Claim 2 ([**3**, p. 38])**.** If R is semiprime then so is its left Utumi quotient ring. The extended centroid C of a semiprime ring coincides with the center of its left Utumi quotient ring.

Claim 3 ([**3**, p. 42])**.** Let B be the set of all the idempotents in C, the extended centroid of R. Assume R is a B-algebra orthogonal complete. For any maximal ideal P of B, PR forms a minimal prime ideal of R, which is invariant under any derivation of R.

We will prove the following:

Theorem 2.1. *Let R be a 2-torsion free semiprime ring, d a nonzero derivation of R, n a fixed positive integer, U the left Utumi quotient ring of R and $[d([x, y]), [x, y]]^n = 0$, for any $x, y \in R$. Then there exists a central idempotent element e of U such that on the direct sum decomposition $eU \oplus (1-e)U$, d vanishes identically on eU and the ring $(1-e)U$ is commutative.*

Proof. Since R is semiprime, by Claim 2, $Z(U) = C$, the extended centroid of R, and, by Claim 1, the derivation d can be uniquely extended on U. Since U and R satisfy the same differential identities (see [**12**]), then $[d([x, y]), [x, y]]^n = 0$, for all $x, y \in U$. Let B be the complete boolean algebra of idempotents in C and M be any maximal ideal of B.

Since U is a B-algebra orthogonal complete (see [**3**, p. 42, (2) of Fact 1]), by Claim 3, MU is a prime ideal of U, which is d-invariant. Denote $\overline{U} = U/MU$ and $\overline{d}$ the derivation induced by d on $\overline{U}$. For any $\overline{x}, \overline{y} \in \overline{U}$, $[\overline{d}([\overline{x}, \overline{y}]), [\overline{x}, \overline{y}]]^n = 0$. In particular $\overline{U}$ is a prime ring and so, by Lemma 1.1, $\overline{d} = 0$ in $\overline{U}$ or $\overline{U}$ is commutative. This implies that, for any maximal ideal M of B, $d(U) \subseteq MU$ or $[U, U] \subseteq MU$. In any case $d(U)[U, U] \subseteq MU$, for all M. Therefore $d(U)[U, U] \subseteq \bigcap_M MU = 0$.

By using the theory of orthogonal completion for semiprime rings (see [**1**, Chapter 3]), it follows that there exists a central idempotent element e in U such that on the direct sum decomposition $eU \oplus (1 - e)U$, d vanishes identically on eU and the ring $(1 - e)U$ is commutative. □

We come now to our last result:

Theorem 2.2. *Let R be a 2-torsion free semiprime ring, d a nonzero derivation of R, n a fixed positive integer, U the left Utumi quotient ring of R and $[d([x, y]), [x, y]]^n \in Z(R)$, for any $x, y \in R$. Then there exists a central idempotent e of U such that, on the direct sum decomposition $U = eU \oplus (1 - e)U$, the derivation d vanishes identically on eU and the ring $(1 - e)U$ satisfies s_4.*

Proof. By Claim 2, $Z(U) = C$, and by Claim 1 d can be uniquely defined on the whole U. Since U and R satisfy the same differential identities, then $[d([x,y]),[x,y]]^n \in C$, for all $x, y \in U$. Let B be the complete boolean algebra of idempotents in C and M any maximal ideal of B. As already pointed out in the proof of Theorem 2.1, U is a B-algebra orthogonal complete and by Claim 3, MU is a prime ideal of U, which is d-invariant. Let $\overline{d}$ the derivation induced by d on $\overline{U} = U/MU$. Since $Z(\overline{U}) = (C + MU)/MU = C/MU$, then $[\overline{d}([x,y]),[x,y]]^n \in (C + MU)/MU$, for any $x, y \in \overline{U}$. Moreover $\overline{U}$ is a prime ring, hence we may conclude, by Theorem 1.1, that $\overline{d} = 0$ in $\overline{U}$ or $\overline{U}$ satisfies s_4. This implies that, for any maximal ideal M of B, $d(U) \subseteq MU$ or $s_4(x_1, x_2, x_3, x_4) \subseteq MU$, for all $x_1, x_2, x_3, x_4 \in U$. In any case $d(U)s_4(x_1, x_2, x_3, x_4) \subseteq \bigcap_M MU = 0$. From [**1**, Chapter 3], there exists a central idempotent element e of U, the left Utumi quotient ring of R, such that there exists a central idempotent e of U such that $d(eU) = 0$ and $(1-e)U$ satisfies s_4. □

References

[1] K.I. Beidar, W.S. Martindale and V. Mikhalev, *Rings with generalized identities*, Pure and Applied Math., Dekker, New York, 1996.

[2] C.L. Chuang, *GPI's having coefficients in Utumi quotient rings*, Proc. Amer. Math. Soc., **103**(**3**) (1988), 723-728.

[3] ———, *Hypercentral derivations*, J. Algebra, **166** (1994), 34-71.

[4] J.S. Erickson, W.S. Martindale III and J.M. Osborn, *Prime nonassociative algebras*, Pacific J. Math., **60** (1975), 49-63.

[5] C. Faith, *Lecture on Injective Modules and Quotient Rings*, Lecture Notes in Mathematics, **49**, Springer Verlag, New York, 1967.

[6] I.N. Herstein, *Topics in ring theory*, Univ. Chicago Press, 1969.

[7] N. Jacobson, *PI-algebras, an introduction*, Lecture notes in Math., **441**, Springer Verlag, New York, 1975.

[8] V.K. Kharchenko, *Differential identities of prime rings*, Algebra and Logic, **17** (1978), 155-168.

[9] J. Lambek, *Lecture on Rings and Modules*, Blaisdell Waltham, MA, 1966.

[10] C. Lanski, *An Engel condition with derivation*, Proc. Amer. Math. Soc., **118**(**3**) (1993), 731-734.

[11] ———, *Differential identities, Lie ideals and Posner's theorems*, Pacific J. Math., **134**(**2**) (1988), 275-297.

[12] T.K. Lee, *Semiprime rings with differential identities*, Bull. Inst. Math. Acad. Sinica, **20**(**1**) (1992), 27-38.

[13] W.S. Martindale III, *Prime rings satisfying a generalized polynomial identity*, J. Algebra, **12** (1969), 576-584.

[14] E.C. Posner, *Derivations in prime rings*, Proc. Amer. Math. Soc., **8** (1957), 1093-1100.

Received August 19, 1998 and revised November 20, 1998.

Dipartimento di Matematica ed Applicazioni
Università di Palermo
90123 Palermo
Italy
E-mail address: lcarini@dipmat.unime.it

Dipartimento di Matematica
Università di Messina
98166 Messina
Italy
E-mail address: enzo@dipmat.unime.it

PACIFIC JOURNAL OF MATHEMATICS
Vol. 193, No. 2, 2000

GROUPS OF LINEAR ISOMETRIES ON MULTIPLIER C*-ALGEBRAS

Claudio D'Antoni and László Zsidó

For A C^*-algebra and $M(A)$ its multiplier algebra, the weak topologies $\sigma(M(A), A^*)$ and $\sigma(A^*, M(A))$ are shown to have the Krein property, claiming the compactness of the closed convex hull of every compact set. This has relevant consequences concerning the analytic generator of strictly continuous one-parameter groups of strictly continuous linear operators on $M(A)$.

Furthermore, it is proved that there exists an one-to-one correspondence between surjective linear isometries on A and strictly bicontinuous, surjective linear isometries on $M(A)$, as well as between strongly continuous respectively strictly continuous locally compact groups of them. In the case of connected groups, they all arise from $*$-automorphism groups by perturbation with a cocycle.

Introduction.

Recently the search for a sound C^*-algebraic framework for quantum groups has renewed the attention for analysis on multiplier C^*-algebras: The present setting requires a more precise understanding of groups of linear isometries on multiplier C^*-algebras. Of particular relevance seems to be the structure of the analytic generator in the one-parameter case (see e.g., [**Kus**]).

The multiplier algebra $M(A)$ of C^*-algebra A is the non-commutative generalization of the Stone-Čech compactification $\beta\Omega$ of a locally compact topological space Ω. In the commutative setting the canonical pairing between the bounded continuous functions on Ω, identified with the continuous functions on $\beta\Omega$, and the bounded regular Borel measures on Ω has been intensively investigated (see e.g., [**Cnw**] and [**HJ**]). In the non-commutative frame the analogous pairing between $M(A)$ and the dual A^* is obtained from the natural duality between A^{**} and A^* considering $M(A)$ embedded in A^{**}. There exists a natural topology on $M(A)$, called the strict topology, which is compatible with this duality between $M(A)$ and A^*.

Our first goal is to prove in Section 1 that the corresponding weak topologies on $M(A)$ and A^* have the so-called Krein property (indicated in [**C-Z**] as axiom A_1), claiming the compactness of the closed convex hull of every compact set. This entails the Pettis integrability of the $\sigma(M(A), A^*)$-continuous $M(A)$-valued respectively $\sigma(A^*, M(A))$-continuous A^*-valued functions and allows to apply the results from [**C-Z**] and [**Z1**] to the one-parameter operator groups on $M(A)$. We notice that for commutative A the Krein property for $\sigma(M(A), A^*)$ and $\sigma(A^*, M(A))$ is already known (see [**HJ**]).

Subsequently, in Section 2 we investigate the extendibility of bounded linear maps $\Phi : A \to B$, A and B C^*-algebras, to strictly continuous linear maps $M(A) \to M(B)$. We prove that for Φ Jordan $*$-homomorphism, whose range generates B, this extension exists and it is the only Jordan $*$-homomorphism $M(\Phi) : M(A) \to M(B)$ extending Φ. A similar result holds also for surjective linear isometries. Moreover, any strictly bicontinuous surjective linear isometry $M(A) \to M(B)$ maps A onto B, hence it is an extension of a surjective linear isometry $A \to B$. For A and B separable, making use of a result by L.G. Brown (see [**Br**]), we get the automatic strict bicontinuity of all surjective linear isometries $M(A) \to M(B)$. The same holds, for a different reason, if A and B are simple.

In Section 3 we study families $(\Phi_t)_t$ of linear isometries depending on a parameter t. We prove that continuous dependence in the strong operator topology of $(\Phi_t)_t$ goes in pointwise strictly continuous dependence of $M(\Phi_t)_t$. As a consequence, we get a one-to-one correspondence between the strongly continuous representations of a locally compact group G by linear isometries on a C^*-algebra A and the strictly continuous representations of G by strictly bicontinuous (automatic, if A separable or simple!) linear isometries on $M(A)$. In the one-parameter case the graph of the analytic generator of the extension turns out to be the strict closure of the graph of the analytic generator of the original group.

Using results due to R.V. Kadison (see [**Kad2**]) we prove also a structure theorem for strongly continuous representations of connected topological groups G by linear isometries on a C^*-algebra A. Namely, they all arise from strongly continuous representations of G by $*$-automorphisms of A, perturbing them with a cocycle. We notice that the W^*-algebra counterpart of this result holds only under additional assumptions, for example, assuming that the centre of the algebra is atomic.

Finally we show that representations of locally compact groups by linear isometries on separable C^*-algebras are strongly continuous under minimal regularity assumptions and they induce strongly continuous representations on separable, invariant C^*-subalgebras of the corona algebra.

1. The Krein Property.

A relevant property for a locally convex topological vector space X is the following one:

(K) the closed convex hull of every weakly compact subset of X is compact.

According to a well known result of M.G. Krein, V.L. Smulian and A. Grothendieck (see e.g., [**Sch**], Th. IV.11.4), the closed convex hull of a weakly compact subset of X is weakly compact if (and only if) it is complete with respect to the associated Mackey topology. Therefore let us call (K) the "Krein property".

We notice two consequences of the Krein property:

If X has the Krein property then every weakly continuous map $\Omega \to X, \Omega$ compact topological space, is Pettis integrable with respect to any Radon measure on Ω (see e.g., [**Rud**], Th. 3.27, [**Sch**], Exercise IV.39 (a), [**Ar**], Prop. 1.2, [**C-Z**], Prop. 1.4).

A second consequence concerns "dual pairs of Banach spaces", that is pairs $(X, \mathcal{F})$ of Banach spaces together with a bilinear functional

$$X \times \mathcal{F} \ni (x, \varphi) \mapsto \langle x, \varphi \rangle$$

satisfying

$$\|x\| = \sup\{|\langle x, \varphi\rangle|;\ \varphi \in \mathcal{F},\ \|\varphi\| \leq 1\}, \quad x \in X,$$
$$\|x\| = \sup\{|\langle x, \varphi\rangle|;\ x \in \mathcal{F},\ \|x\| \leq 1\}, \quad \varphi \in \mathcal{F}.$$

If X, endowed with the weak topology $\sigma(X, \mathcal{F})$, and $\mathcal{F}$, endowed with $\sigma(\mathcal{F}, X)$, have the Krein property then, for any $\sigma(X, \mathcal{F})$-continuous one-parameter group $t \mapsto U_t$ of $\sigma(X, \mathcal{F})$-continuous linear maps on X, the analytic extension operators $U_z, z \in \mathbb{C}$, are (besides being $\sigma(X, \mathcal{F})$-densely defined, a consequence of the Pettis integrability of U) $\sigma(X, \mathcal{F})$-closed (see [**C-Z**], Th. 2.4). Moreover, denoting by $U^{\mathcal{F}}$ the adjoint group on $\mathcal{F}$, $U_z^{\mathcal{F}}$ is the adjoint of U_z in $\mathcal{F}$ (see [**Z1**], Th. 1.1).

If $\mathcal{F} = X^*$ then, according to the Krein theorem, the principle of uniform boundedness and the Alaoglu theorem, $\sigma(X, \mathcal{F})$ and $\sigma(\mathcal{F}, X)$ have the Krein property. We shall prove this for an other kind of dual pair of Banach spaces, in general not of the above form: We shall deal with dual pairs consisting of the multiplier algebra $M(A)$ and the dual space A^* of a C^*-algebra A.

The multiplier algebra $M(A)$, first considered in [**Buc**] (for commutative A) and [**Bus**] (for general C^*-algebra A), is ($*$-isomorphic to) the C^*-subalgebra

$$\{x \in A^{**};\ xa,\ ax \in A \text{ for all } a \in A\}$$

of the second dual A^{**}. For its basic theory we send to [**Ped1**], 3.12 and [**WO**], Chapter 2.

The canonical duality between A^{**} and A^* induces a pairing

$$M(A) \times A^* \ni (x, \varphi) \mapsto \langle x, \varphi \rangle$$

which makes $(M(A), A^*)$ a dual pair of Banach spaces. We recall that the strict topology β on $M(A)$ is the locally convex vector space topology defined by the seminorms

$$x \mapsto \|xa\| \text{ and } x \mapsto \|ax\|, \quad a \in A.$$

It is complete and compatible with the duality between $M(A)$ and A^* (see [**T1**], Cor. 2.3). Hence the strict topology is weaker that the norm-topology on $M(A)$, but stronger than the restriction to $M(A)$ of the weak * topology of A^{**}. A is a strictly dense, norm-closed two-sided ideal of $M(A)$. More precisely, if $(u_\iota)_\iota$ is any two-sided approximate unit for A then $u_\iota \to 1_{A^{**}} \in M(A)$ strictly.

The goal of this section is to prove:

Theorem 1.1 (on the Krein property for multipliers)**.** *Let A be an arbitrary C^*-algebra. Then $M(A)$ with $\sigma(M(A), A^*)$ and A^* with $\sigma(A^*, M(A))$ have the Krein property.*

Since the strict topology is complete, also the Mackey topology $\tau(M(A), A^*)$ is complete (see e.g., [**B**], IV.5, Remark 2). Therefore the Krein property of $\sigma(M(A), A^*)$ follows directly from the Krein theorem.

The proof of the Krein property of $\sigma(A^*, M(A))$ is more involved. Let us shortly comment the already known results concerning it. We notice that the Krein property for $\sigma(A^*, M(A))$ is equivalent to saying that all seminorms

$$M(A) \in x \mapsto \sup\{|\langle x, \varphi\rangle|;\ \varphi \in \mathcal{K}\}$$

with $\mathcal{K} \subset A^*$ $\sigma(A^*, M(A))$-compact are $\tau(M(A), A^*)$-continuous, that is to the $\tau(M(A), A^*)$-equicontinuity of any $\sigma(A^*, M(A))$-compact subset of A^*.

If τ is a locally convex vector space topology on $M(A)$, compatible with the duality between $M(A)$ and A^*, then $M(A), \tau$ is usually called strong Mackey space whenever any $\sigma(A^*, M(A))$-compact subset of A^* is τ-equicontinuous. In other words, $M(A), \tau$ is strong Mackey if and only if $\tau = \tau(M(A), A^*)$ and $\sigma(A^*, M(A))$ has the Krein property.

There are criteria in order that the strict topology β on $M(A)$ be strong Mackey: This happens for commutative A whenever the Gelfand spectrum of A is paracompact ([**Cnw**], Th. 2.6) and, more generally, for arbitrary A, whenever A has a "well behaved" approximate unit ([**T2**], Cor. 3.4). In particular, in this case $\sigma(A^*, M(A))$ has the Krein property. However, even for commutative A, β is not always equal to $\tau(M(A)A^*)$, so $M(A), \beta$ is not always strong Mackey space (see [**Cnw**], Remarks on p. 481). Nevertheless, it was proved that $\sigma(A^*, M(A))$ has the Krein property for any commutative A (see [**HJ**], Th. 2). Actually this proof inspired our proof of the Krein property of $\sigma(A^*, M(A))$ for general A.

The main ingredient in proving the Krein property of $\sigma(A^*, M(A))$ is the following convergence result:

Lemma 1.2. *Let be A a C^*-algebra, $\mathcal{K}$ a $\sigma(A^*, M(A))$-compact subset of $A_h^* = \{\varphi \in A^*;\ \varphi = \varphi^*\}$, and μ a positive Radon measure on the compact topological space $\mathcal{K}, \sigma(A^*, M(A))$. Then, for any (increasing, positive) approximate unit $(u_\iota)_\iota$ of A and any $x \in M(A)$,*

$$\int_{\mathcal{K}} \langle x^* u_\iota x, \varphi\rangle d\mu(\varphi) \xrightarrow{\iota} \int_{\mathcal{K}} \langle x^* x, \varphi\rangle d\mu(\varphi).$$

We notice that if $\mathcal{K}$ would be contained in A_+^* then, according to the Dini theorem, we would have $\langle x^* u_\iota x, \varphi\rangle \xrightarrow{\iota} \langle x^* x, \varphi\rangle$ uniformly for $\varphi \in \mathcal{K}$ and the statement of the Lemma would follow trivially.

Proof. First we consider the case $x = 1_{A^{**}}$.

Let us define the lower $\sigma(A^*, M(A))$-semicontinuous functions $g_\iota : \mathcal{K} \to \mathbb{R}$ by

$$g_\iota(\varphi) = \sup\{|\langle a, \varphi\rangle|;\ a^* = a \in A,\ -u_\iota \le a \le u_\iota\} \le \|\varphi\|.$$

Then $(g_\iota)_\iota$ is upward directed and pointwise convergent to $\mathcal{K} \in \varphi \mapsto \|\varphi\|$, so

$$\int_{\mathcal{K}} g_\iota(\varphi) d\mu(\varphi) \xrightarrow{\iota} \int_{\mathcal{K}} \|\varphi\| d\mu(\varphi). \tag{1.1}$$

For let $\varphi \in \mathcal{K}$ and $\varepsilon > 0$ be arbitrary. Choosing $a^* = a \in A, \|a\| \le 1$, with

$$|\langle a, \varphi\rangle| \ge \|\varphi\| - \frac{\varepsilon}{2},$$

we have

$$(u_\iota^{1/2} a u_\iota^{1/2})^* = u_\iota^{1/2} a u_\iota^{1/2} \in A,$$

so

$$\begin{aligned} g_\iota(\varphi) &\ge |\langle u_\iota^{1/2} a u_\iota^{1/2},\ \varphi\rangle| \\ &\ge |\langle a, \varphi\rangle| - |\langle a - u_\iota^{1/2} a u_\iota^{1/2},\ \varphi\rangle| \\ &\ge \|\varphi\| - \frac{\varepsilon}{2} - \|\varphi\| \cdot \|a - u_\iota^{1/2} a u_\iota^{1/2}\|. \end{aligned}$$

Since

$$\begin{aligned} \|a - u_\iota^{1/2} a u_\iota^{1/2}\| &\le \|(1_{A^{**}} - u_\iota^{1/2}) a\| + \|u_\iota^{1/2} a (1_{A^{**}} - u_\iota^{1/2})\| \\ &\le 2 \cdot \|a(1_{A^{**}} - u_\iota^{1/2})^2 a\|^{1/2} \\ &\le 2 \cdot \|a(1_{A^{**}} - u_\iota) a\|^{1/2} \\ &\le 2 \cdot \|a - u_\iota a\|^{1/2} \end{aligned}$$

and $(u_\iota)_\iota$ is an approximate unit for A, it follows the existence of some ι_ε such that

$$g_\iota(\varphi) \ge \|\varphi\| - \varepsilon \ \text{ for all } \ \iota \ge \iota_\varepsilon.$$

On the other hand,

$$|\langle 1_{A^{**}} - u_\iota, \varphi\rangle| \le \|\varphi\| - g_\iota(\varphi) \quad \text{for all} \quad \varphi \in \mathcal{K} \quad \text{and} \quad \iota.$$

Indeed, for every $a^* = a \in A$ with $-u_\iota \le a \le u_\iota$,

$$\begin{aligned} |\langle a, \varphi\rangle| &\le |\langle a, \varphi_+\rangle| + |\langle a, \varphi_-\rangle| \\ &\le \langle u_\iota, \varphi_+\rangle + \langle u_\iota, \varphi_-\rangle \\ &= \langle u_\iota, |\varphi|\rangle, \end{aligned}$$

$$\begin{aligned} |\langle 1_{A^{**}} - u_\iota, \varphi\rangle| + |\langle a, \varphi\rangle| &\le \langle 1_{A^{**}} - u_\iota, |\varphi|\rangle + \langle u_\iota, |\varphi|\rangle \\ &= \langle 1_{A^{**}}, |\varphi|\rangle \\ &= \|\varphi\|, \end{aligned}$$

hence

$$|\langle 1_{A^{**}} - u_\iota, \varphi\rangle| + g_\iota(\varphi) \le \|\varphi\|.$$

Therefore

$$\begin{aligned} &\left| \int_{\mathcal{K}} \langle 1_{A^{**}}, \varphi\rangle d\mu(\varphi) - \int_{\mathcal{K}} \langle u_\iota, \varphi\rangle d\mu(\varphi) \right| \\ &\le \int_{\mathcal{K}} |\langle 1_{A^{**}} - u_\iota, \varphi\rangle| d\mu(\varphi) \\ &\le \int_{\mathcal{K}} (\|\varphi\| - g_\iota(\varphi)) d\mu(\varphi) \end{aligned}$$

and (1.1) yields

$$\int_{\mathcal{K}} \langle u_\iota, \varphi\rangle d\mu(\varphi) \xrightarrow{\iota} \int_{\mathcal{K}} \langle 1_{A^{**}}, \varphi\rangle d\mu(\varphi).$$

Now let $x \in M(A)$ be arbitrary.

The map

$$\Phi_x : A_h^* \ni \varphi \mapsto \varphi(x^* \cdot x) \in A_h^*$$

is $\sigma(A^*, M(A))$-continuous, so $\mathcal{K}_x = \Phi_x(\mathcal{K})$ is a $\sigma(A^*, M(A))$-compact subset of A_h^*. Denoting by μ_x the image of μ under Φ, μ_x is a positive Radon measure on $\mathcal{K}_x, \sigma(A^*, M(A))$ and

$$\int_{\mathcal{K}_x} f(\psi) d\mu_x(\psi) = \int_{\mathcal{K}} f(\Phi_x(\varphi)) d\mu(\varphi)$$

for all $\sigma(A^*, M(A))$-continuous $f : \mathcal{K}_x \to \mathbb{R}$. Using the above equality with $f(\psi) = \langle u_\iota, \psi\rangle$ respectively $f(\psi) = \langle 1_{A^{**}}, \psi\rangle$ and applying the first part of the proof to $\mathcal{K}_x, \mu_x$, we get

$$\begin{aligned} &\int_{\mathcal{K}} \langle x^* u_\iota x, \varphi\rangle d\mu(\varphi) \\ &= \int_{\mathcal{K}} \langle u_\iota, \psi\rangle d\mu_x(\psi) \xrightarrow{\iota} \int_{\mathcal{K}_x} \langle 1_{A^{**}}, \psi\rangle d\mu_x(\psi) \end{aligned}$$

$$= \int_{\mathcal{K}} \langle x^* x, \varphi \rangle d\mu(\varphi).$$

□

Proof of the theorem. We have already seen that the completeness of the strict topology on $M(A)$ implies the Krein property for $\sigma(M(A), A^*)$.

The Krein property for $\sigma(A^*, M(A))$ means that every $\sigma(A^*, M(A))$-compact $\mathcal{K} \subset A^*$ is contained in a $\sigma(A^*, M(A))$-compact convex subset of A^*. Let us first consider the case of a $\sigma(A^*, M(A))$-compact $\mathcal{K} \subset A^*_h$.

Let $P(\mathcal{K})$ denote the weak* compact convex set of all probability Radon measures on $\mathcal{K}$, endowed with $\sigma(A^*, M(A))$. For every $\mu \in P(\mathcal{K})$ we can define $\Phi(\mu) \in A^*$ by putting

$$\langle a, \Phi(\mu) \rangle = \int_{\mathcal{K}} \langle a, \varphi \rangle d\mu(\varphi), \quad a \in A.$$

Then actually holds:

$$\langle x, \Phi(\mu) \rangle = \int_{\mathcal{K}} \langle x, \varphi \rangle d\mu(\varphi), \quad x \in M(A) \tag{1.2}$$

(in other words there exists the Pettis integral $\sigma(A^*, M(A)) - \int_{\mathcal{K}} \varphi d\mu(\varphi) \in A^*$). Indeed, choosing an approximate unit $(u_\iota)_\iota$ of A and applying the above lemma, we have for every $0 \leq x \in M(A)$

$$\langle x^{1/2} u_\iota x^{1/2}, \Phi(\mu) \rangle = \int_{\mathcal{K}} \langle x^{1/2} u_\iota x^{1/2}, \varphi \rangle d\mu(\varphi) \xrightarrow{\iota} \int_{\mathcal{K}} \langle x, \varphi \rangle d\mu(\varphi).$$

But on the other hand, since $x^{1/2} u_\iota x^{1/2} \to x$ in the weak* topology of A^{**}, we have also

$$\langle x^{1/2} u_\iota x^{1/2}, \Phi(\mu) \rangle \xrightarrow{\iota} \langle x, \Phi(\mu) \rangle.$$

By (1.2) the affine map $\Phi : P(\mathcal{K}) \to A^*$ is continuous with respect to the weak* topology on $P(\mathcal{K})$ and $\sigma(A^*, M(A))$ on A^*. Consequently $\Phi P(\mathcal{K})$ is a $\sigma(A^*, M(A))$-compact convex subset of A^* containing

$$\{\Phi(\delta_\varphi);\ \delta_\varphi \text{ the Dirac measure in } \varphi \in \mathcal{K}\} = \mathcal{K}.$$

Now let the $\sigma(A^*, M(A))$-compact set $\mathcal{K} \subset A^*$ be arbitrary. Since $A^* \ni \varphi \mapsto \varphi^* \in A^*$ is $\sigma(A^*, M(A))$-continuous, $\operatorname{Re}\mathcal{K}, \operatorname{Im}\mathcal{K} \subset A_h$ are also $\sigma(A^*, M(A))$-compact. According to the above part of the proof, there are $\sigma(A^*, M(A))$-compact convex sets $\operatorname{Re}\mathcal{K} \subset \mathcal{K}_1 \subset A^*$ and $\operatorname{Im}\mathcal{K} \subset \mathcal{K}_2 \subset A^*$. Then $\mathcal{K}_1 + i\mathcal{K}_2$ is a $\sigma(A^*, M(A))$-compact convex subset of A^* containing $\mathcal{K}$. □

By the above theorem the results from [**C-Z**] and [**Z1**] are available for $\sigma(M(A), A^*)$-continuous one-parameter groups of $\sigma(M(A), A^*)$-continuous linear operators on $M(A)$. In particular, [**C-Z**], Th. 2.4 and [**Z1**], Th. 1.1 yield:

Corollary. *Let A be a C^*-algebra, $t \to \alpha_t$ a $\sigma(M(A), A^*)$-continuous one-parameter group of $\sigma(M(A), A^*)$-continuous linear operators on $M(A), t \mapsto \alpha_t^*$ the adjoint group on A^*, and $z \in \mathbb{C}$. Define the linear operator α_z in $M(A)$ by putting $(x, y) \in$ graph (α_z) whenever $\mathbb{R} \ni t \mapsto \alpha_t(x) \in M(A)$ has a $\sigma(M(A), A^*)$-continuous extension on the closed horizontal strip between 0 and $\operatorname{Im} z$, which is analytic in the interior and takes the value y in z. The linear operator α_z^* in A^* is defined similarly. Then α_z is strictly densely defined, its graph is closed with respect to the product of the strict topologies on $M(A) \times M(A)$, and its adjoint in A^* is α_z^*.*

2. Jordan $*$-homomorphisms and linear isometries between multiplier algebras.

In this section we investigate the extendibility of Jordan $*$-homomorphisms and linear isometries between C^*-algebras to similar maps between the respective multiplier algebras. We also describe those linear isometries between multiplier algebras, which arise as extensions.

Let A, B be C^*-algebras. $\pi : A \to B$ is called Jordan $*$-homomorphism if it is a linear $*$-map satisfying

$$\pi(x^2) = \pi(x)^2, \quad x \in A.$$

It is well known that then (see e.g., [**S-Z**], 6.6)

$$\pi(xy + yx) = \pi(x)\pi(y) + \pi(y)\pi(x),$$
$$\pi(xyx) = \pi(x)\pi(y)\pi(x),$$
$$xy = yx \Rightarrow \pi(xy) = \pi(x)\pi(y)$$

with x, y elements of A. The last statement implies immediately that if A is unital then $\pi(1_A)$ is unit for the hereditary C^*-subalgebra $\operatorname{Her}_B(\pi(A)) \subset B$ generated by $\pi(A)$. In particular, π being positive, $\|\pi\| = \|\pi(1_A)\| = 1$ or 0. For arbitrary A, the positive map π being bounded, we can consider the Jordan $*$-homomorphism $\pi^{**} : A^{**} \to B^{**}$ and the above remarks yields $\|\pi\| = \|\pi^{**}\| = 1$ or 0.

Let $\pi : A \to B$ be a Jordan $*$-homomorphism. Then

$$\|\pi(x)\pi(y)\| \leq \|xy\|, \quad x, y \in A_h, \tag{2.1}$$

where A_h denotes the Hermitian part $\{a \in A; a = a^*\}$ of A. Indeed,

$$\|\pi(x)\pi(y)\|^2 = \|\pi(y)\pi(x)^2\pi(y)\| = \|\pi(yx^2y)\|$$
$$\leq \|yx^2y\| = \|xy\|^2.$$

Furthermore,

$$\pi(x)^*\pi(x) \leq \pi(x^*x + xx^*), \quad x \in A. \tag{2.2}$$

For we just have to notice that

$$\pi(x)^*\pi(x) \leq \pi(x)^*\pi(x) + \pi(x)\pi(x)^* = \pi(x^*x + xx^*).$$

It follows that

$$(2.3) \qquad \begin{cases} J \subset B \ \text{ norm closed two-sided ideal } \Rightarrow \\ \pi^{-1}(J) \subset A \ \text{ norm closed two-sided ideal.} \end{cases}$$

Indeed, for any $a^* = a \in \pi^{-1}(J)$ and $x^* = x \in A$ we have by (2.2)

$$\begin{aligned} \pi(ax)^*\pi(ax) &\leq \pi((ax)^*ax + ax(ax)^*) \\ &= \pi(xa^2x + ax^2a) \\ &= \pi(x)\pi(a)^2\pi(x) + \pi(a)\pi(x)^2\pi(a) \in J, \end{aligned}$$

so, according to [**Ped1**], Prop. 1.4.5,

$$\pi(ax) \in J, \ \text{ i.e., } \ ax \in \pi^{-1}(J).$$

In particular, $\operatorname{Ker}\pi = \pi^{-1}(\{0\})$ is a norm closed two-sided ideal in A. Thus π factorizes in

$$A \to A/\operatorname{Ker}\pi \xrightarrow{\tilde{\pi}} B$$

where the first arrow denotes the quotient $*$-homomorphism and $\tilde{\pi}$ is an injective Jordan*-homomorphism. If π is surjective then $\tilde{\pi}$ is Jordan $*$-isomorphism, hence isometrical. Therefore in this case

$$\|\pi(x)\|^2 = \|\pi(x^*x)\|, \quad x \in A.$$

The next result extends [**P-S**], Th. 2 and [**Ped2**], Th. 10 to Jordan *-homomorphisms between C^*-algebras. We recall that a C^*-algebra is called $\sigma-$unital whenever it contains a strictly positive element or, equivalently, it has a countable approximate unit (see [**Ped1**], 3.10.4, 3.10.5).

Proposition 2.1. *Let A, B be C^*-algebras, and $\pi : A \to B$ a Jordan $*$-homomorphism with $\operatorname{Her}_B(\pi(A)) = B$. Then*

(i) *there exists a unique extension of π to a Jordan $*$-homomorphism $M(\pi) : M(A) \to M(B)$, namely $M(\pi) = \pi^{**}|M(A)$, which is strictly continuous and unital, hence carrying two-sided approximate units of A in two-sided approximate units of B;*

(ii) *assuming that A is σ-unital, $M(\pi)$ is surjective whenever π is surjective.*

Proof. (i) Let us first prove that if $A \subset A_0 \subset A^{**}$ is a C^*-algebra, $\rho : A_0 \to B^{**}$ is a Jordan $*$-homomorphism extending π and $(x_\iota)_\iota$ is a norm bounded net in $(A_0)_h$ with $\|x_\iota a\| \xrightarrow{\iota} 0$ for all $a \in A$, then $\|\rho(x_\iota)b\| \xrightarrow{\iota} 0$ for all $b \in B$. Indeed, applying (2.1) to ρ, we have for every $a \in A_h$

$$\|\rho(x_\iota)\pi(a)\| = \|\rho(x_\iota)\rho(a)\| \leq \|x_\iota a\| \xrightarrow{\iota} 0.$$

Therefore the hereditary C^*-subalgebra

$$\{b \in B; \|\rho(x_\iota)b\| \xrightarrow{\iota} 0 \text{ and } \|b\rho(x_\iota)\| \xrightarrow{\iota} 0\} \subset B \text{ contains } \pi(A).$$

Now let $x^* = x \in M(A)$ be arbitrary. Choosing a norm bounded net $(a_\iota)_\iota$ in A_h such that $a_\iota \overset{\iota}{\longrightarrow} x$ strictly, by the above proved statement $\pi^{**}(x)b = \text{norm} - \lim\limits_\iota \pi(a_\iota)b \in B$ for all $b \in B$, hence $\pi^{**}(x) \in M(B)$. Thus $\pi^{**}M(A) \subset M(B)$.

Again by the first part of the proof, any Jordan $*$-homomorphism $\rho : M(A) \to M(B)$ extending π, in particular $M(\pi) = \pi^{**}|M(A)$, is strictly continuous on every norm bounded subset of $M(A)$. It follows that $M(\pi)$ is the only such ρ. Moreover, since the strict topology β on $M(A)$ is the finest locally convex vector space topology on $M(A)$ that agrees with β on the norm bounded subsets of $M(A)$ (see [**T1**], Cor. 2.7), $M(\pi)$ is actually strictly continuous.

On the other hand, $\pi^{**}(1_{A^{**}})$ is unit for $\text{Her}_{B^{**}}(\pi^{**}(A^{**})) \supset \text{Her}_B(\pi(A)) = B$, hence also for its weak* closure in B^{**}:

$$\pi^{**}(1_{A^{**}}) = 1_{B^{**}}.$$

In other words, $M(\pi)$ is unital.

(ii) Let

$$A \overset{\pi_0}{\to} A/\text{Ker}\pi \overset{\tilde{\pi}}{\to} B$$

be the canonical factorization of π. By Pedersen's Tietze type extension theorem (see [**Ped2**], Th. 10 or [**WO**], Th. 2.3.9), $M(\pi_0)$ is surjective. On the other hand, $\tilde{\pi}$ being Jordan $*$-isomorphism, $M(\tilde{\pi})$ is Jordan $*$-isomorphism. We conclude that $M(\pi) = M(\tilde{\pi})M(\pi_0)$ is surjective. □

Now we prove a partial converse to Proposition 2.1:

Proposition 2.2. *Let A, B be C^*-algebras. Then any strictly bicontinuous Jordan $*$-isomorphism $\rho : M(A) \to M(B)$ maps A onto B, so $\rho = M(\rho|A : A \to B)$. If A and B are both separable or both (topologically) simple, then we have this for any Jordan $*$-isomorphism $\rho : M(A) \to M(B)$, whose strict bicontinuity is hence automatical.*

Proof. Let $(u_\iota)_\iota$ and $(v_k)_k$ be increasing, positive approximate units for A respectively B. Since A is two-sided ideal in $M(A)$ and ρ^{-1} is strictly continuous,

$$A \ni w_{\iota,k} = u_\iota \rho^{-1}(v_k) u_\iota \overset{\iota,k}{\to} 1_{A^{**}} \quad \text{strictly}.$$

In other words $(w_{\iota,k})_{\iota,k}$ is a two-sided approximate unit for A. It follows for every $0 \le a \in A$

$$\|a - a^{1/2} w_{\iota,k} a^{1/2}\| \overset{\iota,k}{\to} 0,$$

$$\|\rho(a) - \rho(a^{1/2} w_{\iota,k} a^{1/2})\| \overset{\iota,k}{\to} 0,$$

so, having

$$\rho(a^{1/2} w_{\iota,k} a^{1/2}) = \rho(a^{1/2})\rho(u_\iota) v_k \rho(u_\iota)\rho(a^{1/2}) \in B$$

for all ι and k, $\rho(a) \in B$. Consequently

$$\rho(A) \subset B.$$

We get similarly also

$$\rho^{-1}(B) \subset A, \quad \text{i.e.,} \quad \rho(A) \supset B.$$

Now let us assume that A and B are separable and $\rho : M(A) \to M(B)$ is an arbitrary Jordan $*$-isomorphism. By [**Br**], Cor. 6 (see also the remarks after Corollary 1.4 in [**D-Z**]) A is the largest separable, norm closed, two-sided ideal of $M(A)$, so (2.3) implies that $\rho(A)$ is the largest separable, norm closed, two-sided ideal of $M(B)$. Using again [**Br**], Cor. 6, we conclude that $\rho(A) = B$.

Assuming finally that A and B are simple and nonzero, we can argue as above, using the fact that A is the smallest nonzero, closed, two-sided ideal of $M(A)$, and similarly for B. Indeed, is J is any nonzero, closed, two-sided ideal of $M(A)$, then the essentialness of A yields

$$A \cap J \supset AJ \neq \{0\}$$

and it follows by the simplicity of A that $A \cap J = A$. □

Now let $\Phi : A \to B$ be an isometrical linear bijection between C^*-algebras. Then there exist $\pi : A \to B$ Jordan $*$-isomorphism and $u \in M(B)$ unitary such that

$$\Phi(x) = u\pi(x), \quad x \in A \tag{2.4}$$

(see [**Kad1**] for the case of unital C^*-algebras and [**P-S**], Th. 1 for the general case). We notice that u, hence also π, is uniquely determined by Φ : according to Proposition 2.1., for any two-sided approximate unit $(u_\iota)_\iota$ of A we have

$$\pi(u_\iota) \to 1_{B^{**}} \text{ strictly,}$$

so

$$\Phi(u_\iota) \to u \text{ strictly.}$$

We call (2.4) the Kadison decomposition of Φ.

The following theorem is the main result of this section (cf. with [**P-S**], Th. 2):

Theorem 2.3 (on extension of linear isometries)**.** *Let A, B be C^*-algebras. Then every isometric linear bijection $\Phi : A \to B$ has a unique extension to an isometric linear bijection $M(\Phi) : M(A) \to M(B)$, which is strictly bicontinuous. Conversely, every strictly bicontinuous, isometrical linear bijection between $M(A)$ and $M(B)$ is of the above form. Moreover, if A and B are both separable or both* (*topologically*) *simple then the strict bicontinuity of any isometric linear bijection between $M(A)$ and $M(B)$ is automatical.*

Proof. Let $\Phi : A \to B$ be an isometric linear bijection, and $\Phi = u\pi$ its Kadison decomposition. Proposition 2.1 entails that, letting $M(\Phi) = uM(\pi), M(\Phi)$ is a strictly bicontinuous, isometrical linear map of $M(A)$ onto $M(B)$. If $\Psi : M(A) \to M(B)$ is any isometric linear bijection extending Φ and $\Psi = v\rho$ is its Kadison decomposition, then $\Psi(A) = \Phi(A) = B$ implies $\rho(A) = v^*\Psi(A) = B$, so $\rho|A : A \to B$ is a Jordan $*$-isomorphism. The uniqueness of the Kadison decomposition of $\Phi = \Psi|A$ yields

$$\pi = \rho|A, \quad u = v.$$

But, again by Proposition 2.1, we have then $M(\pi) = \rho$ and it follows

$$M(\Phi) = uM(\pi) = v\rho = \Psi.$$

The second and the third statement of the theorem follow by taking the Kadison decomposition and applying Proposition 2.2 to the Jordan $*$-isomorphism in the decomposition. □

We notice that every isometric linear bijection $\Phi : A \to B$ between C^*-algebras induces also an isometric linear bijection $C(\Phi) : C(A) \to C(B)$ between the corresponding corona algebras $C(A) = M(A)/A$ and $C(B) = M(B)/B : C(\Phi)$ carries the canonical image of $x \in M(A)$ in $C(A)$ in the canoncal image of $M(\Phi)(x)$ in $C(B)$.

3. Groups of linear isometries on multiplier algebras.

After having established in the preceeding section that surjective linear isometries between C^*-algebras and strictly bicontinuous surjective linear isometries between multiplier C^*-algebras are in one-to-one correspondence, let us now investigate the interplay of the continuity and analyticity properties of groups of such maps. We also investigate the structure of strongly continuous representations of connected groups by linear isometries on C^*-algebras.

Let A, B be C^*-algebras, and $\pi : A \to B$ a Jordan $*$-homomorphism. Then the following variant of (2.1) holds:

$$\begin{cases} \|\pi(x)\pi(y)\| \le (\|xy\|^2 + \|yx\|^2)^{1/2} \le \|xy\| + \|yx\| \\ \text{for all } x, y \in A, \ \text{one of which is self-adjoint.} \end{cases} \tag{3.1}$$

Indeed, let us assume, for example, that $y \in A_h$. By (2.2) we have $\pi(x)^*\pi(x) \le \pi(x^*x + xx^*)$, so

$$\begin{aligned} \|\pi(x)\pi(y)\|^2 &= \|\pi(y)\pi(x)^*\pi(x)\pi(y)\| \\ &\le \|\pi(y)\pi(x^*x + xx^*)\pi(y)\| \\ &= \|\pi(y(x^*x + xx^*)y)\| \\ &\le \|yx^*xy + yxx^*y\| \end{aligned}$$

$$\leq \|xy\|^2 + \|yx\|^2.$$

First we prove:

Lemma 3.1 (on the continuity of the Kadison decomposition)**.** *Let be Ω a topological space, A, B C^*-algebras, and $\Phi_t : A \to B, t \in \Omega$, surjective linear isometries such that $\Omega \ni t \mapsto \Phi_t(a)$ is norm continuous for every $a \in A$. Let us denote by $\Phi_t = u_t\pi_t$ the Kadison decomposition of $\Phi_t, t \in \Omega$. Then*

$$(1) \qquad \begin{cases} \Omega \ni t \mapsto \pi_t(a) \text{ is norm continuous for every } a \in A \\ \Omega \ni t \mapsto \pi_t^{-1}(b) \text{ is norm continuous for every } b \in B; \end{cases}$$

$$(2) \qquad \Omega \ni t \mapsto M(\pi_t)(x) \text{ is strictly continuous for every } x \in M(A)$$

$$(3) \qquad \Omega \ni t \mapsto u_t \text{ is strictly continuous.}$$

Proof. The first statement in (1) follows by noticing that, for every $0 \leq a \in A$,

$$\pi_t(a) = \pi_t(a^{1/2})^2 = \pi_t(a^{1/2})u_t^*u_t\pi_t(a^{1/2}) = \Phi_t(a^{1/2})^*\Phi_t(a^{1/2}).$$

The second statement is implied by the first one: For every $b \in B$,

$$\|\pi_t^{-1}(b) - \pi_s^{-1}(b)\| = \|\pi_s(\pi_t^{-1}(b)) - b\| \overset{s \to t}{\to} \|\pi_t(\pi_t^{-1}(b)) - b\| = 0.$$

For (2) let $x^* = x \in M(A), t_\iota \to t$ in $\Omega, b^* = b \in B$ and $\varepsilon > 0$ be arbitrary. Let us also consider an (increasing, positive) approximate unit $(a_k)_k$ for A. Then, for some k,

$$\|(1_{A^{**}} - a_k)x\pi_t^{-1}(b)\|\|\pi_t^{-1}(b)(1_{A^{**}} - a_k)\| \leq \frac{\varepsilon}{8\|x\|}.$$

By the just proved (1) there exists ι_ε such that, for every $\iota \geq \iota_\varepsilon$,

$$\|\pi_{t_\iota}^{-1}(b) - \pi_t^{-1}(b)\| \leq \frac{\varepsilon}{8\|x\|}, \quad \|\pi_{t_\iota}(a_kx) - \pi_t(a_kx)\| \leq \frac{\varepsilon}{4\|b\|}.$$

Then we have for every $\iota \geq \iota_\varepsilon$

$$\begin{aligned}
&\|(M(\pi_{t_\iota})(x) - M(\pi_t)(x))b\| \\
&\leq \|M(\pi_{t_\iota})(x - a_kx)b\| + \|(\pi_{t_\iota}(a_kx) - \pi_t(a_kx))b\| \\
&\quad + \|M(\pi_t)(a_kx - x)b\| \\
&= \|M(\pi_{t_\iota})(x - a_kx) \cdot M(\pi_{t_\iota})(\pi_{t_\iota}^{-1}(b))\| \\
&\quad + \|M(\pi_t)(x - a_kx) \cdot M(\pi_t)(\pi_t^{-1}(b))\| \\
&\quad + \|(\pi_{t_\iota}(a_kx) - \pi_t(a_kx))b\| \text{ by } (3.1) \\
&\leq \|(x - a_kx)\pi_{t_\iota}^{-1}(b)\| + \|\pi_{t_\iota}^{-1}(b)(x - a_kx)\| \\
&\quad + \|(x - a_kx)\pi_t^{-1}(b)\| + \|\pi_t^{-1}(b)(x - a_kx)\| \\
&\quad + \|\pi_{t_\iota}(a_kx) - \pi_t(a_kx))b\|
\end{aligned}$$

$$\begin{aligned}&\le \|(x-a_kx)(\pi_{t_\iota}^{-1}(b)-\pi_t^{-1}(b))\| + \|(\pi_{t_\iota}^{-1}(b)-\pi_t^{-1}(b))(x-a_kx)\|\\&\quad + 2\|(1_{A^{**}}-a_k)x\pi_t^{-1}(b)\| + 2\|\pi_t^{-1}(b)(1_{A^{**}}-a_k)x\|\\&\quad + \|(\pi_{t_\iota}(a_kx)-\pi_t(a_kx))b\| \text{ and } \iota_\varepsilon\\&\le \|x\|\frac{\varepsilon}{8\|x\|}+\frac{\varepsilon}{8\|x\|}\|x\|+2\frac{\varepsilon}{8}+2\frac{\varepsilon}{8\|x\|}\|x\|+\frac{\varepsilon}{4\|b\|}\|b\|=\varepsilon.\end{aligned}$$

For (3) let $b \in B$ be arbitrary. Since

$$\begin{aligned}&\|u_tb-u_sb\|\\&= \|u_t\pi_t(\pi_t^{-1}(b))-u_s\pi_s(\pi_s^{-1}(b))\|\\&= \|\Phi_t(\pi_t^{-1}(b))-\Phi_s(\pi_s^{-1}(b))\|\\&\le \|\Phi_t(\pi_t^{-1}(b))-\Phi_s(\pi_t^{-1}(b))\| + \|\pi_t^{-1}(b)-\pi_s^{-1}(b)\|,\end{aligned}$$

by the continuity assumption on $t \mapsto \Phi_t$ and by (1), $\|u_tb-u_sb\| \overset{s\to t}{\to} 0$. Applying the above to b replaced by $u_t^*b^*$, it follows also that

$$\|bu_t-bu_s\| = \|u_t^*b^*-u_s^*b^*\| = \|u_su_t^*b^*-b^*\|$$

converges to $\|u_tu_t^*b^*-b^*\| = 0$ for $s \to t$. □

It follows in a straightforward way the following:

Theorem 3.2 (on extension of groups of linear isometries)**.** *Let A be a C^*-algebra, and G a topological group. For every strongly continuous group $(\Phi_t)_{t\in G}$ of linear isometries on A, the group $(M(\Phi_t))_{t\in G}$ is strictly continuous. If G is locally compact, then also conversely, every $\sigma(M(A),A^*)$-continuous group $(\tilde\Phi_t)_{t\in G}$ of $\sigma(M(A),A^*)$-continuous* (*automatic, if A separable or simple!*) *linear isometries on $M(A)$ leaves A invariant and induces on A a strongly continuous group $(\tilde\Phi|A)_{t\in G}$ of linear isometries. In particular, $(\tilde\Phi_t = M(\tilde\Phi|A))_{t\in G}$ is strictly continuous.*

Proof. Denoting by $\Phi_t = u_t\pi_t$ the Kadison decomposition of Φ_t, the above lemma yields the strict continuity of

$$G \ni t \mapsto M(\Phi_t)(x) = u_t \cdot M(\pi_t)(x)$$

for all $x \in M(A)$.

The converse statement follows from the theorem in the preceeding section and from the well known fact, according to which all $\sigma(A,A^*)$-continuous locally compact groups of bounded linear operators on A are strongly continuous. □

For $(\Phi_t)_{t\in G}$ as above it is convenient to denote the extension $(M(\Phi_t))_{t\in G}$ also by $(M(\Phi)_t)_{t\in G}$, underlining in this way that not only the individual Φ_t's are extended by strict continuity, but the whole strongly continuous group Φ is extended to a strictly continuous group $M(\Phi)$.

Next we look for connection between the analytic generator Φ_{-i} of a strongly continuous one-parameter group $(\Phi_t)_{t\in\mathbb{R}}$ of linear isometries on a C^*-algebra and the analytic generator $M(\Phi)_{-i}$ of the strictly continuous one-parameter group $(M(\Phi)_t)_{t\in\mathbb{R}}$ (compare with [**Kus**], Th. 2.41):

Theorem 3.3 (on extension of analytic generators)**.** *Let A be a C^*-algebra, $(\Phi_t)_{t\in\mathbb{R}}$ a strongly continuous one-parameter group of linear isometries on A, and $z\in\mathbb{C}$. Then the graph of $M(\Phi)_z$ is the closure of the graph of Φ_z with respect to the product of the strict topologies on $M(A)\times M(A)$.*

Proof. First of all, the graph of $M(\Phi)_z$ is closed with respect to the product of the strict topologies on $M(A)\times M(A)$ according to the corollary in the first section. Since the strict dual of $M(A)$ is A^*, it is enough to show that the graph of Φ_z is $\sigma(M(A),A^*)\times\sigma(M(A),A^*)$-dense in the graph of $M(\Phi)_z$.

By Theorem 1.1 both $\sigma(M(A),A^*)$ and $\sigma(A^*,M(A))$ have the Krein property, so the formula

$$\langle T_n(x),\varphi\rangle=\sqrt{\frac{n}{\pi}}\int_{-\infty}^{+\infty}e^{-nt^2}\langle M(\Phi)_t(x),\varphi\rangle dt,$$

$$\langle T_{n,z}(x),\varphi\rangle=\sqrt{\frac{n}{\pi}}\int_{-\infty}^{+\infty}e^{-n(t-z)^2}\langle M(\Phi)_t(x),\varphi\rangle dt,\ x\in M(A),\ \varphi\in A^*$$

define $\sigma(M(A),A^*)$-continuous linear maps

$$T_n,\ T_{n,z}:M(A)\to M(A)$$

(see e.g., [**C-Z**], Prop. 1.4). In particular,

$$\Gamma_n=\{(T_n(a),T_{n,z}(a));a\in A\}\text{ is }\sigma(M(A),A^*)\times\sigma(M(A),A^*)\text{-dense in }\{(T_n(x),T_{n,z}(x));\ x\in M(A)\}.$$

According to [**C-Z**], Cor 2.5 and the proof of Lemma 2.2,

$$T_nM(\Phi)_z\subset M(\Phi)_zT_n=T_{n,z},\quad n\ge 1,$$
$$\sigma(M(A),A^*)-\lim_{n\to\infty}T_n(x)=x,\quad x\in M(A).$$

It follows that

$$\bigcup_{n\ge1}\{(T_n(x),T_{n,z}(x));x\in M(A)\}=\{(T_n(x),M(\Phi)_zT_n(x));x\in M(A)\}$$
$$\supset\bigcup_{n\ge1}\{(T_n(x),T_nM(\Phi)_z(x));\ x\in\mathcal{D}_{M(\Phi)_z}\},$$

hence also

$$\bigcup_{n\ge1}\Gamma_n$$

is $\sigma(M(A),A^*)\times\sigma(M(A),A^*)$-dense in the graph of $M(\Phi)_z$.

But $\bigcup\limits_{n\geq 1}\Gamma_n$ is contained in the graph of Φ_z. Indeed, since also $\sigma(A,A^*)$ and $\sigma(A^*,A)$ have the Krein property, applying the above quoted results from [**C-Z**] to $(\Phi_t)_{t\in\mathbb{R}}$ instead of $(M(\Phi)_t)_{t\in\mathbb{R}}$, we get

$$T_nA\subset A,\ T_{n,z}A\subset A \text{ and } \Phi_zT_n|A=T_{n,z}|A,\quad n\geq 1.$$

Therefore

$$(T_n(a),T_{n,z}(a))=(T_n(a),\Phi_zT_n(a))\in \text{ graph of } \Phi_z \text{ for all } n\geq 1 \text{ and } a\in A.$$

□

Now let A,B be C^*-algebras, and $\pi : A\to B$ a Jordan $*$-isomorphism. According to the properties of Jordan $*$-homomorphisms, listed at the beginning of the second section,

$$\pi^{**}(xz)=\pi^{**}(x)\pi^{**}(z)$$

for every $x\in A^{**}$ and z in the centre $Z(A^{**})$ of A^{**}. Thus, denoting by $p_0(A)$ the greatest projection in $Z(A^{**})$ with $p_0(A)A^{**}\subset Z(A^{**})$, and by $p_0(B)$ the similar projection in $Z(B^{**})$, we have

$$\pi^{**}(p_0(A))=p_0(B).$$

By a well known result due to N. Jacobson, C. Rickart and R.V. Kadison ([**Kad1**], Th. 10, see also [**Kad2**], [**Stm**], [**Th**]), there exists a unique projection $p=p(\pi)\in Z(B^{**})$ such that

$$\text{(3.2)}\quad \begin{cases} p\leq 1_{B^{**}}-p_0(B),\\ (\pi(a_1a_2)-\pi(a_1)\pi(a_2))p=0,\quad a_1,a_2\in A,\\ (\pi(a_1a_2)-\pi(a_2)\pi(a_1))(1_{B^{**}}-p_0(B)-p)=0,\quad a_1,a_2\in A.\end{cases}$$

We call $p(\pi)$ the strict homomorphic carrier of π (compare with [**Kad2**], Remark (2.7), where the maximal homomorphic carrier $p_0(B)+p(\pi)$ is considered).

We notice that actually $p(\pi)$ is the unique element $0\leq p\in Z(B^{**})$ satisfying (3.2). Indeed, for any such element p, denoting by p_1 and p_2 the supports of p respectively $1_{B^{**}}-p_0(B)-p$, we have for all $a_1,a_2\in A$

$$(\pi(a_1a_2)-\pi(a_1)\pi(a_2))p_1=0,$$
$$(\pi(a_1a_2)-\pi(a_2)\pi(a_1))p_2=0,$$

hence

$$(\pi(a_1)\pi(a_2))-\pi(a_2)\pi(a_1))p_1p_2=0.$$

It follows successively

$$p_1p_2 \le p_0(B),$$
$$p_1p_2 = p_1p_2(1_{B^{**}} - p_0(B)) = 0,$$
$$p_1 - p = p_1(1_{B^{**}} - p_0(B) - p) = p_1p_2(1_{B^{**}} - p_0(B) - p) = 0,$$
$$p = p_1 \text{ is a projection.}$$

The next result is covered by [**Kad2**], Lemmas (4.8) and (4.10), but we give a proof for the convenience of the reader:

Lemma 3.4 (on the continuous dependence of the homomorphic carrier)**.** *Let be Ω a topological space, A, B C^*-algebras, and $\pi_t : A \to B, t \in \Omega$, Jordan $*$-isomorphisms such that*

$$\Omega \ni t \mapsto \pi_t(a) \in B^{**} \text{ is weak*-continuous for every } a \in A.$$

Then

$$\Omega \ni t \mapsto \pi_t(a) \in B^{**} \text{ is s*-continuous for every } a \in A,$$
$$\Omega \ni t \mapsto p(\pi_t) \in Z(B^{**}) \text{ is s-continuous.}$$

Proof. Clearly, it is enough to prove the s*-continuity of $\Omega \ni t \mapsto \pi_t(a) \in B^{**}$ for all self-adjoint $a \in A$. But in this case

$$(\pi_t(a) - \pi_s(a))^2 = \pi_t(a^2) + \pi_s(a^2) - \pi_t(a)\pi_s(a) - \pi_s(a)\pi_t(a)$$

is weak*-convergent to

$$2\pi_t(a^2) - 2\pi_t(a)\pi_t(a) = 0$$

for $s \to t$.

Let be $t_\iota \to t$ in Ω. If p is any weak*-limit point of $(p(\pi_{t_\iota}))_\iota$ then

$$p(\pi_{t_\iota}) \le 1_{B^{**}} - p_0(B),$$
$$(\pi_{t_\iota}(a_1a_2) - \pi_{t_\iota}(a_1)\pi_{t_\iota}(a_2))p(\pi_{t_\iota}) = 0$$
$$(\pi_{t_\iota}(a_1a_2) - \pi_{t_\iota}(a_2)\pi_{t_\iota}(a_1))(1_{B^{**}} - p_0(B) - p(\pi_{t_\iota})) = 0,$$

true for all ι and $a_1, a_2 \in A$, together with

$$\pi_{t_\iota}(a) \xrightarrow{s^*} \pi_t(a), \quad a \in A,$$

implies

$$p \le 1_{B^{**}} - p_0(B),$$
$$(\pi_t(a_1a_2) - \pi_t(a_1)\pi_t(a_2))p = 0$$
$$(\pi_t(a_1a_2) - \pi_t(a_2)\pi_t(a_1))(1_{B^{**}} - p_0(B) - p) = 0$$

for all $a_1, a_2 \in A$. By the remark before the statement of the lemma it follows that $p = p(\pi_t)$. Thus $p(\pi_{t_\iota}) \to p(\pi_t)$ in the weak* topology.

But on the projections of B^{**} the weak* topology coincides with the s-topology: For $e, f \in B^{**}$ projections

$$(e - f)^2 = e - f - f(e - f) - (e - f)f \xrightarrow{\text{weak}^*} 0$$

whenever $e \overset{\text{weak}^*}{\longrightarrow} f$. Consequently

$$p(\pi_{t_\iota}) \to p(\pi_t) \text{ in the } s\text{-topology.}$$

□

The above lemmas on the continuity of the Kadison decomposition respectively on the continuous dependence of the homomorphic carrier allow us to prove the following:

Theorem 3.5 (on the structure of strongly continuous connected groups of linear isometries). *Let A be a C^*-algebra, G a connected topological group, and $(\Phi_t)_{t\in G}$ a strongly continuous group of linear isometries on A. Let*

$$\Phi_t = u_t\pi_t, \quad t \in G$$

denote the corresponding Kadison decompositions. Then $(\pi_t)_{t\in G}$ is a strongly continuous group of $*$-automorphisms of A, while $G \ni t \mapsto u_t \in M(A)$ is a strictly continuous unitary $M(\pi)$-cocycle:

$$u_{ts} = u_t \cdot M(\pi_t)(u_s), \quad t, s \in G.$$

Proof. We recall that the strict continuity of $G \ni t \mapsto u_t$ is exactly statement (3) in Lemma 3.1. Also the $M(\pi)$-cocycle property follows easily:

$$u_{ts} = M(\Phi_{ts})(1_{A^{**}}) = M(\Phi_t)(M(\Phi_s)(1_{A^{**}})) = u_t M(\pi_t)(u_s).$$

Using it, we get successively for every $a \in A$

$$u_{ts}\pi_{ts}(a) = \Phi_{ts}(a) = \Phi_t(\Phi_s(a)) = u_t\pi_t(u_s\pi_s(a)),$$

$$\pi_{ts}(a) = u_{ts}^* u_t\pi_t(u_s\pi_s(a)) = M(\pi_t)(u_s^*) \cdot \pi_t(u_s\pi_s(a)).$$

Consequently,

$$\pi_t \text{ multiplicative } \Rightarrow \pi_{ts} = \pi_t \cdot \pi_s \text{ for all } s \in G.$$

According to this and to statement (1) in Lemma 3.1, it remains only to prove the multiplicativity of every π_t.

For we shall use the arguments from the proof of [**Kad2**], Th. 3.4:

Let e denote the neutral element of G. By the two continuity lemmas in this section, the map

$$G \ni t \mapsto p_t = p_0(A) + p(\pi_t) \in Z(A^{**})$$

is s-continuous. Since $\pi_e = id_A$ is multiplicative, we have $p_e = 1_{A^{**}}$.

Now every pure state φ of A is multiplicative on $Z(A^{**})$, so

$$\langle p, \varphi\rangle = 1 \text{ or } 0 \text{ for all projections } p \in Z(A^{**}).$$

G being connected, the continuous function

$$G \ni t \mapsto \langle p_t, \varphi\rangle \in \{1.0\}$$

is constant, therefore

$$\langle p_t, \varphi \rangle = \langle p_e, \varphi \rangle = 1 \text{ for every } t \in G.$$

We conclude that, for every $t \in G$, every $a_1, a_2 \in A$ and every pure state φ of A,

$$\begin{aligned} &\langle |\pi_t(a_1 a_2) - \pi_t(a_1)\pi_t(a_2)|^2, \varphi \rangle \\ &= \langle (1_{A^{**}} - p_t)|\pi_t(a_1 a_2) - \pi_t(a_1)\pi_t(a_2)|^2 (1_{A^{**}} - p_t), \varphi \rangle \\ &\leq 4\|a_1\|^2 \|a_2\|^2 \langle (1_{A^{**}} - p_t), \varphi \rangle = 0, \end{aligned}$$

hence

$$\pi_t(a_1 a_2) - \pi_t(a_1)\pi_t(a_2) = 0.$$

□

Let A be a C^*-algebra, and G a topological group. A strongly continuous group $(\Phi_t)_{t\in G}$ of linear isometries on A is called in [**Kus**], Section 2 semi-multiplicative whenever there are strongly continuous groups $(\Phi_t^{(j)})_{t\in G}, j = 1, 2$, of linear isometries on A such that

$$\Phi_t(ab) = \Phi_t(a)\Phi_t^{(1)}(b) = \Phi_t^{(2)}(a)\Phi_t(b), \ t \in G, \ a, b \in A. \tag{3.3}$$

But it is easy to see that, with $\Phi_t = u_t \pi_t$ the Kadison decomposition of Φ_t, for any maps $\Phi_t^{(j)} : A \to A, \ t \in G, \ j = 1, 2$, satisfying (3.3), we have necessarily

$$\Phi_t^{(1)} = \pi_t \text{ and } \Phi_t^{(2)} = u_t \pi_t u_t^*, \quad t \in G.$$

Thus $(\Phi_t)_{t\in G}$ is semi-multiplicative if and only if all π_t's are $*$-automorphisms and the above theorem claims essentially that, assuming G connected, every strongly continuous representation of G by linear isometries on A is automatically semi-multiplicative.

We notice that our Theorem 3.3 on the extension of analytic generators was proved in [**Kus**], Th. 2.41 under the additional assumption of semi-multiplicativity of the one-parameter group, which turns out to be automatic by the above remark. However, our approach is more natural for the prevailingly linear character of the statement of the theorem. Indeed, our proof works for any (even not bounded) strongly continuous one-parameter group of bounded linear operators on A, which extends by strict continuity to a $\sigma(A^*, M(A))$-continuous one-parameter group of bounded linear operators on $M(A)$.

Let us consider also the W^*-algebra counterpart of the above theorem (compare with [**Kad2**], Th. 3.4):

Theorem 3.6 (on the structure of weak* continuous connected groups of linear isometries)**.** *Let M be a W^*-algebra, whose centre is atomic, G a* *connected topological group, and $(\Phi_t)_{t\in G}$ a weak* continuous group of linear isometries on M. Let*

$$\Phi_t = u_t\pi_t, \quad t \in G$$

denote the corresponding Kadison decompositions. Then $(\pi_t)_{t\in G}$ is an s^*-continuous group of $*$-automorphisms of M, while $G \ni t \mapsto u_t \in M$ is an s^*-continuous unitary π-cocycle:

$$u_{ts} = u_t\pi_t(u_s), \quad t, s \in G.$$

Proof. First of all, $G \ni t \mapsto u_t = \Phi_t(1_M)$ is weak*-continuous, hence also s^*-continuous. Indeed, on the unitaries of M the weak* topology coincides with the s^*-topology: For $u, v \in M$ unitaries

$$(u-v)^*(u-v) + (u-v)(u-v)^* = 4 \cdot 1_M - u^*v - v^*u - uv^* - vu^* \overset{\text{weak}^*}{\longrightarrow} 0$$

whenever $v \overset{\text{weak}^*}{\longrightarrow} u$.

Next, for every $x \in M$,

$$G \ni t \mapsto \pi_t(x) = u_t^*\Phi_t(x) \in M$$

is weak*-continuous according to the weak*-continuity of $(\Phi_t)_{t\in G}$ and the s-continuity of $G \ni t \mapsto u_t$. It follows by straightforward arguments that the maps

$$G \ni t \mapsto \pi_t(x) \in M, \quad x \in M$$

are even s^*-continuous (see [**Kad2**], Lemma 4.10 or the first paragraph of the proof of Lemma 3.4).

Similarly, as in the proof of the preceeding theorem, we get also

$$u_{ts} = u_t\pi_t(u_s), \quad t, s \in G,$$

and then

$$\pi_t \text{ multiplicative } \Rightarrow \pi_{ts} = \pi_t \cdot \pi_s \text{ for all } s \in G.$$

Thus it remains only to prove the multiplicativity of every π_t.

Let p_0 denote the greatest projection in the centre $Z(M)$ of M satisfying $p_0 M \subset Z(M)$. Then

$$\pi_t(p_0) = p_0 \quad \text{for all } t \in G.$$

According to [**Kad1**], Th. 10, for every $t \in G$ there exists a unique projection $p(\pi_t) \in Z(M)$ such that

$$p(\pi_t) \le 1_M - p_0,$$
$$(\pi_t(x_1x_2) - \pi_t(x_1)\pi_t(x_2))p(\pi_t) = 0, \quad x_1, x_2 \in M,$$
$$(\pi_t(x_1x_2) - \pi_t(x_2)\pi_t(x_1))(1_M - p_0 - p(\pi_t)) = 0, \quad x_1, x_2 \in M.$$

Moreover, arguing similarly as in the proof of Lemma 3.4, it is easy to verify that the map

$$G \ni t \mapsto p(\pi_t) \in Z(M),$$

hence also

$$G \ni t \mapsto p_t = p_0 + p(\pi_t) \in Z(M)$$

is s-continuous. We notice that $\pi_e = id_M$ implies $p_e = 1_M$, where e stands for the neutral element of G.

Now, for every normal state φ on M, whose central support is a minimal projection of $Z(M)$, the continuous function

$$G \ni t \mapsto \langle p_t, \varphi \rangle$$

takes values in $\{1, 0\}$. G being connected, it follows that

$$\langle p_t, \varphi \rangle = \langle p_e, \varphi \rangle = 1 \text{ for all } t \in G.$$

Consequently, for every $t \in G$ and $x_1, x_2 \in M$,

$$\begin{aligned} &\langle |\pi_t(x_1 x_2) - \pi_t(x_1)\pi_t(x_2)|^2, \varphi \rangle \\ &= \langle (1_M - p_t)|\pi_t(x_1 x_2) - \pi_t(x_1)\pi_t(x_2)|^2 (1_M - p_t), \varphi \rangle \\ &\leq 4\|x_1\|^2 \|x_2\|^2 \langle 1_M - p_t, \varphi \rangle = 0. \end{aligned}$$

Since $Z(M)$ is atomic, the normal states on M with central supports minimal in $Z(M)$ separate the points of M and we conclude that

$$\pi_t(x_1 x_2) - \pi_t(x_1)\pi_t(x_2) = 0 \text{ for all } t \in G \text{ and } x_1, x_2 \in M.$$

□

In particular, if M is a factor, G is a connected topological group and $(\Phi_t)_{t\in G}$ is a weak*-continuous group of linear isometries on M, then

$$\Phi_t = u_t \pi_t, \quad t \in G$$

with $(\pi_t)_{t\in G}$ an s^*-continuous group of $*$-automorphisms of M and $G \ni t \mapsto u_t \in M$ an s^*-continuous unitary π-cocycle.

However, the above theorem does not hold without the assumption on the centre of M, even we assume M of type I_2, $G = \mathbb{R}$ and all Φ_t Jordan $*$-automorphisms. Indeed, denoting by M the W^*-algebra $L^\infty(\mathbb{R};\ \mathrm{Mat}_2(\mathbb{C}))$ of all essentially bounded (equivalence classes of) 2×2 matrix valued measurable functions on $\mathbb{R}$, choosing some $*$-anti-automorphism σ of $\mathrm{Mat}_2(\mathbb{C})$ (e.g., the transpose map) and putting for $F \in L^\infty(\mathbb{R};\ \mathrm{Mat}_2(\mathbb{C}))$

$$\Phi_t(F)(s) = \begin{cases} \sigma(F(s-t)) \text{ if } 0 < s < t \\ F(s-t) \text{ otherwise,} \end{cases} \qquad t \geq 0,\ s \in \mathbb{R},$$

$$\Phi_t(F)(s) = \begin{cases} \sigma^{-1}(F(s-t)) \text{ if } t < s < 0 \\ F(s-t) \text{ otherwise,} \end{cases} \qquad t \leq 0,\ s \in \mathbb{R},$$

$(\Phi_t)_{t\in\mathbb{R}}$ is a weak*-continuous one-parameter group of Jordan $*$-automorphisms of M, but no Φ_t with $t \neq 0$ is multiplicative.

We finish the section (and the paper) with a continuity criterion for groups of linear isometries on multiplier algebras of separable C^*-algebras and with a result concerning continuity on corona algebras of separable C^*-algebras.

They will follow from the next two lemmas of general character:

Lemma 3.7 ([**Z2**]). *Let $(X, \mathcal{F})$ be a dual pair of Banach spaces, G a locally compact group, and $(U_t)_{t\in G}$ a group of $\sigma(X,\mathcal{F})$-continuous linear operators on X. If there exists a compact set $K \subset G$ of nonzero Haar measure, such that all functions*

$$K \ni t \mapsto \langle U_t(x), \varphi\rangle, \quad x \in X, \varphi \in \mathcal{F}$$

are continuous, then $(U_t)_{t\in G}$ is $\sigma(X,\mathcal{F})$-continuous.

Proof. Let m denote a left Haar measure on G. By [**He-R**], Cor. 20.17 there exists $t_0 \in G$ such that $m(K \cap t_0K^{-1}) > 0$. Then the support of the restriction of m to $K \cap t_0K^{-1}$ is a compact subset K_0 of $K \cap t_0K^{-1}$ such that

$$m(K_0) = m(K \cap t_0K^{-1}) > 0, \text{ in particular, } K_0 \neq \emptyset,$$
$$V \subset G \text{ open, } V \cap K_0 \neq \emptyset \Rightarrow m(V \cap K_0) > 0.$$

Now let $x \in X, \varphi \in \mathcal{F}$ and $\varepsilon > 0$ be arbitrary. Defining $F : K_0 \times K_0 \to \mathbb{C}$ by

$$F(t,s) = \langle U_{ts^{-1}}(x), \varphi\rangle \quad t, s \in K_0,$$

F is separately continuous. Indeed, by the continuity assumption on U,

$$K_0 \subset K \ni t \mapsto \langle U_{ts^{-1}}(x), \varphi\rangle = \langle U_t(U_{s^{-1}}(x)), \varphi\rangle$$

is continuous for every $s \in G$, and

$$K_0 \subset t_0K^{-1} \ni s \mapsto \langle U_{ts^{-1}}(x), \varphi\rangle = \langle U_{s^{-1}t_0}(U_{t_0^{-1}}(x)), U_t^{\mathcal{F}}(\varphi)\rangle$$

is continuous for every $t \in G$. According to a theorem of I. Namioka ([**N**], Th. 1.2, see also [**Chr**], Th. 1) it follows the existence of a dense G_δ set D in K_0 such that F is jointly continuous in every point of $D \times K_0$. In particular, choosing some $t_1 \in D$, F is continuous in (t_1, t_1), so there exists some open set $t_1 \in V_1 \subset G$ such that

$$(t,s) \in (V_1 \cap K_0) \times (V_1 \cap K_0) \Rightarrow |F(t,s) - F(t_1,t_1)| < \varepsilon,$$

that is

$$t \in (V_1 \cap K_0) \cdot (V_1 \cap K_0)^{-1} \Rightarrow |\langle U_t(x) - x, \varphi\rangle| < \varepsilon.$$

But $V_1 \cap K_0 \ni t$, being not empty, we have $m(V_1 \cap K_0) > 0$ and [**He-R**], Cor. 20.17 implies that $(V_1 \cap K_0) \cdot (V_1 \cap K_0)^{-1}$ is a neighborhood of the neutral element of G.

We conclude that the mappings

$$G \ni t \to U_t(x) \in X, \quad x \in X$$

are $\sigma(X,\mathcal{F})$-continuous in the neutral element of G, hence everywhere. $\square$

Lemma 3.8 ([**Z2**], compare with [**Hi-P**], Th. 3.5.3 and Th. 10.2.3)**.** *Let X be a separable Banach space, $\mathcal{F}$ a linear subspace of the dual X^*, satisfying*

$$\|x\| = \sup\{|\langle x,\varphi\rangle|; \varphi\in\mathcal{F}, \|\varphi\|\le 1\}, \quad x\in X,$$

G a locally compact group with left Haar measure m, and $(U_t)_{t\in G}$ a group of bounded linear operators on X. If there exists a Haar-measurable $B\subset G$ with $0<m(B)<+\infty$, such that all functions

$$B\ni t\mapsto \langle U_t(x),\varphi\rangle, \quad x\in X,\ \varphi\in\mathcal{F}$$

are Haar-measurable, then $(U_t)_{t\in G}$ is strongly continuous.

Proof. Let $(x_k)_{k\ge1}$ be a norm dense sequence in $X\backslash\{0\}$. For every $k,n\ge1$ there exists $\varphi_{k,n}\in\mathcal{F}$, $\|\varphi_{k,n}\|\le1$, such that

$$|\langle x_k,\varphi_{k,n}\rangle| > \|x_k\| - \frac{1}{n}.$$

Then

$$\|x\| = \sup\{|\langle x,\varphi_{k,n}\rangle|; k,n\ge1\}, \quad x\in X.$$

Let $x\in X$ and $\delta,\varepsilon>0$ be arbitrary.

By the measurability assumption on U, the functions

$$B\ni t\mapsto \|U_t(x)-x_k\| = \sup\{|\langle U_t(x)-x_k,\varphi_{p,q}\rangle|; p,q\ge1\}, k\ge1$$

are Haar measurable, so all sets

$$S_k = \{t\in B; \|U_t(x)-x_k\|<\varepsilon\}\subset B, \quad k\ge1$$

are also Haar-measurable. Since $(x_k)_{k\ge1}$ is norm dense in X, we have

$$\bigcup_{k\ge1} S_k = B.$$

Thus, putting

$$R_k = S_k\backslash \bigcup_{1\le l<k} S_l, \quad k\ge1,$$

we get a partition $R_1,R_2,\ldots$ of B in Haar-measurable sets such that

$$t\in R_k \Rightarrow \|U_t(x)-x_k\|<\varepsilon, \quad k\ge1.$$

By the countable additivity of m there exists $k_0\ge1$ with

$$m\left(B\backslash\bigcup_{k=1}^{k_0} R_k\right) < \frac{\delta}{3}m(B),$$

and then, by the regularity of m, a compact set $K_0 \subset \bigcup_{k=1}^{k_0} R_k$ with

$$m\left(\bigcup_{k=1}^{k_0} R_k \setminus K_0\right) < \frac{\delta}{3} m(B).$$

It follows that

$$m(B \setminus K_0) < \frac{2\delta}{3} m(B)$$

and the Haar-measurable sets

$$E_k = R_k \cap K_0, \quad 1 \leq k \leq k_0$$

yield a partition of K_0 such that

$$t \in E_k \Rightarrow \|U_t(x) - x_k\| < \varepsilon, \quad 1 \leq k \leq k_0.$$

Let χ_{E_k} denote the characteristic function of E_k. By the Lusin theorem there are compact subsets $K_1, \ldots, K_{k_0}$ of K_0 such that, for every $1 \leq k \leq k_0$,

$$m(K_0 \setminus K_k) < \frac{\delta}{3k_0} m(B),$$

$$\chi_{E_k} | K_k : K_k \to [0,1] \text{ is continuous.}$$

Then

$$K_{x,\delta,\varepsilon} = \bigcap_{k=1}^{k_0} K_k \subset K_0$$

is a compact subset of B with

$$m(B \setminus K_{x,\delta,\varepsilon}) < \delta m(B)$$

and all functions

$$\chi_{E_k} | K_{x,\delta,\varepsilon} : K_{x,\delta,\varepsilon} \to [0,1]$$

are continuous. Therefore

$$K_{x,\delta,\varepsilon} \ni t \mapsto F_{x,\delta,\varepsilon}(t) = \sum_{k=1}^{k_0} \chi_{E_k}(t) x_k \in X$$

is norm continuous. Moreover,

$$\|U_t(x) - F_{x,\delta,\varepsilon}(t)\| < \varepsilon \text{ for all } t \in K_{x,\delta,\varepsilon}.$$

Summing up the aboves: For every $x \in X$ and every $\delta, \varepsilon > 0$ there exists a compact set $K_{x,\delta,\varepsilon} \subset B$ with

$$m(B \setminus K_{x,\delta,\varepsilon}) < \delta m(B)$$

and a norm continuous map $F_{x,\delta,\varepsilon} : K_{x,\delta,\varepsilon} \to X$ with

$$\|U_t(x) - F_{x,\delta,\varepsilon}(t)\| < \varepsilon \text{ for all } t \in K_{x,\delta,\varepsilon}.$$

Then, for every $x \in X$ and $\delta > 0$,

$$K_{x,\delta} = \bigcap_{n=1}^{\infty} K_{x,\delta 2^{-n},n^{-1}} \subset B$$

is a compact set with

$$m(B \backslash K_{x,\delta}) < \delta m(B)$$

and

$$K_{x,\delta} \ni t \to U_t(x) \in X$$

is norm continuous as the uniform limit of the norm continuous maps $F_{x,\delta 2^{-n},n^{-1}} | K_{x,\delta}$, $n \geq 1$.

Denoting now

$$L_1 = \bigcap_{k=1}^{\infty} K_{x_k, \frac{1}{3} 2^{-k}},$$

L_1 is compact subset of B with

$$m(B \backslash L_1) < \frac{1}{3} m(B)$$

and all mappings

$$L_1 \ni t \mapsto U_t(x_k) \in X, \quad k \geq 1$$

are norm continuous.

On the other hand, again by the measurability assumption on U, the function

$$B \ni t \mapsto \|U_t\| = \sup\{\|x_k\|^{-1} \cdot |\langle U_t(x_k), \varphi_{p,q} \rangle|; k, p, q \geq 1\}$$

is Haar-measurable. By the countable additivity of m there exists a sufficiently large $c > 0$ with

$$m(\{t \in B; \|U_t\| \leq c\}) > \frac{2}{3} m(B),$$

and then, by the regularity of m, a compact set

$$L_2 \subset \{t \in B; \|U_t\| \leq c\} \text{ with } m(L_2) > \frac{2}{3} m(B).$$

We conclude that $L = L_1 \cap L_2$ is a compact subset of B such that

$$\begin{aligned} m(L) &= m(L_2) - m(L_2 \backslash L_1) \\ &\geq m(L_2) - m(B \backslash L_1) \\ &> \frac{2}{3} m(B) - \frac{1}{3} m(B) > 0, \end{aligned}$$

$$\|U_t\| \leq c \text{ for all } t \in L$$

and the mappings

$$L \ni t \mapsto U_t(x_k) \in X, \quad k \geq 1$$

are norm continuous. Since $(x_k)_{k\geq 1}$ is dense in X, it follows the norm continuity of all mappings

$$L \ni t \mapsto U_t(x) \in X, \quad x \in X.$$

Now Lemma 3.7 yields the $\sigma(X, X^*)$-continuity, hence the strong continuity of $(U_t)_{t\in G}$. □

By the above result, every representation of a locally compact group by bounded linear operators on a separable Banach space, which has a "minimal regularity property", is automatically strongly continuous.

Theorem 3.9 (on characterizing continuity of linear isometry groups on separable C^*-algebras and their multipliers algebras)**.** *Let be H a complex Hilbert space, $A \subset B(H)$ a non-degenerate, separable C^*-subalgebra, and*

$$M(A) = \{x \in B(H); xa,\ ax \in A \text{ for all } a \in A\}.$$

Then $\Phi \mapsto \Phi|A$ establishes an one-to-one correspondence between the surjective linear isometries Φ on $M(A)$ and those on A.

Let further be G a locally compact group and $(\Phi_t)_{t\in G}$ a group of linear isometries on $M(A)$. If there are

- *a linear subspace $\mathcal{F} \subset B(H)^*$, satisfying $\|a\| = \sup\{|\langle a, \varphi\rangle|; \varphi \in \mathcal{F}, \|\varphi\| \leq 1\}$ for all $a \in A$,*
- *a Haar-measurable $B \subset G$ of nonzero, finite Haar measure,*

such that all functions

$$B \ni t \mapsto \langle \Phi_t(a), \varphi\rangle, \quad a \in A,\ \varphi \in \mathcal{F}$$

are Haar-measurable, then $(\Phi|A)_{t\in G}$ is strongly continuous and $(\Phi_t)_{t\in G}$ is strictly continuous, i.e., the maps

$$G \ni t \mapsto \Phi_t(x)a, \quad G \ni t \mapsto a\Phi_t(x)$$

are norm-continuous for all $x \in M(A)$ and $a \in A$.

Proof. The first statement follows from Theorem 2.3, taking into account the canonical identification of the here defined $M(A)$ with the multiplier algebra defined in the first section (see [**Ped1**], Prop. 3.12.3 or [**WO**], Prop. 2.2.11).

Now let $(\Phi_t)_{t\in G}$ be as in the theorem. Then Lemma 3.8 implies the strong continuity of $(\Phi_t|A)_{t\in G}$, and then the strict continuity of $(\Phi_t)_{t\in G}$ is entailed by Theorem 3.2. □

In general, for $(\Phi_t)_{t\in G}$ strongly continuous group of linear isometries on a separable C^*-algebra A, the group $(C(\Phi_t))_{t\in G}$ is not strongly continuous. However we have strong continuity on separable, invariant C^*-algebras of $C(A)$:

Theorem 3.10 (on continuity properties of linear isometry groups on corona algebras of separable C^*-algebras). *Let be G a locally compact group, $(\Phi_t)_{t\in G}$ a strongly continuous group of linear isometries on a separable C^*-algebra A, and D a separable C^*-subalgebra of $C(A) = M(A)/A$, left invariant by $(C(\Phi_t))_{t\in G}$. Then the group $(C(\Phi_t)|D)_{t\in G}$ of linear isometries on D is strongly continuous.*

Proof. Let us denote by π the quotient map $M(A) \to C(A)$ and let $(x_n)_{n\geq 1}$ be a sequence in $M(A)$ such that $(\pi(x_n))_{n\geq 1}$ is dense in D. Then the C^*-subalgebra $B \subset M(A)$ generated by $A \cup \{x_n; n \geq 1\}$ is separable and it is easy to see that

$$B = \pi^{-1}(D).$$

It follows that B is left invariant by $(M(\Phi_t))_{t\in G}$. Since $(M(\Phi_t)|B)_{t\in G}$ is $\sigma(B, A^*)$-continuous, according to Lemma 3.8 it is strongly continuous and the strong continuity of $(C(\Phi_t)|D)_{t\in G}$ follows. □

Actually we have proved more: If $G, (\Phi_t)_{t\in G}$ and π are as above, then

$$\pi^{-1}\left(\cup\left\{D; \begin{matrix} D \subset C(A) \text{ separable } C^*\text{-subalgebra} \\ C(\Phi_t)D \subset D \text{ for all } t \in G \end{matrix}\right\}\right)$$

is contained in

$$\{x \in M(A);\ G \ni t \mapsto M(\Phi_t)(x) \text{ is norm-continuous}\}.$$

Clearly, the above two C^*-algebras of $M(A)$ are equal whenever G is separable.

References

[Ar] W. Arveson, *On groups of automorphisms of operator algebras*, J. Funct. Analysis, **15** (1974), 217-243.

[B] N. Bourbaki, *Topological Vector Spaces*, Springer-Verlag, 1987.

[Br] L.G. Brown, *Determination of A from M(A) and related matters*, C.R. Math. Rep. Acad. Sci. Canada, **10** (1988), 273-278.

[Buc] R.C. Buck, *Bounded continuous functions on a locally compact space*, Michigan Math. J., **5** (1958), 95-104.

[Bus] R.C. Busby, *Double centralizers and extensions of C^*-algebras*, Trans. Amer. Math. Soc., **132** (1968), 79-99.

[Chr] J.P.R. Christensen, *Joint continuity of separately continuous functions*, Proc. Amer. Math. Soc., **82** (1981), 455-461.

[C-Z] I. Ciorănescu and L. Zsidó, *Analytic generators for one-parameter groups*, Tohoku Math. J., **28** (1976), 327-362.

[Cnw] J.B. Conway, *The strict topology and compactness in the space of measures*, II, Trans. Amer. Math. Soc., **126** (1967), 474-486.

[D-Z] C. D'Antoni and L. Zsidó, *Abelian strict approximation in multiplier C^*-algebras*, preprint, 1998.

[He-R] E. Hewitt and K.A. Ross, *Abstract Harmonic Analysis*, I-II, Springer-Verlag, 1963/1970.

[Hi-P] E. Hille and R.S. Phillips, *Functional Analysis and Semi-Groups*, Amer. Math. Soc. Colloquium Publ., **31** (1981).

[HJ] J. Hoffmann-Jørgensen, *Weak compactness and tightness of subsets of $M(X)$*, Math. Scand., **31** (1972), 127-150.

[Kad1] R.V. Kadison, *Isometries of operator algebras*, Ann. of Math., **54** (1951), 325-338.

[Kad2] ______, *Transformations of states in operator theory and dynamics*, Topology, **3** (1965), 177-198.

[Kus] J. Kustermans, *One-parameter representations on C^*-algebras*, preprint (see funct - an/9707009).

[N] I. Namioka, *Separate continuity and joint continuity*, Pacific J. Math., **51** (1974), 515-531.

[P-S] A.L.T. Paterson and A.M. Sinclair, *Characterization of isometries between C^*-algebras*, J. London Math. Soc., **5**(**2**) (1976), 755-761.

[Ped1] G.K. Pedersen, *C^*-Algebras and their Automorphism Groups*, Academic Press, 1979.

[Ped2] ______, *SAW*-algebras and corona C^*-algebras*, J. Operator Theory, **15** (1986), 15-32.

[Rud] W. Rudin, *Functional Analysis*, McGraw-Hill, 1973.

[Sch] H.H. Schaefer, *Topological Vector Spaces*, Springer-Verlag, 1980.

[Stm] E. Størmer, *On the Jordan structure of C^*-algebras*, Trans. Amer. Math. Soc., **120** (1965), 438-447.

[S-Z] Ş. Strătilă and L. Zsidó, *Operator algebras*, INCREST Prepublication, (1977-1979), 1-511.

[T1] D.C. Taylor, *The strict topology for double centralizer algebras*, Trans. Amer. math. Soc., **150** (1970), 633-643.

[T2] ______, *A general Phillips theorem for C^*-algebras and some applications*, Pacific J. Math., **40** (1970), 477-488.

[Th] K. Thomsen, *Jordan-morphisms in *-algebras*, Proc. Amer. Math. Soc., **86** (1982), 283-286.

[WO] N.E. Wegge-Olsen, *K-theory and C^*-algebras*, Oxford University Press, 1993.

[Z1] L. Zsidó, *Spectral and ergodic properties of the analytic generator*, J. Approximation Theory, **20** (1977), 77-138.

[Z2] ______, *unpublished manuscript*, 1979.

Received September 30, 1998. Both authors were Supported by M.U.R.S.T., C.N.R. and E.U.

Università di Roma "Tor Vergata"
Via della Ricerca Scientifica
00133 Roma
Italia

E-mail address: dantoni@axp.mat.uniroma2.it

Università di Roma "Tor Vergata"
Via della Ricerca Scientifica
00133 Roma
Italia
E-mail address: zsido@axp.mat.uniroma2.it

PACIFIC JOURNAL OF MATHEMATICS
Vol. 193, No. 2, 2000

A GENERALIZATION OF CURVE GENUS FOR AMPLE VECTOR BUNDLES, II

Yoshiaki Fukuma and Hironobu Ishihara

Let X be a compact complex manifold of dimension $n \geq 2$ and $\mathcal{E}$ an ample vector bundle of rank $r < n$ on X. As the continuation of Part I, we further study the properties of $g(X, \mathcal{E})$ that is an invariant for pairs $(X, \mathcal{E})$ and is equal to curve genus when $r = n - 1$. Main results are the classifications of $(X, \mathcal{E})$ with $g(X, \mathcal{E}) = 2$ (resp. 3) when $\mathcal{E}$ has a regular section (resp. $\mathcal{E}$ is ample and spanned) and $1 < r < n - 1$.

Introduction.

The present paper is a continuation of [**I**]. For a pair $(X, \mathcal{E})$ which consists of a compact complex manifold X of dimension $n \geq 2$ and an ample vector bundle $\mathcal{E}$ of rank $r < n$ on X, we defined in [**I**] an invariant $g(X, \mathcal{E})$ by the formula

$$2g(X, \mathcal{E}) - 2 := (K_X + (n - r)c_1(\mathcal{E}))c_1(\mathcal{E})^{n-r-1}c_r(\mathcal{E}).$$

We note that $g(X, \mathcal{E})$ is a nonnegative integer, and $g(X, \mathcal{E})$ is equal to the curve genus of $(X, \mathcal{E})$ when $r = n - 1$. As in the case of curve genus, above $(X, \mathcal{E})$ with $g(X, \mathcal{E}) \leq 1$ have been classified in [**I**]; moreover, it is shown that $g(X, \mathcal{E}) \geq q(X)$ for spanned $\mathcal{E}$ and its equality condition is given in [**I**]. ($q(X)$ is the irregularity of X.)

After we recall some preliminary results in Section 1, we consider the cases $g(X, \mathcal{E}) = 2$ and $g(X, \mathcal{E}) = 3$ when $1 < r < n-1$ in Section 2. Corresponding results for c_1-sectional genus are given in [**Fj2**] and [**BiLL**] respectively. In Section 3 we consider the cases $g(X, \mathcal{E}) = q(X) + 1$ and $g(X, \mathcal{E}) = q(X) + 2$ when $1 < r < n - 1$. Related results for c_1-sectional genus are given in [**R**]. In Section 4 we give another relation between $g(X, \mathcal{E})$ and $q(X)$, namely $g(X, \mathcal{E}) \geq 2q(X) - 1$ for $1 < r < n - 1$. When $r = 1$, this inequality is satisfied except one case. In Section 5 we show that $g(X, \mathcal{E}) \geq g(C)$ when there exists a fibration $f : X \to C$ over a curve. We also give its equality condition. Finally in Appendix we give a classification of (X, L) with $g(X, L) = q(X) + 2$ and $n = 2$ for ample and spanned line bundles L on X.

The authors would like to express their gratitude to Professor Takao Fujita for his valuable comments, especially for informing them of Lemma 2.2.5. They are grateful to the referee for reading the manuscript carefully.

1. Preliminaries.

We use a notation similar to that in [**I**]. For example, we denote by $H(\mathcal{E})$ the tautological line bundle on $\mathbb{P}_X(\mathcal{E})$, the projective space bundle associated to a vector bundle $\mathcal{E}$ on a variety X. We say that a vector bundle $\mathcal{E}$ is spanned if $H(\mathcal{E})$ is spanned. A polarized manifold (X, L) is said to be a scroll over a variety W if $(X, L) \simeq (\mathbb{P}_W(\mathcal{F}), H(\mathcal{F}))$ for some ample vector bundle $\mathcal{F}$ on W. We denote by $\mathbb{F}_e$ the Hirzebruch surfaces $\mathbb{P}_{\mathbb{P}^1}(\mathcal{O}_{\mathbb{P}^1} \oplus \mathcal{O}_{\mathbb{P}^1}(-e))$ $(e > 0)$, by σ a minimal section, and by f a fiber of the ruling $\mathbb{F}_e \to \mathbb{P}^1$. Numerical equivalence is denoted by $\equiv$.

Definition 1.1. Let X be a compact complex manifold of dimension $n \geq 2$ and $\mathcal{E}$ an ample vector bundle of rank $r < n$ on X. We define a rational number $g(X, \mathcal{E})$ for the pair $(X, \mathcal{E})$ by the formula

$$2g(X, \mathcal{E}) - 2 := (K_X + (n-r)c_1(\mathcal{E}))c_1(\mathcal{E})^{n-r-1}c_r(\mathcal{E}).$$

It turns out that $g(X, \mathcal{E})$ is a nonnegative integer (see [**I**]). When $r = 1$ (resp. $r = n-1$), $g(X, \mathcal{E})$ is nothing but the sectional genus (resp. curve genus) of $(X, \mathcal{E})$.

Remark 1.2. Let $(X, \mathcal{E})$ be as above. Suppose that $(X, \mathcal{E})$ satisfies the condition

($*$) There exists a section $s \in H^0(X, \mathcal{E})$ whose zero locus $Z := (s)_0$ is a smooth submanifold of X of the expected dimension $n - r$.

Then we have $g(X, \mathcal{E}) = g(Z, \det \mathcal{E}_Z)$ (see [**I**]). If $\mathcal{E}$ is spanned, then $\mathcal{E}$ satisfies ($*$) by Bertini's theorem.

The following facts are used in the subsequent sections.

Proposition 1.3. *Let X be an n-dimensional compact complex manifold and $\mathcal{E}$ an ample vector bundle of rank $r < n$ on X with the property ($*$) in (1.2). Let $\iota : Z \hookrightarrow X$ be the embedding. Then*

(1) *$H^i(\iota) : H^i(X, \mathbb{Z}) \to H^i(Z, \mathbb{Z})$ is an isomorphism for $i < n - r$.*
(2) *$H^i(\iota)$ is injective and its cokernel is torsion free for $i = n - r$.*
(3) *$\mathrm{Pic}(\iota) : \mathrm{Pic}(X) \to \mathrm{Pic}(Z)$ is an isomorphism for $n - r > 2$.*
(4) *$\mathrm{Pic}(\iota)$ is injective and its cokernel is torsion free for $n - r = 2$.*

Proof. See Theorem 1.3 in [**LM1**] and see also Theorem 1.1 in [**LM2**]. □

Proposition 1.4. *Let X be an n-dimensional compact complex manifold and $\mathcal{E}$ an ample vector bundle of rank $r \geq 2$ on X with the property ($*$).*

If $Z \simeq \mathbb{P}^{n-r}(n - r \geq 1)$, then $(X, \mathcal{E})$ is one of the following:

(P1) $(\mathbb{P}^n, \mathcal{O}_{\mathbb{P}^n}(1)^{\oplus r})$;
(P2) $(\mathbb{P}^n, \mathcal{O}_{\mathbb{P}^n}(2) \oplus \mathcal{O}_{\mathbb{P}^n}(1)^{\oplus(n-2)})$;
(P3) $(\mathbb{Q}^n, \mathcal{O}_{\mathbb{Q}^n}(1)^{\oplus(n-1)})$;
(P4) $X \simeq \mathbb{P}_{\mathbb{P}^1}(\mathcal{F})$ *for some vector bundle* $\mathcal{F}$ *of rank* n *on* $\mathbb{P}^1$ *and* $\mathcal{E} = \oplus_{j=1}^{n-1}(H(\mathcal{F}) + \pi^*\mathcal{O}_{\mathbb{P}^1}(b_j))$*, where* $\pi : X \to \mathbb{P}^1$ *is the bundle projection.*

If $Z \simeq \mathbb{Q}^{n-r}(n - r \geq 2)$*, then* $(X, \mathcal{E})$ *is one of the following:*

(Q1) $(\mathbb{P}^n, \mathcal{O}_{\mathbb{P}^n}(2) \oplus \mathcal{O}_{\mathbb{P}^n}(1)^{\oplus(r-1)})$;
(Q2) $(\mathbb{Q}^n, \mathcal{O}_{\mathbb{Q}^n}(1)^{\oplus r})$;
(Q3) $X \simeq \mathbb{P}_{\mathbb{P}^1}(\mathcal{F})$ *and* $\mathcal{E} = \oplus_{j=1}^{n-2}(H(\mathcal{F}) + \pi^*\mathcal{O}_{\mathbb{P}^1}(b_j))$*, where* $\mathcal{F}$ *is the same as that in* (P4).

Proof. See Theorem A and Theorem B in [**LM1**]. □

Proposition 1.5. *Let* X *be a complex projective manifold of dimension* n *and let* $\mathcal{E}$ *be an ample vector bundle of rank* $n - 2 \geq 2$ *on* X *satisfying* $(*)$.

(1) *If* Z *is a geometrically ruled surface over a smooth curve* B *such that* $Z \neq \mathbb{F}_0, \mathbb{F}_1$*, then* X *is a* $\mathbb{P}^{n-1}$*-bundle over* B *and* $\mathcal{E}_F = \mathcal{O}_{\mathbb{P}^{n-1}}(1)^{\oplus(n-2)}$ *for every fiber* F *of the bundle map* $X \to B$.
(2) *If* $Z = \mathbb{F}_0$*, then* $(X, \mathcal{E})$ *is either the type in* (1) *with* $B = \mathbb{P}^1$ *or* $(\mathbb{P}^n, \mathcal{O}_{\mathbb{P}^n}(2) \oplus \mathcal{O}_{\mathbb{P}^n}(1)^{\oplus(n-3)})$ *or* $(\mathbb{Q}^n, \mathcal{O}_{\mathbb{Q}^n}(1)^{\oplus(n-2)})$.
(3) *If* $Z = \mathbb{F}_1$*, then* $(X, \mathcal{E})$ *is either the type in* (1) *with* $B = \mathbb{P}^1$ *or possibly* $X \simeq \mathbb{P}_{\mathbb{P}^2}(\mathcal{F})$ *for some ample vector bundle* $\mathcal{F}$ *on* $\mathbb{P}^2$ *with* $c_1(\mathcal{F}) = k(n-2)+3$ *for some positive integer* k *and* $\mathcal{E}_F = \mathcal{O}_{\mathbb{P}^{n-2}}(1)^{\oplus(n-2)}$ *for every fiber* F *of the bundle map* $X \to \mathbb{P}^2$.

Proof. See [**LM3**]. □

Proposition 1.6. *Let* X *be a complex projective manifold of dimension* n *and let* $\mathcal{E}$ *be an ample vector bundle of rank* $r \geq 2$ *on* X*. If* $g(X, \det \mathcal{E}) = 2$*, then* $n = 2$ *and* $(X, \mathcal{E})$ *is one of the following:*

(1) X *is the Jacobian variety of a smooth curve* B *of genus* 2 *and* $\mathcal{E} \simeq \mathcal{E}_r(B, o) \otimes N$ *for some* $N \in \operatorname{Pic} X$ *with* $N \equiv 0$*, where* $\mathcal{E}_r(B, o)$ *is the Jacobian bundle for some point* o *on* B;
(2) $X \simeq \mathbb{P}_B(\mathcal{F})$ *for some stable vector bundle* $\mathcal{F}$ *of rank* 2 *on an elliptic curve* B *with* $c_1(\mathcal{F}) = 1$*. There is an exact sequence*
$$0 \to \mathcal{O}_X[2H(\mathcal{F}) + \rho^*G] \to \mathcal{E} \to \mathcal{O}_X[H(\mathcal{F}) + \rho^*T] \to 0,$$
where $G, T \in \operatorname{Pic} B$ *and* ρ *is the projection* $X \to B$*. We have* $(\deg G, \deg T) = (-2, 1)$ *or* $(-1, 0)$;
($2^\sharp$) $X, \mathcal{F}, B$ *and* ρ *are as in* (2) *and* $\mathcal{E} \simeq \rho^*\mathcal{G} \otimes H(\mathcal{F})$ *for some stable vector bundle* $\mathcal{G}$ *of rank* 3 *on* B *with* $c_1(\mathcal{G}) = -1$;
(3) $X \simeq \mathbb{P}_B(\mathcal{F})$ *and* $\mathcal{E} \simeq \rho^*\mathcal{G} \otimes H(\mathcal{F})$ *for some semistable vector bundles* $\mathcal{F}$ *and* $\mathcal{G}$ *of rank* 2 *on an elliptic curve* B *with* $(c_1(\mathcal{F}), c_1(\mathcal{G})) = (1, 0)$ *or* $(0, 1)$;

(4) *$-K_X$ is ample, $K_X^2 = 1$ and $\det \mathcal{E} = -2K_X$. We have $\mathcal{E} \simeq [-K_X]^{\oplus 2}$, or $c_2(\mathcal{E}) = 3$ and $r = 2$;*
(5_0) *$X \simeq \mathbb{P}^1 \times \mathbb{P}^1$ and $\mathcal{E} \simeq \mathcal{O}(1,1) \oplus \mathcal{O}(1,2)$;*
(5_1) *X is the blowing-up of $\mathbb{P}^2$ at a point and $\mathcal{E} \simeq [2L - E]^{\oplus 2}$, where L is the pull-back of $\mathcal{O}_{\mathbb{P}^2}(1)$ and E is the exceptional curve.*

Proof. See (2.25) Theorem in [**Fj2**]. □

Proposition 1.7. *Let X be a complex projective manifold of dimension n and let $\mathcal{E}$ be an ample and spanned vector bundle of rank $r \geq 2$ on X. If $g(X, \det \mathcal{E}) = 3$, then $n = 2$ and $(X, \mathcal{E})$ is one of the following:*
(1a) *$X = \mathbb{P}^2$, $\mathcal{E} = \mathcal{O}_{\mathbb{P}^2}(1)^{\oplus 4}$;*
(1b) *$X = \mathbb{P}^2$, and either $\mathcal{E} = \mathcal{O}_{\mathbb{P}^2}(1)^{\oplus 2} \oplus \mathcal{O}_{\mathbb{P}^2}(2)$ or $\mathcal{E} = T_{\mathbb{P}^2} \oplus \mathcal{O}_{\mathbb{P}^2}(1)$;*
(1c) *$X = \mathbb{P}^2$, $\operatorname{rank} \mathcal{E} = 2$ and $\det \mathcal{E} = \mathcal{O}_{\mathbb{P}^2}(4)$;*
(2a) *$X = \mathbb{F}_0$, and either $\mathcal{E} = [\sigma + f] \oplus [\sigma + 3f]$ or $\mathcal{E} = [\sigma + 2f]^{\oplus 2}$;*
(2b) *$X = \mathbb{F}_1$, $\mathcal{E} = [\sigma + 2f] \oplus [\sigma + 3f]$;*
(2c) *$X = \mathbb{F}_2$, $\mathcal{E} = [\sigma + 3f]^{\oplus 2}$;*
(3) *X is a Del Pezzo surface with $K_X^2 = 2$ and either $\mathcal{E} = [-K_X]^{\oplus 2}$, or $\mathcal{E} = \psi^*(\mathcal{Q}|_Y)$, where ψ is a birational morphism from X to a surface Y of bidegree $(4,4)$ in the Grassmannian of lines of $\mathbb{P}^3$, and $\mathcal{Q}$ is the universal rank 2 quotient bundle;*
(4) *$X = \mathbb{P}(\mathcal{F})$, where $\mathcal{F}$ is a rank 2 vector bundle on an elliptic curve B with $c_1(\mathcal{F}) = 1$ and $\mathcal{E} = H(\mathcal{F}) \otimes \rho^*\mathcal{G}$, where $\rho : X \to B$ is the bundle projection and $\mathcal{G}$ is any rank 2 vector bundle on B defined by a nonsplitting exact sequence $0 \to \mathcal{O}_B \to \mathcal{G} \to \mathcal{O}_B(x) \to 0$, where $x \in B$.*

Proof. See (1.10) Theorem in [**BiLL**]. □

2. The cases $g(X, \mathcal{E}) = 2$ and $g(X, \mathcal{E}) = 3$.

Theorem 2.1. *Let X be a compact complex manifold of dimension n and $\mathcal{E}$ an ample vector bundle of rank r on X with $1 < r < n - 1$ and the property $(*)$ in (1.2). If $g(X, \mathcal{E}) = 2$, then $(X, \mathcal{E})$ is one of the following:*
(i) *There exists an ample line bundle A on X such that (X, A) is a Del Pezzo 4-fold of degree 1 and $\mathcal{E} = A^{\oplus 2}$ (see also (2.2.1));*
(ii) *$X \simeq \mathbb{P}_B(\mathcal{F})$ and $\mathcal{E} = H(\mathcal{F}) \otimes \pi^*\mathcal{G}$, where $\mathcal{F}$ and $\mathcal{G}$ are vector bundles on an elliptic curve B such that $\operatorname{rank} \mathcal{F} = 4$, $\operatorname{rank} \mathcal{G} = 2$, $c_1(\mathcal{F}) + 2c_1(\mathcal{G}) = 1$, and $\pi : X \to B$ is the bundle projection;*
(iii) *$X \simeq \mathbb{P}_B(\mathcal{F})$ and $\mathcal{E} = H(\mathcal{F}) \otimes \pi^*\mathcal{G}$, where $\mathcal{F}$ and $\mathcal{G}$ are vector bundles on an elliptic curve B such that $\operatorname{rank} \mathcal{F} = 5$, $\operatorname{rank} \mathcal{G} = 3$, $3c_1(\mathcal{F}) + 5c_1(\mathcal{G}) = 1$, and $\pi : X \to B$ is the bundle projection.*

Proof. Suppose that $g(X, \mathcal{E}) = 2$. Since $\mathcal{E}$ satisfies $(*)$, there exists a nonzero section $s \in H^0(X, \mathcal{E})$ whose zero locus $Z := (s)_0$ is a smooth submanifold of X of dimension $n - r$ and $2 = g(X, \mathcal{E}) = g(Z, \det \mathcal{E}_Z)$. From (1.6) we see that $n-r=2$ and $(Z,\mathcal{E}_Z)$ is one of the cases in (1.6). We make a case by case analysis in the following.

(2.1.1) If $(Z,\mathcal{E}_Z)$ is in case (1.6;1), then $K_Z=\mathcal{O}_Z$. We have $K_X+\det\mathcal{E}=\mathcal{O}_X$ since $[K_X+\det\mathcal{E}]_Z=K_Z$ and $\mathrm{Pic}(\iota):\mathrm{Pic}(X)\to\mathrm{Pic}(Z)$ is injective by (1.3). We get also that $H^1(\iota):H^1(X,\mathbb{Z})\to H^1(Z,\mathbb{Z})$ is an isomorphism by (1.3), but this is impossible since X is a Fano manifold and Z is an abelian surface.

(2.1.2) If $(Z,\mathcal{E}_Z)$ is in case (1.6;5_0), we have $r=2$ and $n=4$. By (1.4), $(X,\mathcal{E})$ is one of the cases (Q1),(Q2) and (Q3). We easily see that $g(X,\mathcal{E})\neq 2$ in cases (Q1) and (Q2). In case (Q3), we can write $\mathcal{F}=\oplus_{i=1}^4\mathcal{O}_{\mathbb{P}^1}(a_i)$. Since $\mathcal{E}$ is ample, $H(\mathcal{F})+\pi^*\mathcal{O}_{\mathbb{P}^1}(b_j)$ is ample and so is $\mathcal{F}\otimes\mathcal{O}_{\mathbb{P}^1}(b_j)$. Hence we get $a_i+b_j>0$ for every i and j. Then it follows that

$$\begin{aligned}2=2g(X,\mathcal{E})-2&=(K_X+2c_1(\mathcal{E}))c_1(\mathcal{E})c_2(\mathcal{E})\\&=2\left(-2+\sum_{i=1}^4 a_i+2(b_1+b_2)\right)\geq 4,\end{aligned}$$

a contradiction.

(2.1.3) If $(Z,\mathcal{E}_Z)$ is in case (1.6;5_1), we have $r=2$ and $n=4$. Since $Z=\mathbb{F}_1$, we see that $(X,\mathcal{E})$ is in case (1.5;3). If $(X,\mathcal{E})$ is the type (1.5;1) with $B=\mathbb{P}^1$, then we come to a contradiction by the argument of (2.1.2). Hence we have $X\simeq\mathbb{P}_{\mathbb{P}^2}(\mathcal{F})$ for some ample vector bundle $\mathcal{F}$ on $\mathbb{P}^2$ with $c_1(\mathcal{F})=2k+3$ $(k>0)$, and $\mathcal{E}_F=\mathcal{O}_{\mathbb{P}^2}(1)^{\oplus 2}$ for every fiber F of the bundle map $\pi:X\to\mathbb{P}^2$. We set $H:=H(\mathcal{F})$; we can write $\mathcal{E}=H\otimes\pi^*\mathcal{G}$ for some vector bundle $\mathcal{G}$ of rank 2 on $\mathbb{P}^2$. Since $\mathcal{E}_Z=[2L-E]^{\oplus 2}$, we can write $H_Z=aL-E$ $(2\leq a\in\mathbb{Z})$. Then we get $\mathcal{G}=\mathcal{O}_{\mathbb{P}^2}(2-a)^{\oplus 2}$, hence $\mathcal{E}=[H+\pi^*\mathcal{O}_{\mathbb{P}^2}(2-a)]^{\oplus 2}$ by $(\pi|_Z)^*\mathcal{G}=\mathcal{E}_Z\otimes[-H_Z]=[(2-a)L]^{\oplus 2}$. Since $\mathcal{E}$ is ample, $H+\pi^*\mathcal{O}_{\mathbb{P}^2}(a)$ is ample and so is $\mathcal{F}\otimes\mathcal{O}_{\mathbb{P}^2}(a)$. Then we get $c_1(\mathcal{F}\otimes\mathcal{O}_{\mathbb{P}^2}(2-a))\geq 3$, hence $2k-3a+6\geq 0$. We note that

$$3=(2L-E)^2=c_2(\mathcal{E}_Z)=c_2(\mathcal{E})^2=s_2(\mathcal{F})+4c_1(\mathcal{F})\cdot(2-a)+6(2-a)^2.$$

On the other hand, we have

$$a^2-1=(aL-E)^2=H_Z^2=H^2\cdot c_2(\mathcal{E})=s_2(\mathcal{F})+2c_1(\mathcal{F})\cdot(2-a)+(2-a)^2.$$

From these two equalities we get $(2-a)(2k-3a+7)=0$. Since $2k-3a+6\geq 0$, we have $a=2$ and then $c_2(\mathcal{F})=3$ and $\mathcal{E}=H^{\oplus 2}$. It follows that

$$2=2g(X,\mathcal{E})-2=(K_X+2c_1(\mathcal{E}))c_1(\mathcal{E})c_2(\mathcal{E})=2s_2(\mathcal{F})+4k\geq 10,$$

a contradiction.

(2.1.4) If $(Z,\mathcal{E}_Z)$ is in case (1.6;4), then $r=2$ and $n=4$. We have $2K_X+3\det\mathcal{E}=\mathcal{O}_X$ since, by adjunction, $[2K_X+3\det\mathcal{E}]_Z=2K_Z+\det\mathcal{E}_Z=\mathcal{O}_Z$ and the restriction map $\mathrm{Pic}(X)\to\mathrm{Pic}(Z)$ is injective. By setting $A:=K_X+2\det\mathcal{E}$, we get $\det\mathcal{E}=2A$ and $K_X+3A=\mathcal{O}_X$, hence (X,A) is a

Del Pezzo 4-fold. Then we set $\mathcal{E}' := \mathcal{E} \oplus A$; we get $K_X + \det \mathcal{E}' = \mathcal{O}_X$ and $\mathcal{E}' \simeq A^{\oplus 3}$ by using Proposition 7.4 in [**PSW**]. It follows that $\mathcal{E} \simeq A^{\oplus 2}$ and

$$2 = 2g(X, \mathcal{E}) - 2 = (K_X + 2c_1(\mathcal{E}))c_1(\mathcal{E})c_2(\mathcal{E}) = 2A^4,$$

hence $A^4 = 1$. Thus we obtain that $(X, \mathcal{E})$ is the case (i) of our theorem.

(2.1.5) If $(Z, \mathcal{E}_Z)$ is in case (1.6;2), then $r = 2$ and $n = 4$. Since Z is a geometrically ruled surface over an elliptic curve B, by (1.5), X is a $\mathbb{P}^3$-bundle over B and $\mathcal{E}|_{\widetilde{F}} = \mathcal{O}_{\mathbb{P}^3}(1)^{\oplus 2}$ for every fiber $\widetilde{F}$ of the ruling $\pi : X \to B$. On the other hand, we have $\mathcal{E}_Z|_F = \mathcal{O}_{\mathbb{P}^1}(1) \oplus \mathcal{O}_{\mathbb{P}^1}(2)$ for every fiber F of the ruling $\rho : Z \to B$. This is a contradiction since $\pi|_Z = \rho$. If $(Z, \mathcal{E}_Z)$ is in case (1.6;$2^\sharp$) or (1.6;3), by using (1.5), we obtain that $(X, \mathcal{E})$ is the case (ii) or (iii) of our theorem respectively. This completes the proof. □

Remark 2.2. We make some comments on (2.1).

(2.2.1) In case (2.1; i), Del Pezzo 4-folds of degree 1 have been classified in [**Fj1**], Part III. In particular, they are weighted hypersurfaces of degree 6 in the weighted projective space $\mathbb{P}(3, 2, 1, 1, 1, 1)$.

(2.2.2) We give an example of $(X, \mathcal{E})$ in case (2.1; ii) in the following. Let L_1 and L_2 be line bundles on an elliptic curve B such that $\deg L_1 = \deg L_2$ and $L_1 \neq L_2$ in $\operatorname{Pic} B$. Let $\mathcal{F}$ be an indecomposable vector bundle of rank 4 on B with $c_1(\mathcal{F}) = 1 - 2\deg L_1 - 2\deg L_2$. We set $X := \mathbb{P}_B(\mathcal{F})$, $\mathcal{G} := L_1 \oplus L_2$, and $\mathcal{E} := H(\mathcal{F}) \otimes \pi^*\mathcal{G} = \oplus_{i=1}^2 [H(\mathcal{F}) + \pi^* L_i]$, where $\pi : X \to B$ is the bundle projection. Since $c_1(\mathcal{F} \otimes L_i) = 1$, $\mathcal{F} \otimes L_i$ is ample and $h^0(B, \mathcal{F} \otimes L_i) = 1$. Then there exists an exact sequence

$$0 \to \mathcal{O}_B \to \mathcal{F} \otimes L_i \to Q_i \to 0,$$

where Q_i is a locally free sheaf of rank 3 on B. Since Q_i is ample and $c_1(Q_i) = 1$, we see that Q_i is indecomposable. We set $D_i := \mathbb{P}_B(Q_i)$ and $Z := D_1 \cap D_2$. Since $c_1(Q_2 \otimes [L_1 - L_2]) = 1$, there exists an exact sequence

$$0 \to \mathcal{O}_B \to Q_2 \otimes [L_1 - L_2] \to Q \to 0,$$

where Q is a locally free sheaf of rank 2 on B. Then we have $Z = \mathbb{P}_B(Q)$ in $|H(Q_2) + (\pi|_{D_2})^*(L_1 - L_2)|$. Thus we see that $(X, \mathcal{E})$ satisfies the condition $(*)$ and $(X, \mathcal{E})$ is an example of (2.1; ii).

(2.2.3) The authors have no example for case (2.1; iii). We note that without the condition $(*)$ we have examples for the case. Indeed, we can take semistable vector bundles $\mathcal{F}$ and $\mathcal{G}$ on an elliptic curve B with the property that $\operatorname{rank} \mathcal{F} = 5$, $\operatorname{rank} \mathcal{G} = 3$, and $3c_1(\mathcal{F}) + 5c_1(\mathcal{G}) = 1$. Let $\pi : \mathbb{P}(\mathcal{F}) \to B$ and $\pi' : \mathbb{P}(\mathcal{G}) \to B$ be the bundle projections. Then $5H(\mathcal{F}) - \pi^* \det \mathcal{F}$ is nef on $\mathbb{P}(\mathcal{F})$ and $3H(\mathcal{G}) - (\pi')^* \det \mathcal{G}$ is nef on $\mathbb{P}(\mathcal{G})$ by Theorem 3.1 in [**Mi**]. We set $\mathcal{E} := H(\mathcal{F}) \otimes \pi^*\mathcal{G}$ and let $p : \mathbb{P}(\mathcal{E}) \to B$ be the composition of the projection $\mathbb{P}(\mathcal{E}) \to \mathbb{P}(\mathcal{F})$ and π. Then $15H(\mathcal{E}) - F$ is nef on $\mathbb{P}(\mathcal{E})$ for a fiber F of p, hence $\mathcal{E}$ is ample. But it is uncertain that such $\mathcal{E}$ satisfies $(*)$.

(2.2.4) We see that every vector bundle $\mathcal{E}$ appeared in (2.1) is not spanned. Indeed, it is clear for case (2.1; i). For cases (2.1; ii) and (2.1; iii), we use the following:

Lemma 2.2.5. *Let $\mathcal{F}$ be a vector bundle of rank r on an elliptic curve. Then there exists a line sub-bundle L of $\mathcal{F}$ such that $\deg L \geq [c_1(\mathcal{F})/r]$, where $[c_1(\mathcal{F})/r]$ is the largest integer that is not greater than $c_1(\mathcal{F})/r$.*

This is a consequence of the Mukai-Sakai theorem [**MuS**], hence proof is omitted.

Suppose that $\mathcal{E}$ is spanned in case (2.1; ii). Applying the lemma to $\mathcal{F}^\vee$ and $\mathcal{G}^\vee$, we get quotient line bundles L_1 and L_2 of $\mathcal{F}$ and $\mathcal{G}$ respectively, with the property that $\deg L_1 \leq -[-c_1(\mathcal{F})/4]$ and $\deg L_2 \leq -[-c_1(\mathcal{G})/2]$. The surjection $\mathcal{F} \to L_1$ gives a section $C := \mathbb{P}(L_1)$ of the projection $\pi : \mathbb{P}_B(\mathcal{F}) \to B$. Since $H(\mathcal{F})|_C = (\pi|_C)^* L_1$, we see that $(\pi|_C)^*(L_1 \otimes L_2)$ is a quotient line bundle of $\mathcal{E}_C$, hence $L_1 \otimes L_2$ is ample and spanned. From $c_1(\mathcal{F}) + 2c_1(\mathcal{G}) = 1$ we get $\deg L_1 + \deg L_2 \leq -[(2c_1(\mathcal{G}) - 1)/4] - [-c_1(\mathcal{G})/2] = 1$; this leads to a contradiction since B is an elliptic curve. Similarly we can show that $\mathcal{E}$ is not spanned in case (2.1; iii).

Theorem 2.3. *Let X be a compact complex manifold of dimension n and $\mathcal{E}$ an ample and spanned vector bundle of rank r on X with $1 < r < n-1$. If $g(X, \mathcal{E}) = 3$, then $(X, \mathcal{E})$ is one of the following:*

(i) $(\mathbb{P}^6, \mathcal{O}_{\mathbb{P}^6}(1)^{\oplus 4})$;

(ii) $(\mathbb{P}^1 \times \mathbb{P}^3, \mathcal{O}_{\mathbb{P}^1 \times \mathbb{P}^3}(1,1)^{\oplus 2})$;

(iii) *There exists a double covering $f : X \to \mathbb{P}^4$ with branch locus $B \in |\mathcal{O}_{\mathbb{P}^4}(4)|$ and $\mathcal{E} = f^* \mathcal{O}_{\mathbb{P}^4}(1)^{\oplus 2}$.*

Proof. Suppose that $g(X, \mathcal{E}) = 3$. We argue as in the proof of (2.1). Since $\mathcal{E}$ is spanned, there exists a nonzero section $s \in H^0(X, \mathcal{E})$ whose zero locus $Z := (s)_0$ is a smooth submanifold of X of dimension $n-r$ and $3 = g(X, \mathcal{E}) = g(Z, \det \mathcal{E}_Z)$. From (1.7) we see that $n - r = 2$ and $(Z, \mathcal{E}_Z)$ is one of the cases in (1.7).

(2.3.1) If $(Z, \mathcal{E}_Z)$ is in case (1a), (1b), or (1c) of (1.7), then $Z = \mathbb{P}^2$ and $(X, \mathcal{E})$ is the case (P1) of (1.4) since $n - r = 2$. We obtain that $(X, \mathcal{E})$ is the case (i) of our theorem by $g(X, \mathcal{E}) = 3$.

(2.3.2) If $(Z, \mathcal{E}_Z)$ is in case (3) of (1.7), then $r = 2$ and $n = 4$. By setting $A := K_X + 2\det \mathcal{E}$, we infer that (X, A) is a Del Pezzo manifold and $\mathcal{E} = A^{\oplus 2}$ from the same argument as that in (2.1.4). Then we find that $A^4 = 2$ since $g(X, \mathcal{E}) = 3$. Hence we obtain that $(X, \mathcal{E})$ is the case (iii) of our theorem by [**Fj1**], Part I.

(2.3.3) If $(Z, \mathcal{E}_Z)$ is in case (2a), (2b), (2c), or (4) of (1.7), then $r = 2$ and $n = 4$. Since Z is a geometrically ruled surface, by (1.5), $(X, \mathcal{E})$ is one of the following:

(R1) $(\mathbb{P}^4, \mathcal{O}_{\mathbb{P}^4}(1) \oplus \mathcal{O}_{\mathbb{P}^4}(2))$;
(R2) $(\mathbb{Q}^4, \mathcal{O}_{\mathbb{Q}^4}(1)^{\oplus 2})$;
(R3) X is a $\mathbb{P}^3$-bundle over a smooth curve B and $\mathcal{E}_{\widetilde{F}} = \mathcal{O}_{\mathbb{P}^3}(1)^{\oplus 2}$ for every fiber $\widetilde{F}$ of the bundle map $\pi : X \to B$;
(R4) $Z = \mathbb{F}_1$, $X \simeq \mathbb{P}_{\mathbb{P}^2}(\mathcal{F})$ for some ample vector bundle $\mathcal{F}$ on $\mathbb{P}^2$ with $c_1(\mathcal{F}) = 2k + 3$ ($k > 0$), and $\mathcal{E}_{\widetilde{F}} = \mathcal{O}_{\mathbb{P}^2}(1)^{\oplus 2}$ for every fiber $\widetilde{F}$ of the bundle map $\pi : X \to \mathbb{P}^2$.

Cases (R1) and (R2) are ruled out by $g(X, \mathcal{E}) = 3$. Case (R4) comes from (2b) of (1.7), hence $\pi|_Z$ is the blowing-up $\mathbb{F}_1 \to \mathbb{P}^2$ and $\mathcal{E}_Z = [\sigma + 2f] \oplus [\sigma + 3f]$. We can write $\mathcal{E} = H(\mathcal{F}) \otimes \pi^*\mathcal{G}$ for some vector bundle $\mathcal{G}$ of rank 2 on $\mathbb{P}^2$ and $H(\mathcal{F})_Z = a\sigma + bf$ for some $a, b \in \mathbb{Z}$. Then

$$\begin{aligned} 2\sigma + 5f = \det \mathcal{E}_Z &= 2H(\mathcal{F})_Z + (\pi|_Z)^* \det \mathcal{G} \\ &= (2a + c_1(\mathcal{G}))\sigma + (2b + c_1(\mathcal{G}))f, \end{aligned}$$

hence $2a - 2b = -3$, a contradiction. In case (R3), we have $X \simeq \mathbb{P}_B(\mathcal{F})$ and $\mathcal{E} = H(\mathcal{F}) \otimes \pi^*\mathcal{G}$ for some vector bundles $\mathcal{F}$ and $\mathcal{G}$ on B such that $\operatorname{rank} \mathcal{F} = 4$ and $\operatorname{rank} \mathcal{G} = 2$. Then

$$\begin{aligned} 4 = 2g(X, \mathcal{E}) - 2 &= (K_X + 2c_1(\mathcal{E}))c_1(\mathcal{E})c_2(\mathcal{E}) \\ &= 2(2g(B) - 2 + c_1(\mathcal{F}) + 2c_1(\mathcal{G})), \end{aligned}$$

where $g(B)$ is the genus of B. Since $\mathcal{E}$ is ample, we find that $c_1(\mathcal{F}) + 2c_1(\mathcal{G}) > 0$ from $(\det \mathcal{E})^4 > 0$. It follows that $g(B) \leq 1$. In case $g(B) = 0$, we have $B \simeq \mathbb{P}^1$ and $c_1(\mathcal{F}) + 2c_1(\mathcal{G}) = 4$. Then we can write $\mathcal{F} = \sum_{i=1}^4 \mathcal{O}(a_i)$ and $\mathcal{G} = \sum_{j=1}^2 \mathcal{O}(b_j)$. By the same argument as that in (2.1.2), we infer that $a_i + b_j = 1$ for every i and j. It follows that $a_1 = \cdots = a_4$ and $b_1 = b_2$, hence $\mathbb{P}_B(\mathcal{F}) \simeq \mathbb{P}^1 \times \mathbb{P}^3$ and $\mathcal{E} = \mathcal{O}_{\mathbb{P}^1 \times \mathbb{P}^3}(1, 1)^{\oplus 2}$, which is the case (ii) of our theorem. In case $g(B) = 1$, we have $c_1(\mathcal{F}) + 2c_1(\mathcal{G}) = 2$. Then we get a contradiction by the same argument as that in (2.2.4). We have thus completed the proof. □

3. The cases $g(X, \mathcal{E}) = q(X) + 1$ and $g(X, \mathcal{E}) = q(X) + 2$.

Theorem 3.1. *Let X be a compact complex manifold of dimension n and let $\mathcal{E}$ be an ample and spanned vector bundle of rank r with $1 < r < n - 1$. Then $g(X, \mathcal{E}) = q(X) + 1$ if and only if $(X, \mathcal{E})$ is one of the following:*
(1) $(\mathbb{P}^5, \mathcal{O}_{\mathbb{P}^5}(1)^{\oplus 2})$;
(2) $(\mathbb{P}^5, \mathcal{O}_{\mathbb{P}^5}(1)^{\oplus 3})$;
(3) $(\mathbb{Q}^4, \mathcal{O}_{\mathbb{Q}^4}(1)^{\oplus 2})$.

Proof. First we note that if $(X, \mathcal{E})$ is one of the cases (1),(2) and (3) of our theorem, then we easily see that $g(X, \mathcal{E}) = 1 = q(X) + 1$. Suppose that $g(X, \mathcal{E}) = q(X) + 1$ on the contrary. Let Z be a smooth submanifold of X with $\dim Z = n - r$ defined as the zero locus of some $s \in H^0(X, \mathcal{E})$.

Then $g(X,\mathcal{E}) = g(Z, \det \mathcal{E}_Z)$. We put $A := \det \mathcal{E}_Z$; then A is ample and spanned. If $n - r \geq 3$, we take general members $D_1, \dots, D_{n-r-2} \in |A|$ with the property that $S := D_1 \cap \dots \cap D_{n-r-2}$ is a smooth surface. If $n - r = 2$, we set $S = Z$. Hence there exists a polarized surface (S, A_S) such that $g(Z, A) = g(S, A_S)$. We get $q(X) = q(Z) = q(S)$ by using (1.3). Thus we get $g(S, A_S) = q(S) + 1$.

We show that $h^0(K_S) = 0$. Indeed, it is obvious if $\kappa(S) = -\infty$, where $\kappa(S)$ is the Kodaira dimension of S. When $\kappa(S) \geq 0$, by Riemann-Roch Theorem and Vanishing Theorem, we get

$$h^0(K_S + A_S) - h^0(K_S) = g(S, A_S) - q(S) = 1.$$

If $h^0(K_S) > 0$, then

$$h^0(K_S + A_S) \geq h^0(K_S) + h^0(A_S) - 1.$$

But this is impossible since $h^0(A_S) \geq 3$. Hence $h^0(K_S) = 0$. Thus we get $g(S, A_S) \geq 2q(S)$ by Lemma 1.4 in [**Ma1**] since (S, A_S) is not a scroll over a smooth curve. Then $q(S) \leq 1$ and $g(X, \mathcal{E}) \leq 2$ by the above argument. So we obtain that $(X, \mathcal{E})$ is the case (1),(2), or (3) of our theorem by using (2.1), (2.2.4) and [**I**]. □

Remark 3.2. Let L be an ample and spanned line bundle on a compact complex manifold X of dimension $n \geq 2$. When $n \geq 3$, we have $g(X, L) = q(X) + 1$ if and only if (X, L) is a Del Pezzo manifold (see [**Fk3**]). When $n = 2$, we have $g(X, L) = q(X) + 1$ if and only if (X, L) is a Del Pezzo surface (i.e., $L = -K_X$) or $X \simeq \mathbb{P}_B(\mathcal{F})$ and $L \equiv 2H(\mathcal{F})$ for some ample vector bundle $\mathcal{F}$ of rank 2 on an elliptic curve B with $c_1(\mathcal{F}) = 1$. We can prove this by the argument in (3.1) and Theorem 3.1 in [**LP**].

Proposition 3.3. *Let X be a compact complex manifold of dimension n and let $\mathcal{E}$ be an ample and spanned vector bundle of rank r with $1 < r < n-1$. Then we have $g(X, \mathcal{E}) \neq q(X) + 2$.*

Proof. The following argument is similar to the proof of (3.1). Suppose that $g(X, \mathcal{E}) = q(X)+2$. Let Z be a smooth submanifold of X with $\dim Z = n-r$ defined as the zero locus of some $s \in H^0(X, \mathcal{E})$. Then $g(X, \mathcal{E}) = g(Z, \det \mathcal{E}_Z)$ and $\det \mathcal{E}_Z$ is ample and spanned. As in the proof of (3.1), we get a smooth surface S such that $g(Z, \det \mathcal{E}_Z) = g(S, \det \mathcal{E}_S)$. We have $q(X) = q(Z) = q(S)$, thus we get $g(S, \det \mathcal{E}_S) = q(S) + 2$. Then we find that $q(S) \leq 1$ by Theorem 3.4 in [**R**]. It follows that $g(X, \mathcal{E}) \leq 3$ and we infer that $(X, \mathcal{E})$ does not exist from (2.1), (2.2.4) and (2.3). This completes the proof. □

Remark 3.4. We see that the pairs $(X, \mathcal{E})$ in (2.3) satisfy $g(X, \mathcal{E}) = q(X)+3$. In Appendix we give a classification of polarized surfaces (X, L) such that $g(X, L) = q(X) + 2$ and L is spanned.

4. Another Lower bound for $g(X, \mathcal{E})$.

Proposition 4.1. *Let L be an ample and spanned line bundle on a compact complex manifold X with $\dim X = n \geq 2$. Then $g(X, L) \geq 2q(X) - 1$ unless (X, L) is a scroll over a smooth curve B of genus $g(B) \geq 2$.*

Proof. Since L is ample and spanned, if $n \geq 3$, we can take general members $D_1, \dots, D_{n-2} \in |L|$ such that $S := D_1 \cap \cdots \cap D_{n-2}$ is a smooth surface. If $n = 2$, we set $S = X$. Then we get $g(X, L) = g(S, L_S)$ and $q(X) = q(S)$.

If $\kappa(S) \geq 0$, then $g(X, L) = g(S, L_S) \geq 2q(S) - 1 = 2q(X) - 1$ by Corollary 3.2 in [**Fk1**].

If $\kappa(S) = -\infty$ and (S, L_S) is not a scroll over a smooth curve, then $g(X, L) = g(S, L_S) \geq 2q(S) = 2q(X)$ by Lemma 1.4 in [**Ma1**].

If $\kappa(S) = -\infty$ and (S, L_S) is a scroll over a smooth curve, then $g(X, L) = g(S, L_S) = q(S) = q(X)$. Hence we get $g(X, L) \geq 2q(X) - 1$ if $q(S) \leq 1$. So we may assume that $q(S) \geq 2$. Then we obtain that (X, L) is a scroll over a smooth curve B of genus $g(B) \geq 2$ by using Theorem 3 in [**Bă**]. □

Theorem 4.2. *Let X be a compact complex manifold with $\dim X = n$ and let $\mathcal{E}$ be an ample and spanned vector bundle of rank r with $1 < r < n - 1$. Then $g(X, \mathcal{E}) \geq 2q(X) - 1$.*

Proof. Let Z be the zero locus of some $s \in H^0(X, \mathcal{E})$ such that Z is a smooth submanifold of X with $\dim Z = n - r$. Then $g(X, \mathcal{E}) = g(Z, \det \mathcal{E}_Z)$ and $q(X) = q(Z)$. We put $A := \det \mathcal{E}_Z$; then A is ample and spanned. Since (Z, A) is not a scroll, by (4.1), we obtain that $g(X, \mathcal{E}) = g(Z, A) \geq 2q(Z) - 1 = 2q(X) - 1$. □

5. The case of a fiber space over a curve.

Definition 5.1. Here we say that a quartet $(f, X, C, \mathcal{E})$ is a *generalized polarized fiber space over a curve* if:

1. X and C are compact complex manifolds with $1 = \dim C < \dim X = n$,
2. $f : X \to C$ is a surjective morphism with connected fibers, and
3. $\mathcal{E}$ is an ample vector bundle of rank r on X.

Theorem 5.2. *Let $(f, X, C, \mathcal{E})$ be a generalized polarized fiber space over a curve with $r \leq n - 1$. Then $g(X, \mathcal{E}) \geq g(C)$.*

Proof. First we remark that the following equality holds:

$$(5.2.1) \qquad g(X, \mathcal{E}) = g(C) + \frac{1}{2}(K_{X/C} + (n - r)c_1(\mathcal{E}))c_1(\mathcal{E})^{n-r-1}c_r(\mathcal{E}) + (g(C) - 1)(c_1(\mathcal{E})^{n-r-1}c_r(\mathcal{E})F - 1),$$

where $K_{X/C} := K_X - f^*(K_C)$ and F is a general fiber of f.

If $g(C) = 0$, then Theorem 5.2 is true by [**I**]. So we may assume that $g(C) \geq 1$.

(I) The case in which $K_{X/C} + (n-r)c_1(\mathcal{E})$ is f-nef.

Then there exists a surjective map

$$f^* \circ f_*(\mathcal{O}(m(K_{X/C} + (n-r)c_1(\mathcal{E})))) \to \mathcal{O}(m(K_{X/C} + (n-r)c_1(\mathcal{E})))$$

for any large m by base point free theorem.

By Theorem A in Appendix in [**Fk2**], $f_*(\mathcal{O}(m(K_{X/C} + (n-r)c_1(\mathcal{E}))))$ is semipositive. Hence $K_{X/C} + (n-r)c_1(\mathcal{E})$ is nef. So we get

$$(K_{X/C} + (n-r)c_1(\mathcal{E}))c_1(\mathcal{E})^{n-r-1}c_r(\mathcal{E}) \geq 0.$$

Hence we obtain $g(X, \mathcal{E}) \geq g(C)$ because of (5.2.1) and $c_1(\mathcal{E})^{n-r-1}c_r(\mathcal{E})F \geq 1$.

(II) The case in which $K_{X/C} + (n-r)c_1(\mathcal{E})$ is not f-nef.

Then $K_X + (n-r)c_1(\mathcal{E})$ is not nef. So by Mori Theory, there exists an extremal rational curve l such that $(K_X + (n-r)c_1(\mathcal{E}))l < 0$. Hence

$$n+1 \geq -K_X l > (n-r)c_1(\mathcal{E})l \geq (n-r)r \geq n-1.$$

If $(n-r)r = n$, then $(n,r) = (4,2)$.

If $(n-r)r = n-1$, then $r = 1$ or $r = n-1$.

(II-1) The case where $(n,r) = (4,2)$.

Then $-K_X l = 5 = n+1$. So we have $\operatorname{Pic} X \cong \mathbb{Z}$ by [**W**]. But this is impossible because X has a nontrivial fibration.

(II-2) The case in which $r = 1$.

Then Theorem 5.2 is true by Theorem 1.2.1 in [**Fk2**].

(II-3) The case in which $r = n-1$.

If $n = 2$, then $r = 1$ and so we may assume that $n \geq 3$. Since X has a nontrivial fibration, $(X, \mathcal{E})$ is the following type by [**YZ**]: There exists a surjective morphism $\pi : X \to B$ such that any fiber of π is $\mathbb{P}^{n-1}$ and $\mathcal{E}|_{F_\pi} \cong \mathcal{O}(1)^{\oplus n-1}$, where B is a smooth curve and F_π is a fiber of π.

Since any fiber of π is $\mathbb{P}^{n-1}$, there exists a morphism $\delta : B \to C$ such that $f = \delta \circ \pi$. Because f has connected fibers, δ is an isomorphism. In particular, $g(B) = g(C)$. On the other hand, by [**Ma2**], $g(X, \mathcal{E}) = g(B)$. Hence $g(X, \mathcal{E}) = g(C)$. This completes the proof of Theorem 5.2. □

Theorem 5.3. *Let $(f, X, C, \mathcal{E})$ be a generalized polarized fiber space over a curve with $2 \leq r \leq n-1$. If $g(X, \mathcal{E}) = g(C)$, then $r = n-1$, any fiber F of f is isomorphic to $\mathbb{P}^{n-1}$ and $\mathcal{E}|_F \cong \oplus_{i=1}^{n-1}\mathcal{O}_{\mathbb{P}^{n-1}}(1)$.*

Proof. (I) The case in which $g(C) \leq 1$.

Then $g(X, \mathcal{E}) = g(C) \leq 1$, and by the classification results of [**I**] and [**Ma2**], we get the following: X is a $\mathbb{P}^{n-1}$-bundle over $\mathbb{P}^1$ or a smooth elliptic curve and $\mathcal{E}|_{F_\pi} \cong \mathcal{O}_{\mathbb{P}^{n-1}}(1)^{\oplus n-1}$, where F_π is a fiber of its bundle map

$\pi : X \to B$ and B is $\mathbb{P}^1$ or a smooth elliptic curve. Since any fiber of π is $\mathbb{P}^{n-1}$, there exists a morphism $\delta : B \to C$ such that $f = \delta \circ \pi$. Because f has connected fibers, δ is an isomorphism. Therefore we get the assertion.

(II) The case in which $g(C) \geq 2$.

(II-1) $n - r \geq 2$ case.

If $K_{X/C} + (n-r-1)c_1(\mathcal{E})$ is f-nef, then by the same argument as in the proof of Theorem 5.2 we get

$$(K_{X/C} + (n-r-1)c_1(\mathcal{E}))c_1(\mathcal{E})^{n-r-1}c_r(\mathcal{E}) \geq 0$$

and

$$(K_{X/C} + (n-r)c_1(\mathcal{E}))c_1(\mathcal{E})^{n-r-1}c_r(\mathcal{E}) \geq 1.$$

Hence we obtain that $g(X, \mathcal{E}) > g(C)$ by (5.2.1). So we may assume that $K_{X/C} + (n-r-1)c_1(\mathcal{E})$ is not f-nef. Then by Mori Theory, there exists an extremal rational curve l such that $(K_X + (n-r-1)c_1(\mathcal{E}))l < 0$. Hence we get

$$n + 1 \geq -K_X l > (n-r-1)c_1(\mathcal{E})l \geq (n-r-1)r \geq n-2.$$

If $(n-r-1)r = n$, then $-K_X l = n+1$ and $\operatorname{Pic} X \cong \mathbb{Z}$ by [**W**]. But this is impossible.

If $(n-r-1)r = n-1$, then $n = 5$ and $r = 2$.

Here we prove the following Lemma.

Lemma 5.4. *Let $(f, X, C, \mathcal{E})$ be a generalized polarized fiber space over a curve with $2 \leq r \leq n-1$ and $g(C) \geq 1$. If $\kappa(K_F + xc_1(\mathcal{E}_F)) \geq 0$ for a rational number x with $x < n-r$ and a general fiber F of f, then $g(X, \mathcal{E}) \geq g(C) + 1$.*

Proof. By assumption, there exists a natural number N such that

$$f_*(\mathcal{O}(N(K_{X/C} + xc_1(\mathcal{E})))) \neq 0.$$

By Remark 1.3.2 in [**Fk2**], $N(K_{X/C} + xc_1(\mathcal{E}))$ is pseudo effective. Therefore

$$(K_{X/C} + xc_1(\mathcal{E}))c_1(\mathcal{E})^{n-r-1}c_r(\mathcal{E}) \geq 0$$

and we get

$$(K_{X/C} + (n-r)c_1(\mathcal{E}))c_1(\mathcal{E})^{n-r-1}c_r(\mathcal{E}) \geq 1.$$

Since $g(C) \geq 1$, we get that $g(X, \mathcal{E}) \geq g(C) + 1$ by (5.2.1). $\square$

We continue the proof of Theorem 5.3. If $K_F + xc_1(\mathcal{E}_F)$ is nef for a rational number x with $x < 3$, then we can prove that $g(X, \mathcal{E}) > g(C)$ by Lemma 5.4.

Assume that $K_F + xc_1(\mathcal{E}_F)$ is not nef for a rational number x with $x < 3$. Then there exists an extremal rational curve l on F such that $n \geq -K_F l > xc_1(\mathcal{E}_F)l \geq rx$. Since $n = 5$ and $r = 2$, we have $x < 5/2$. Therefore there exists a rational number $y < 3$ such that $K_F + yc_1(\mathcal{E}_F)$ is nef, and we get $g(X, \mathcal{E}) > g(C)$.

If $(n-r-1)r = n-2$, then $r = n-2$ by assumption. Assume that $K_F + xc_1(\mathcal{E}_F)$ is not nef for a rational number x with $x < 2$. Then we get $n > rx$ by the same argument as above. Since $r = n-2$, we get $x < n/(n-2) = 1 + 2/(n-2)$. By assumption, we get $n \geq 4$. So we have $x < 2$. Therefore there exists a rational number $y < 2$ such that $K_F + yc_1(\mathcal{E}_F)$ is nef. Hence we have $g(X, \mathcal{E}) > g(C)$.

(II-2) $n - r = 1$ case.

First we assume that $K_F + c_1(\mathcal{E}_F)$ is nef for a general fiber F of f. If $K_F + c_1(\mathcal{E}_F)$ is ample, then there exists a rational number $t > 0$ such that $\kappa(K_F + (1-t)c_1(\mathcal{E}_F)) \geq 0$ by Kodaira's Lemma. So we get that $g(X, \mathcal{E}) > g(C)$ by the same argument as above. Assume that $K_F + c_1(\mathcal{E}_F)$ is not ample. Since $\dim F = \operatorname{rank} \mathcal{E}_F$, by [**Fj3**], we get that $(F, \mathcal{E}_F)$ is one of the following:

(A) $(\mathbb{P}^{n-1}, \mathcal{O}_{\mathbb{P}^{n-1}}(2) \oplus \mathcal{O}_{\mathbb{P}^{n-1}}(1)^{\oplus n-2})$,
(B) $(\mathbb{P}^{n-1}, T_{\mathbb{P}^{n-1}})$,
(C) $(\mathbb{Q}^{n-1}, \mathcal{O}_{\mathbb{Q}^{n-1}}(1)^{\oplus n-1})$,
(D) F is a $\mathbb{P}^{n-2}$-bundle over a smooth curve B and $\mathcal{E}_{F_\pi} = \mathcal{O}_{\mathbb{P}^1}(1)^{\oplus n-1}$ for every fiber F_π of the projection $\pi : F \to B$.

If $(F, \mathcal{E}_F)$ is one of the type (A), (B), or (C), then $h^0(K_F + c_1(\mathcal{E}_F)) > 0$ by easy calculation. Here we prove the following Lemma.

Lemma 5.5. *Let $(f, X, C, \mathcal{E})$ be a generalized polarized fiber space over a curve with $2 \leq r \leq n-1$. If $h^0(K_F + c_1(\mathcal{E}_F)) > 0$ for a general fiber F of f, then $(K_{X/C} + c_1(\mathcal{E}))c_1(\mathcal{E})^{n-r-1}c_r(\mathcal{E}) > 0$.*

Proof. By hypothesis, $f_*\mathcal{O}(K_{X/C} + c_1(\mathcal{E})) \neq 0$. By Theorem 2.4 and Corollary 2.5 in [**EV**], we get that $f_*\mathcal{O}(K_{X/C} + c_1(\mathcal{E}))$ is ample. By the proof of Lemma 1.4.1 in [**Fk2**], we get that $m(K_{X/C} + c_1(\mathcal{E})) - f^*A$ is an effective divisor for a large number m and an ample divisor A on C. Hence we obtain $(K_{X/C} + c_1(\mathcal{E}))c_1(\mathcal{E})^{n-r-1}c_r(\mathcal{E}) > 0$. □

By Lemma 5.5, we get that $g(X, \mathcal{E}) > g(C)$ if $(F, \mathcal{E}_F)$ is one of the type (A), (B), or (C).

Assume that $(F, \mathcal{E}_F)$ is the type (D). Then there exist vector bundles $\mathcal{F}$ and $\mathcal{G}$ on B with $\operatorname{rank} \mathcal{F} = \operatorname{rank} \mathcal{G} = n-1$ such that $\mathcal{E}_F \cong H(\mathcal{F}) \otimes \pi^*(\mathcal{G})$, where $H(\mathcal{F})$ is the tautological line bundle of $\mathbb{P}(\mathcal{F})$. Then $K_F + c_1(\mathcal{E}_F) = \pi^*(K_B + \det \mathcal{F} + \det \mathcal{G})$. Since $K_F + c_1(\mathcal{E}_F)$ is nef, we get $(K_{X/C} + c_1(\mathcal{E}))c_r(\mathcal{E}) \geq 0$ by the proof of Lemma 5.4. We have $g(X, \mathcal{E}) = g(C)$, then

$c_r(\mathcal{E})F = 1$ by (5.2.1). Since $1 = c_r(\mathcal{E}_F) = c_1(\mathcal{F}) + c_1(\mathcal{G})$, we obtain that

$$\begin{aligned} &h^0(K_B + \det \mathcal{F} + \det \mathcal{G}) \\ &\geq 1 - g(B) + \deg(K_B + \det \mathcal{F} + \det \mathcal{G}) \\ &= g(B) - 1 + c_1(\mathcal{F}) + c_1(\mathcal{G}) \\ &= g(B). \end{aligned}$$

Because $K_F + c_1(\mathcal{E}_F)$ is nef, we obtain that $\deg(K_B + \det \mathcal{F} + \det \mathcal{G}) \geq 0$. Hence $g(B) \geq 1$. Therefore $h^0(K_F + c_1(\mathcal{E}_F)) \geq 1$. By Lemma 5.5 we obtain that $g(X, \mathcal{E}) > g(C)$ and this is a contradiction.

Next we assume that $K_F + c_1(\mathcal{E}_F)$ is not nef. Then $K_X + c_1(\mathcal{E})$ is not nef and the same argument as in the proof of Theorem 5.2, case (II-3), shows that $(f, X, C, \mathcal{E})$ is as required. This completes the proof of Theorem 5.3. □

Remark 5.6. Let $(f, X, C, \mathcal{E})$ be as in Theorem 5.2. Suppose that $g(X, \mathcal{E}) = g(C)$ and $r = 1$. Then by Theorem 1.4.2 and Proposition 1.4.3 in [**Fk2**], $(f, X, C, \mathcal{E})$ is a scroll (in the sense of [**Fk2**], §0) unless $n = 2$ and $(f, X, C, \mathcal{E}) \cong (\pi, \mathbb{P}^1 \times \mathbb{P}^1, \mathbb{P}^1, L)$, where π is one projection such that $LF_\pi \geq 2$ for a fiber F_π of π. By the other projection ρ, however, $(\rho, \mathbb{P}^1 \times \mathbb{P}^1, \mathbb{P}^1, L)$ becomes a scroll.

Appendix.

Proposition A. *Let (X, L) be a quasi-polarized surface (i.e., L is a nef and big line bundle on a smooth surface X) such that $\kappa(X) = 2$ and $h^0(L) \geq 2$. Then $K_X L \geq 2q(X) - 2$. If equality holds and (X, L) is L-minimal (i.e., $LE > 0$ for any (-1)-curve E on X), then (X, L) is the following:*

$X \cong F \times C$ and $L \equiv C + 2F$, where F and C are smooth curves with $g(F) = 2$ and $g(C) \geq 2$.

Proof. See [**Fk4**]. □

Proposition B. *Let (X, L) be a polarized surface with $\kappa(X) = 0$ or 1. Assume that L is spanned. Then $g(L) := g(X, L) \geq 2q(X)$. Furthermore if $g(L) = 2q(X)$, then (X, L) is one of the following:*

(1) *(X, L) is a polarized abelian surface with $L^2 = 6$ such that $(X, L) \not\cong (E_1 \times E_2, p_1^*(D_1) + p_2^*(D_2))$, where E_i is a smooth elliptic curve, p_i is the i-th projection, and $D_i \in \mathrm{Pic}(E_i)$ for $i = 1, 2$ with $\deg D_1 = 1$ and $\deg D_2 = 3$.*

(2) *X is a one point blowing up of S, and $L = \mu^* A - 2E$, where S is an abelian surface, A is an ample line bundle with $A^2 = 8$, $\mu : X \to S$ is its blowing up, and E is a (-1)-curve of μ.*

(3) *$\kappa(X) = 1$, $L^2 = 4$, $q(X) = 3$, X has a locally trivial elliptic fibration $f : X \to C$, and $LF = 3$ for a fiber F of f, where C is a smooth curve with $g(C) = 2$.*

Proof. See [**Fk5**]. □

Theorem. *Let X be a smooth projective surface and let L be an ample and spanned line bundle on X. If $g(L) = q(X) + 2$, then (X, L) is one of the following:*

(1) *(X, L) is a relatively minimal conic bundle over a smooth curve B of genus two (i.e., X is a $\mathbb{P}^1$-bundle over B and $L_F = \mathcal{O}_{\mathbb{P}^1}(2)$ for every fiber F of the ruling).*

(2) *X is a $\mathbb{P}^1$-bundle X_0 blown-up at s $(0 \le s \le 4)$ points $p_1, \dots, p_s$ on distinct fibers and $L = \pi^* L_0 - E_1 - \cdots - E_s$, where $\pi : X \to X_0$ is the blowing up, $E_i = \pi^{-1}(p_i)$, X_0 is an elliptic $\mathbb{P}^1$-bundle of invariant $e \le 0$, and $L_0 \equiv 2\sigma + (e+2)f$ (σ is a minimal section with $\sigma^2 = -e$ and f is a fiber).*

(3) *X is an $\mathbb{F}_e$ $(e \le 2)$ blown-up at s $(0 \le s \le 9)$ points $p_1, \dots, p_s$ on distinct fibers and $L = \pi^* L_0 - E_1 - \cdots - E_s$, where $\pi : X \to \mathbb{F}_e$ is the blowing up, $E_i = \pi^{-1}(p_i)$, and $L_0 \equiv 2\sigma + (e+3)f$.*

(4) *X is a Del Pezzo surface of degree one and there exists a double covering $\pi : X \to \mathcal{Q} \subset \mathbb{P}^3$ of a quadric cone $\mathcal{Q}$ branched at the vertex and along the transverse intersection of $\mathcal{Q}$ with a cubic surface and $L = \pi^*(\mathcal{O}_{\mathcal{Q}}(1))$.*

(5) *(X, L) is a polarized abelian surface with $L^2 = 6$ such that $(X, L) \not\cong (E_1 \times E_2, p_1^*(D_1) + p_2^*(D_2))$, where E_i is a smooth elliptic curve, p_i is the i-th projection, and $D_i \in \operatorname{Pic}(E_i)$ for $i = 1, 2$ with $\deg D_1 = 1$ and $\deg D_2 = 3$.*

(6) *X is a blowing up of an abelian surface S at one point p and $L = \pi^* A - 2E$, where $\pi : X \to S$ is the blowing up, $E = \pi^{-1}(p)$, and A is an ample line bundle on S with $A^2 = 8$.*

(7) *X is a $K3$ surface which is a double covering of $\mathbb{P}^2$ branched along a smooth curve of degree six and L is the pull back of $\mathcal{O}_{\mathbb{P}^2}(1)$.*

Proof. (I) The case in which $\kappa(X) = 0$ or 1.

Then by Proposition B, we get that $g(L) \ge 2q(X)$. So we obtain $q(X) \le 2$ by assumption.

(I-1) If $q(X) = 2$, then $g(L) = q(X) + 2 = 2q(X)$ and by Proposition B we get the type (5) and (6) in Theorem.

(I-2) If $q(X) \le 1$, then $g(L) \le 3$ and $L^2 \le 4$ by $K_X L \ge 0$.

(I-2-1) If $L^2 = 4$, then $\kappa(X) = 0$ and X is minimal since $K_X L = 0$. So by Kodaira vanishing Theorem and Riemann-Roch Theorem, we get the equality: $h^0(L) = L^2/2 + \chi(\mathcal{O}_X) = 2 + \chi(\mathcal{O}_X)$. Because L is ample and spanned, we obtain $h^0(L) \ge 3$ and $\chi(\mathcal{O}_X) \ge 1$. But then $q(X) = 0$ by the classification theory of surfaces and this is impossible.

(I-2-2) If $L^2 = 3$, then $g(L) = 3$, $K_X L = 1$, and $q(X) = 1$. We have $h^0(L) \ge 3$ since L is ample spanned.

If $h^0(L) \geq 4$, then $g(L) > \Delta(L)$ and $L^2 \geq 2\Delta(L) + 1$, where $\Delta(L) := 2 + L^2 - h^0(L)$ is the Δ-genus of L. But then $q(X) = 0$ (see e.g. (I.3.5) in [**Fj4**]).

If $h^0(L) = 3$, then there is a triple covering $\varphi_{|L|} : X \to \mathbb{P}^2$ which is defined by $|L|$. Let $\mathcal{E}$ be a vector bundle of rank two on $\mathbb{P}^2$ such that $\pi_*\mathcal{O}_X = \mathcal{O}_{\mathbb{P}^2} \oplus \mathcal{E}$. By Lemma 3.2 in [**Be**], we get the following two equalities:

(i) $\chi(\mathcal{O}_X) = (1/2)g(L)(g(L) + 1) + 2 - c_2$,
(ii) $K_X^2 = 2g(L)^2 - 4g(L) + 11 - 3c_2$,

where $c_2 := c_2(\mathcal{E})$. Since $g(L) = 3$, we get that $3\chi(\mathcal{O}_X) - K_X^2 = 7$ by the above equalities.

If $\kappa(X) = 0$, then $K_X^2 = -1$ because $K_X L = 1$. So we get $\chi(\mathcal{O}_X) = 2$. But by the classification theory of surfaces, this is impossible because $q(X) = 1$.

If $\kappa(X) = 1$, then X is minimal and $K_X^2 = 0$ because $K_X L = 1$. But then $3\chi(\mathcal{O}_X) = 7$ and this is impossible.

(I-2-3) If $L^2 = 2$, then $K_X L = 0$ or 2. Since $\kappa(X) \geq 0$, we get that $\Delta(L) \geq 1$ and $h^0(L) = 3$. Then there exists a double covering $\varphi_{|L|} : X \to \mathbb{P}^2$ which is defined by $|L|$. We remark that $K_X = \varphi_{|L|}^*(K_{\mathbb{P}^2} + D)$ for some $D \in \mathrm{Pic}(\mathbb{P}^2)$. Since $\kappa(X) = 0$ or 1, we get that $\kappa(X) = 0$ and so X is minimal. In particular $K_X = \mathcal{O}_X$. Therefore $K_X L = 0$ and $g(L) = 2$. Since $h^0(L) = L^2/2 + \chi(\mathcal{O}_X) = 1 + \chi(\mathcal{O}_X)$, we get $\chi(\mathcal{O}_X) = 2$. Hence X is a K3 surface by the Classification theory of surfaces. This is the type (7) in Theorem.

(II) The case in which $\kappa(X) = 2$.

Then by Corollary 3.2 in [**Fk1**], we get $g(L) \geq 2q(X) - 1$. So we obtain $q(X) \leq 3$ and $g(L) \leq 5$ by assumption. Furthermore $L^2 \leq 3$ by Proposition A because L is spanned. (We remark that L is L-minimal if L is ample.)

If $h^0(L) \geq 4$, then $g(L) > 1 \geq \Delta(L)$ and $L^2 \geq 2\Delta(L) + 1$. On the other hand, since $\kappa(X) \geq 0$, we obtain that $\Delta(L) = 1$ and $L^2 = 3$. So we get $q(X) = 0$ and $g(L) \geq 3$ and this is impossible. Therefore $h^0(L) = 3$.

If $L^2 = 3$, then there exists a triple covering $\varphi_{|L|} : X \to \mathbb{P}^2$ which is defined by $|L|$. In this case, by the same argument as above, we get

$$2(K_X^2 - 3\chi(\mathcal{O}_X)) = (g(L) - 1)(g(L) - 10).$$

Since $3 \leq g(L) \leq 5$, we get the following:

(α) $(g(L), q(X), K_X L, K_X^2 - 3\chi(\mathcal{O}_X)) = (5, 3, 5, -10)$,
(β) $(g(L), q(X), K_X L, K_X^2 - 3\chi(\mathcal{O}_X)) = (4, 2, 3, -9)$,
(γ) $(g(L), q(X), K_X L, K_X^2 - 3\chi(\mathcal{O}_X)) = (3, 1, 1, -7)$.

Claim. *The above three cases cannot occur.*

Proof. (II-1) The case (γ).

In this case X is minimal because $K_X L = 1$. But then this is impossible by Hodge index Theorem.

(II-2) The case (β).

If X is minimal, then $K_X^2 \geq 2q(X) = 4$ by Théorème 6.1 in [**D**]. On the other hand, $K_X^2 \leq 3$ by Hodge index Theorem and this is a contradiction.

So we get that X is not minimal. Let $\mu := \mu_r \circ \cdots \circ \mu_1 : X := X_0 \to X_1 \to \cdots \to X_{r-1} \to X_r =: X'$ be an admissible minimalization of X and let $m = (m_r, \ldots, m_1)$ be the weight sequence of this minimalization (see (II.14.4) in [**Fj4**]). We remark that $m_r \geq \cdots \geq m_1$.

If $m_1 = 1$, then $g(L_1) = q(X_1) + 1$ and $h^0(L_1) \geq 2$, where $L_1 := (\mu_1)_*(L)$ in the sense of cycle theory. But then this is impossible by Proposition A because $2 = K_X L > K_{X_1} L_1$. So we get $m_1 \geq 2$. Then $L_1^2 \geq 7$ and $K_{X_1} L_1 \leq 1$. Hence X_1 is minimal and this is a contradiction by Hodge index Theorem.

(II-3) The case (α).

If X is minimal, then $\chi(\mathcal{O}_X) \geq 4$ because $3\chi(\mathcal{O}_X) = K_X^2 + 10$. Furthermore $p_g(X) \geq 6$ since $q(X) = 3$. Hence $K_X^2 \geq 2p_g(X) \geq 12$ by Théorème 6.1 in [**D**]. But this is impossible by Hodge index Theorem. So we get that X is not minimal. By the same argument as in the case (II-2) we get a contradiction. □

We continue the proof of Theorem.

If $L^2 = 2$, then there exists a double covering $\varphi_{|L|} : X \to \mathbb{P}^2$ which is defined by $|L|$. Let $\mathcal{O}_{\mathbb{P}^2}(a)$ be a line bundle on $\mathbb{P}^2$ such that $B \in |\mathcal{O}_{\mathbb{P}^2}(2a)|$, where B is the branch locus. Then $(\varphi_{|L|})_*(\mathcal{O}_X) = \mathcal{O}_{\mathbb{P}^2} \oplus \mathcal{O}_{\mathbb{P}^2}(-a)$. Hence

$$h^1(\mathcal{O}_X) = h^1((\varphi_{|L|})_*(\mathcal{O}_X)) = h^1(\mathcal{O}_{\mathbb{P}^2}) + h^1(\mathcal{O}_{\mathbb{P}^2}(-a)) = 0.$$

So we get $g(L) = 2$. But since $K_X L > 0$ and $L^2 = 2$, this is impossible.

(III) The case in which $\kappa(X) = -\infty$.

Since (X, L) is not a scroll over a smooth curve, we get $g(L) \geq 2q(X)$ by Lemma 1.4 in [**Ma1**]. So $q(X) \leq 2$.

(III-1) The case in which $q(X) = 2$.

In this case, $g(L) = q(X) + 2 = 2q(X)$. Since $K_X + L$ is nef, we get

$$\begin{aligned} 0 \leq (K_X + L)^2 &= (K_X)^2 + 2(K_X + L)L - L^2 \\ &\leq 8(1 - q(X)) + 4(g(L) - 1) - L^2 \\ &= 4(g(L) - 2q(X) + 1) - L^2. \end{aligned}$$

Hence $L^2 \leq 4$ in this case.

If $L^2 = 4$, then X is relatively minimal and $(K_X + L)^2 = 0$, that is, (X, L) is a relatively minimal conic bundle over a smooth curve. This is the type (1) in Theorem.

If $L^2 \leq 3$ and $h^0(L) \geq 4$, then we get a contradiction as in (I-2-2). So we may assume that $L^2 \leq 3$ and $h^0(L) = 3$.

If $L^2 = 3$, then $K_X L = 3$ and there is a triple covering $\varphi_{|L|} : X \to \mathbb{P}^2$ which is defined by $|L|$. Since $\chi(\mathcal{O}_X) = -1$, we get that $K_X^2 = -12$ by Lemma 3.2 in [**Be**]. Here we calculate $(K_X + L)^2$;

$$(K_X + L)^2 = K_X^2 + 2K_X L + L^2 = -12 + 6 + 3 < 0.$$

But this is a contradiction because $K_X + L$ is nef.

If $L^2 = 2$, then there is a double covering $\varphi_{|L|} : X \to \mathbb{P}^2$ which is defined by $|L|$. But then $q(X) = 0$ and this is a contradiction.

(III-2) The case in which $q(X) = 1$.

Then $g(L) = 3$. Here we use the classification of polarized surfaces with sectional genus three by [**LL**].

Claim. *The case in which $L^2 = 3$ cannot occur.*

Proof. If $L^2 = 3$ and $h^0(L) \geq 4$, then $g(L) > 1 \geq \Delta(L)$ and $L^2 \geq 2\Delta(L)+1$. But this is impossible because $q(X) = 1$. So we may assume that $h^0(L) = 3$. Then there is a triple covering $\varphi_{|L|} : X \to \mathbb{P}^2$ which is defined by $|L|$. Since $\chi(\mathcal{O}_X) = 0$, we get $K_X^2 = -7$ by Lemma 3.2 in [**Be**]. But in the table II of [**LL**], the case in which $L^2 = 3$ cannot occur. □

Next we prove that the following case cannot occur (see (2.6) in [**LL**]):

X is an elliptic $\mathbb{P}^1$-bundle $X_\sharp$ of invariant $e = 0$, blown up at a single point p not lying on a curve $D \in |m\sigma|$, $m \leq 2$ and $L = \eta^*[4\sigma + (2e+1)f] \otimes [E]^{-2}$. (Here we use the same notations as in [**LL**].)

Let σ' be the strict transform of σ under η. Since

$$0 < L\sigma' = (4\sigma + f)\sigma - 2E\sigma' = 1 - 2E\sigma',$$

we see that $E\sigma' = 0$ and $L\sigma' = 1$. It follows that $\sigma \cong \sigma' \cong \mathbb{P}^1$ since L is spanned. This is a contradiction.

By the above argument, we obtain the type (2) in Theorem by the classification of polarized surfaces with sectional genus three (see [**LL**]).

(III-3) The case in which $q(X) = 0$.

Then $g(L) = 2$. So by Theorem 3.1 in [**LP**] we get the type (3) and (4) in Theorem. □

References

[Bă] L. Bădescu, *On ample divisors*: II, in 'Proceedings of the week of Algebraic Geometry', Bucharest, 1980; Teubner Texte Math., Band 40, (1981), 12-32.

[Be] G.M. Besana, *On polarized surfaces of degree three whose adjoint bundles are not spanned*, Arch. Math. (Basel), **65** (1995), 161-167.

[BiLL] A. Biancofiore, A. Lanteri and E.L. Livorni, *Ample and spanned vector bundles of sectional genera three*, Math. Ann., **291** (1991), 87-101.

[D] O. Debarre, *Inégalités numériques pour les surfaces de type général*, Bull. Soc. Math. France, **110** (1982), 319-346; *Addendum*, Bull. Soc. Math. France, **111** (1983), 301-302.

[EV] H. Esnault and E. Viehweg, *Effective bounds for semipositive sheaves and for the height of points on curves over complex function fields*, Compositio Math., **76** (1990), 69-85.

[Fj1] T. Fujita, *On the structure of polarized manifolds with total deficiency one*, I, J. Math. Soc. Japan, **32** (1980), 709-725; II, J. Math. Soc. Japan, **33** (1981), 415-434; III, J. Math. Soc. Japan, **36** (1984), 75-89.

[Fj2] ———, *Ample vector bundles of small c_1-sectional genera*, J. Math. Kyoto Univ., **29** (1989), 1-16.

[Fj3] ———, *On adjoint bundles of ample vector bundles*, in 'Complex Algebraic Varieties', Bayreuth, 1990; Lecture Notes in Math., **1507**, Springer, (1992), 105-112.

[Fj4] ———, *Classification Theories of Polarized Varieties*, London Math. Soc. Lecture Note Ser., **155**, Cambridge Univ. Press, 1990.

[Fk1] Y. Fukuma, *On sectional genus of quasi-polarized manifolds with nonnegative Kodaira dimension*, Math. Nachr., **180** (1996), 75-84.

[Fk2] ———, *A lower bonnd for sectional genus of quasi-polarized manifolds*, J. Math. Soc. Japan, **49** (1997), 339-362.

[Fk3] ———, *On polarized 3-folds (X, L) with $g(L) = q(X) + 1$ and $h^0(L) \geq 4$*, Ark. Mat., **35** (1997), 299-311.

[Fk4] ———, *A lower bound for $K_X L$ of quasi-polarized surfaces (X, L) with nonnegative Kodaira dimension*, Canad. J. Math., **50** (1998), 1209-1235.

[Fk5] ———, *On sectional genus of k-very ample line bundles on smooth surfaces with nonnegative Kodaira dimension*, Kodai. Math. J., **21** (1998), 153-178.

[I] H. Ishihara, *A generalization of curve genus for ample vector bundles*, I, Comm. Algebra, **27** (1999), 4327-4335.

[LL] A. Lanteri and E.L. Livorni, *Complex surfaces polarized by an ample and spanned line bundle of genus three*, Geom. Dedicata, **31** (1989), 267-289.

[LM1] A. Lanteri and H. Maeda, *Ample vector bundles with sections vanishing on projective spaces or quadrics*, Internat. J. Math., **6** (1995), 587-600.

[LM2] ———, *Ample vector bundle characterizations of projective bundles and quadric fibrations over curves*, in 'Higher Dimensional Complex Varieties', Trento, 1994, de Gruyter, (1996), 247-259.

[LM3] ———, *Geometrically ruled surfaces as zero loci of ample vector bundles*, Forum Math., **9** (1997), 1-15.

[LP] A. Lanteri and M. Palleschi, *Adjunction properties of polarized surfaces via Reider's method*, Math. Scand., **65** (1989), 175-188.

[Ma1] H. Maeda, *On polarized surfaces of sectional genus three*, Sci. Papers College Arts Sci. Univ. Tokyo, **37** (1987), 103-112.

[Ma2] ———, *Ample vector bundles of small curve genera*, Arch. Math. (Basel), **70** (1998), 239-243.

[Mi] Y. Miyaoka, *The Chern classes and Kodaira dimension of a minimal variety*, in 'Algebraic Geometry', Sendai, 1985; Adv. Stud. in Pure Math., **10**, Kinokuniya, (1987), 449-476.

[MuS] S. Mukai and F. Sakai, *Maximal subbundles of vector bundles on a curve*, Manuscripta Math., **52** (1985), 251-256.

[PSW] T. Peternell, M. Szurek and J.A. Wiśniewski, *Fano manifolds and vector bundles*, Math. Ann., **294** (1992), 151-165.

[R] F. Russo, *Some inequalities for ample and spanned vector bundles on algebraic surfaces*, Boll. Un. Mat. Ital. A (7), **8** (1994), 323-333.

[W] J.A. Wiśniewski, *Length of extremal rays and generalized adjunction*, Math. Z., **200** (1989), 409-427.

[YZ] Y.G. Ye and Q. Zhang, *On ample vector bundles whose adjunction bundles are not numerically effective*, Duke Math. J., **60** (1990), 671-687.

Received August 5, 1998. Both authors are research Fellows of the Japan Society for the Promotion of Science.

Tokyo Institute of Technology
Oh-okayama, Meguro-ku
Tokyo 152-8551
Japan
E-mail address: fukuma@math.titech.ac.jp

Naruto University of Education
Takashima, Naruto-cho
Naruto-shi 772-8502
Japan

Tokyo Institute of Technology
Oh-okayama, Meguro-ku
Tokyo 152-8551
Japan
E-mail address: ishihara@math.titech.ac.jp

PACIFIC JOURNAL OF MATHEMATICS
Vol. 193, No. 2, 2000

L^2 SPECTRAL DECOMPOSITION ON THE HEISENBERG GROUP ASSOCIATED TO THE ACTION OF $U(p,q)$

T. Godoy and L. Saal

Here we consider the Heisenberg group $H_n = C^n \times \Re$. $U(p,q)$, $p+q=n$, acts by automorphism on H_n by $g\cdot(z,t) = (gz,t)$.

Let $\{X_1,\dots,X_n,Y_1,\dots,Y_n,T\}$ be the standard basis of the Lie algebra of H_n and let

$$L = \sum_{j=1}^{p}\left(X_j^2+Y_j^2\right) - \sum_{j=p+1}^{n}\left(X_j^2+Y_j^2\right).$$

Via the Plancherel inversion formula, we obtain the joint spectral decomposition of $L^2(H_n)$ with respect to L and T

$$f = \sum_{k\in Z}\int_{-\infty}^{+\infty} f * S_{\lambda,k}\,|\lambda|^n\,d\lambda, \quad f\in S(H_n)$$

where each $S_{\lambda.k}$ is a tempered distribution $U(p,q)$ invariant satisfying $iTS_{\lambda,k} = \lambda S_{\lambda,k}$, $LS_{\lambda,k} = -|\lambda|(2k+p-q)S_{\lambda,k}$. We compute explicitly the distributions $S_{\lambda,k}$ and the integral $\mu_k = \int_{-\infty}^{+\infty} f * S_{\lambda,k}\,|\lambda|^n\,d\lambda$.

1. Introduction.

Let $H_n = C^n \times \Re$ with law group $(z,t)(z',t') = \left(z+z', t+t' - \frac{1}{2}\mathrm{Im}B(z,z')\right)$, where $B(z,w) = \sum_{j=1}^{p} z_j\overline{w_j} - \sum_{j=p+1}^{n} z_j\overline{w_j}$. Then H_n can be viewed as the $2n+1$ dimensional Heisenberg group. Indeed, if $n=p+q$, $Q(z,w) = -\mathrm{Im}B(z,w)$ is the standard symplectic form on $\Re^{2(p+q)}$ via the identification $\Psi : \Re^{2(p+q)} \to C^n$ given by

$$\Psi\left(x',x'',y',y''\right) = \left(x'+iy', x''-iy''\right), \quad x',y'\in\Re^p; x'',y''\in\Re^q. \tag{1.1}$$

Moreover, Ψ provides a global coordinate system (x,y,t) with $x=(x',x'')$, $y=(y',y'')$. The vector fields $X_j = -\frac{1}{2}y_j\frac{\partial}{\partial t} + \frac{\partial}{\partial x_j}$, $Y_j = \frac{1}{2}x_j\frac{\partial}{\partial t} + \frac{\partial}{\partial y_j}$, $j=1,\dots,n$ and $T=\frac{\partial}{\partial t}$ form a basis for the Lie algebra h_n of H_n. As usual, $\mathcal{U}(h_n)$ will denote its universal enveloping algebra, which can be identified with the algebra of left invariant differential operators on H_n.

$U(p,q) = \{g \in GL(n,\mathbb{C}) : B(gz,gw) = B(z,w)\}$ acts by automorphism on H_n by

$$g\cdot(z,t) = (gz,t), \qquad g \in U(p,q), (z,t) \in H_n. \tag{1.2}$$

It is well known that the subalgebra $\mathcal{U}(h_n)^{U(n)}$ of the elements which commute with the action of $U(n) = U(n,0)$ given by (1.2), is generated by T and the Heisenberg Laplacian $\sum\limits_{j=1}^{n}\left(X_j^2 + Y_j^2\right)$. The spherical functions associated with the Gelfand pair $(U(n), H_n)$ have been obtained independently by many authors (see e.g., [**H-R**], [**Ko**], [**St**]). Moreover in [**B-J-R**] it is developed a general calculus to provide the bounded K- spherical functions for a Gelfand pair (K, H_n), $K \subset U(n)$.

For general $p, q, p+q = n$, let

$$L = \sum_{j=1}^{p}\left(X_j^2 + Y_j^2\right) - \sum_{j=p+1}^{n}\left(X_j^2 + Y_j^2\right).$$

Then

$$L = \left(\sum_{j=1}^{p}\left(x_j^2 + y_j^2\right) - \sum_{j=p+1}^{n}\left(x_j^2 + y_j^2\right)\right)\frac{\partial^2}{\partial t^2} + \sum_{j=1}^{p}\left(\frac{\partial^2}{\partial {x_j}^2} + \frac{\partial^2}{\partial {y_j}^2}\right) - \sum_{j=p+1}^{n}\left(\frac{\partial^2}{\partial {x_j}^2} + \frac{\partial^2}{\partial {y_j}^2}\right) + \frac{\partial}{\partial t}\sum_{j=1}^{n}\left(x_j\frac{\partial}{\partial y_j} - y_j\frac{\partial}{\partial x_j}\right). \tag{1.3}$$

It is easy to see, reasoning as in the case $p = n, q = 0$, (see Lemma 2.1 below), that the subalgebra $\mathcal{U}(h_n)^{U(p,q)}$, of the left invariant differential operators which commute with the action of $U(p,q)$ is generated by T and L. So, it is natural to ask for the joint eigendistributions of L and T and the associated decomposition of $L^2(H_n)$. In order to do this, we will use, following [**St**], the Plancherel inversion formula to decompose $f \in S(H_n)$ as

$$f = \sum_{k\in Z}\int_{-\infty}^{+\infty} f * S_{\lambda,k}\,|\lambda|^n\,d\lambda$$

where each $S_{\lambda,k}$ is a tempered and $U(p,q)$ invariant distribution satisfying $iTS_{\lambda,k} = \lambda S_{\lambda,k}$, $LS_{\lambda,k} = -|\lambda|(2k+p-q)S_{\lambda,k}$.

Next we will study the confluent hypergeometric equation in a suitable distribution space in order to obtain that, for $k \geq q$

$$\langle S_{\lambda,k}, f\rangle = c\sum_{j=0}^{n-2} c_j(\lambda)\int\limits_{\Re} e^{-i\lambda t}\delta_B^j(f(.,t))\,dt + c\int\limits_{C^n\times\Re} e^{-i\lambda t}e^{-\frac{|\lambda|}{4}B(z)}L_{k-q}^{n-1}\left(\frac{|\lambda|}{2}B(z)\right)H(|\lambda|B(z))\,f(z,t)\,dz\,dt$$

where $B(z) = B(z,z)$, H is the Heaviside function, δ_B^j are canonical distributions associated to the quadratic form B defined as in [**G-Sh**], supported on $\{z \in C^n : B(z) = 0\}$ and where L_{k-q}^{n-1} denotes, as usual, a Laguerre polynomial. The various constants $c, c_j(\lambda)$ are explicitly computed. Similar formulas are obtained if $k \leq -p$. If $-p < k < q$, $S_{\lambda,k}$ is written as a finite sum in terms of the distributions δ_B^j, $j = 1, ..., n-2$. Finally, we compute $\mu_k = \int\limits_{\Re} S_{\lambda,k} |\lambda|^n d\lambda$ and so the projections $\wp_k f = f * \mu_k$, $k \in Z$. In particular we recover the projections onto the kernel of $L + i(2k + p - q)T$, extending the formula given in [**M-R,2**] for $n = 2$, $p = q = 1$, to arbitrary n, p, q.

Acknowledgments. We express our thanks to Fulvio Ricci who introduced and guides us in this beautiful subject and to Jorge Vargas for many useful conversations.

2. Some preliminaries.

As in the case $p = n, q = 0$ we have that $\mathcal{U}(h_n)^{U(p,q)}$ is generated by T and L and the proof follows the same lines but we add it for the sake of completeness.

Lemma 2.1. *$\mathcal{U}(h_n)^{U(p,q)}$ is generated by T and L.*

Proof. Let $S(h_n)$ be the symmetric algebra generated by the set

$$\{X_1, \ldots, X_n, Y_1, \ldots, Y_n, T\}$$

and let $\Lambda : S(h_n) \to \mathcal{U}(h_n)$ be the symmetrizer map. Since $U(p,q)$ acts on $S(h_n)$ and on $\mathcal{U}(h_n)$ by automorphism, the following diagram is commutative (see [**V**], Th. 3.3.4)

$$\begin{array}{ccc} S(h_n) & \xrightarrow{\Lambda} & \mathcal{U}(h_n) \\ \downarrow g & & \downarrow g \\ S(h_n) & \xrightarrow{\Lambda} & \mathcal{U}(h_n) \end{array}, \qquad g \in U(p,q).$$

Λ is a linear isomorphism, thus Λ maps $S(h_n)^{U(p,q)}$ onto $\mathcal{U}(h_n)^{U(p,q)}$. Since the action of $U(p,q)$ preserves degree on $S(h_n)$, the lines of Theorem 3.3.8 in [**V**] say that if $\{1, u_1, \ldots, u_m\}$ is a set of generators of $S(h_n)^{U(p,q)}$, then $\{1, \Lambda(u_1), \ldots, \Lambda(u_m)\}$ generates $\mathcal{U}(h_n)^{U(p,q)}$. If $u \in S(h_n)^{U(p,q)}$ then $u = \sum P_j(X_1, \ldots, X_n, Y_1, \ldots, Y_n,) T^j$ where the sum is finite and each P_j is a polynomial $U(p,q)$ invariant. Decomposing P_j as a sum of homogeneous polynomials, the same is true for all of them. Since $SU(p,q)$ acts transitively on

$$S_1 = \left\{ (x,y) \in \Re^{2n} : \sum_{j=1}^{p} \left(x_j^2 + y_j^2\right) - \sum_{j=p+1}^{n} \left(x_j^2 + y_j^2\right) = 1 \right\}$$

each P_j must be a polynomial in $\sum\limits_{j=1}^{p}\left(x_j^2+y_j^2\right)-\sum\limits_{j=p+1}^{n}\left(x_j^2+y_j^2\right)$. This ends the proof. $\square$

We recall that for $\lambda \in \Re$ $\lambda \neq 0$, the Schrödinger's representation π_λ of the Heisenberg group $\Re^n \times \Re^n \times \Re$ is defined on $L^2\left(\Re^n\right)$ by

$$\pi_\lambda\left(x,y,t\right)h\left(\zeta\right)=e^{-i\left(\lambda t+sg(\lambda)\sqrt{|\lambda|}x\cdot\zeta+\frac{1}{2}\lambda x\cdot y\right)}h\left(\zeta+\sqrt{|\lambda|}y\right). \tag{2.1}$$

We denote by $E_\lambda\left(h_1,h_2\right)$ the matrix entry associated to π_λ and the vectors h_1, h_2, given by

$$E_\lambda\left(h_1,h_2\right)\left(x,y,t\right)=\left\langle \pi_\lambda\left(x,y,t\right)h_1,h_2\right\rangle.$$

We also denote by $d\pi_\lambda$ the infinitesimal representation defined on the space of C^∞ vectors for π_λ, which is, in this case, the space of the rapidly decreasing functions

$$d\pi_\lambda\left(X\right)h=\frac{d}{dt}_{|t=0}\pi_\lambda\left(\exp tX\right)h.$$

We still denote by π_λ the corresponding representation of $H_n = C^n \times \Re$ and by $E_\lambda\left(h_1,h_2\right), d\pi_\lambda$ its associated matrix entries and infinitesimal representation respectively.

It is remarked in [**St**] that

$$XE_\lambda\left(h_1,h_2\right)=E_\lambda\left(d\pi_\lambda\left(X\right)h_1,h_2\right),\quad X\in\mathcal{U}\left(h_n\right).$$

It follows that $iTE_\lambda = \lambda E_\lambda$ and that, in order to obtain matrix entries eigenfuntions of L, we must look for eigenvectors of $d\pi_\lambda\left(L\right)$ in $L^2\left(\Re^n\right)$.

Thus we pick the orthonormal basis of $L^2\left(\Re^n\right)$ given by the Hermite functions: For $\alpha=\left(\alpha_1,\ldots,\alpha_n\right)\in\left(N\cup\{0\}\right)^n$, let

$$h_\alpha\left(\zeta\right)=\left(2^{|\alpha|}\alpha!\sqrt{\pi}\right)^{-\frac{n}{2}}e^{-\frac{|\zeta|^2}{2}}\prod_{j=1}^{n}H_{\alpha_j}\left(\zeta_j\right)$$

with $|\alpha|=\alpha_1+\cdots+\alpha_n$, $\alpha!=\alpha_1!\ldots\alpha_n!$ and where

$$H_k\left(s\right)=\left(-1\right)^k e^{s^2}\frac{d^k}{ds^k}\left(e^{-s^2}\right)$$

is the $k-th$ Hermite polynomial.

It follows from (2.1) that

$$d\pi_\lambda\left(L\right)=-\left|\lambda\right|\left(B\left(\zeta\right)-\left(\sum_{j=1}^{p}\frac{\partial^2}{\partial\zeta_j^2}-\sum_{j=p+1}^{n}\frac{\partial^2}{\partial\zeta_j^2}\right)\right)$$

where $B\left(\zeta\right)=\sum\limits_{j=1}^{p}\zeta_j^2-\sum\limits_{j=p+1}^{n}\zeta_j^2$.

For $\alpha = (\alpha_1, \dots, \alpha_n)$ we set $\|\alpha\| = \sum_{j=1}^{p} \alpha_j - \sum_{j=p+1}^{n} \alpha_j$. Since $\left(\zeta_j^2 - \frac{\partial^2}{\partial \zeta_j^2}\right) h_{\alpha_j}$ $= (2\alpha_j + 1) h_{\alpha_j}$, we have that $d\pi_\lambda (L) h_\alpha = - |\lambda| (2 \|\alpha\| + p - q) h_\alpha$. Thus

$$d\pi_\lambda (L) E_\lambda (h_\alpha, h_\alpha) = - |\lambda| (2 \|\alpha\| + p - q) E_\lambda (h_\alpha, h_\alpha). \tag{2.2}$$

(2.2) and the Plancherel inversion formula lead us to the joint spectral resolution of iT and L.

The inversion formula asserts that, for $f \in S(H_n)$

$$f(x,y,t) = \frac{1}{(2\pi)^{n+1}} \int_{-\infty}^{+\infty} tr\left(\pi_\lambda (f) \pi_\lambda (x,y,t)\right) |\lambda|^n \, d\lambda$$

where $\pi_\lambda (f) = \int_{H_n} f(x,y,t) \pi_\lambda (x,y,t)^{-1} \, dxdydt$. Moreover, for $f \in S(H_n)$, $(x,y,t) \in H_n$, we have that

$$\sum_\alpha \int_{-\infty}^{+\infty} |\langle \pi_\lambda (x,y,t) \pi_\lambda (f) h_\alpha, h_\alpha \rangle| \, |\lambda|^n \, d\lambda \leq M < \infty \tag{2.3}$$

with M independent of (x,y,t) (see [**R**], Th. 10.1).

Taking account of that

$$\langle \pi_\lambda (x,y,t) \pi_\lambda (f) h_\alpha, h_\alpha \rangle = (E_\lambda (h_\alpha, h_\alpha) * f)(x,y,t)$$

and that

$$E_\lambda (h_\alpha, h_\alpha) \left((x,y,t)^{-1}\right) = \overline{E_\lambda (h_\alpha, h_\alpha)(x,y,t)}$$

we have

$$\begin{aligned} f(x,y,t) &= \frac{1}{(2\pi)^{n+1}} \int_{-\infty}^{+\infty} \sum_\alpha \langle \pi_\lambda (x,y,t) \pi_\lambda (f) h_\alpha, h_\alpha \rangle |\lambda|^n \, d\lambda \\ &= \frac{1}{(2\pi)^{n+1}} \sum_\alpha \int_{-\infty}^{+\infty} (f * E_\lambda (h_\alpha, h_\alpha))(x,y,t) |\lambda|^n \, d\lambda \\ &= \frac{1}{(2\pi)^{n+1}} \sum_{k \in Z} \int_{-\infty}^{+\infty} \sum_{\|\alpha\| = k} (f * E_\lambda (h_\alpha, h_\alpha))(x,y,t) |\lambda|^n \, d\lambda. \end{aligned}$$

Lemma 2.2. *Let $\mu_k : S(H_n) \to C$ be defined by*

$$\mu_k (f) = \int_{-\infty}^{+\infty} \sum_{\|\alpha\| = k} \langle f, E_\lambda (h_\alpha, h_\alpha) \rangle |\lambda|^n \, d\lambda, \qquad f \in S(H_n).$$

Then $\mu_k \in S'(H_n)$.

Proof. For $k \in Z$, let H_k be the closed subspace of $L^2(\Re^n)$ generated by $\{h_\alpha : \|\alpha\| = k\}$, thus $L^2(\Re^n) = \bigoplus_{k \in Z} H_k$. Let P_k be the orthogonal projection from $L^2(\Re^n)$ onto H_k. Now, for $f \in S(H_n)$, we define $\wp_k f$ by

$$\pi_\lambda (\wp_k f) = P_k \pi_\lambda (f). \tag{2.4}$$

It follows from (2.3) that

$$\int_{-\infty}^{+\infty} \sum_{\alpha} \left|\left\langle \pi_\lambda\left(\wp_k f\right) \pi_\lambda\left(x,y,t\right) h_\alpha, h_\alpha\right\rangle\right| \left|\lambda\right|^n d\lambda < \infty$$

and so

$$\wp_k f\left(x,y,t\right) = \frac{1}{\left(2\pi\right)^{n+1}} \int_{-\infty}^{+\infty} \sum_{\|\alpha\|=k} \left(f * E_\lambda\left(h_\alpha,h_\alpha\right)\right)\left(x,y,t\right) \left|\lambda\right|^n d\lambda.$$

$\wp_k f$ commutes with left translations and by (2.4) and the Plancherel formula it extends to a bounded operator on $L^2\left(H_n\right)$. So, there exists a unique tempered distribution, which is μ_k such that $\wp_k f = f * \mu_k$. □

We set, for $\lambda \in \Re - \{0\}$ and $f \in S\left(H_n\right)$

$$S_{\lambda,k}\left(f\right) = \sum_{\|\alpha\|=k} \left\langle f, E_\lambda\left(h_\alpha,h_\alpha\right)\right\rangle. \tag{2.5}$$

We claim that $S_{\lambda,k}$ is well defined and belongs to $S'\left(H_n\right)$. In order to see this, we consider $\overline{H_n} = H_n/N$ where $N = \{0\}\times\{0\}\times 2\pi Z$. Then $\overline{H_n} = \Re^n \times \Re^n \times S^1$, where $S^1 = \left\{e^{i\theta} : \theta \in \Re\right\}$. Each irreducible unitary representation of $\overline{H_n}$ is unitarily equivalent to one and only one of the following representations: The representations π_m acting on $L^2\left(\Re^n\right)$ corresponding to $\lambda = 2\pi m$, $m \in Z$ and the one dimensional representations $\sigma_{a,b}\left(x,y,t\right) = e^{i(ax+by)}$, $(a,b) \in \Re^n \times \Re^n$. For f nice enough, $\pi_m\left(f\right)$ is a Hilbert Schmidt operator. We have also $\sigma_{a,b}\left(f\right) = \int_{\Re^n\times\Re^n\times S^1} f\left(x,y,t\right) e^{-i(ax+by)} dxdydt = \widehat{f}\left(a,b,\overline{0}\right)$, where $\widehat{f}$ denotes the euclidean Fourier transform and $\overline{0}$ is the identity in N. The Plancherel identity asserts that

$$\|f\|^2_{L^2\left(\overline{H_n}\right)} = \sum_{m\neq 0} \|\pi_m\left(f\right)\|^2_{HS} \left|m\right|^n + \int_{\Re^n\times\Re^n} \left|\sigma_{a,b}\left(f\right)\right|^2 dadb.$$

Also, setting $\phi\left(a,b\right) = \sigma_{a,b}\left(f\right)$, the inversion formula is in this case

$$f\left(x,y,t\right) = \sum_{m\neq 0} tr\left(\pi_m\left(f\right)\pi_m\left(x,y,t\right)^{-1}\right)\left|m\right|^n + \widehat{\phi}\left(-x,-y\right).$$

So we can consider $L, T = \frac{\partial}{\partial\theta}$ and $\wp_k$ as above, and repeat all the arguments for $\overline{H_n}$ instead of H_n to obtain that $\nu_k\left(f\right) = \sum\limits_{m\neq 0} \left|m\right|^n \sum\limits_{\|\alpha\|=k} \left\langle f, E_m\left(h_\alpha,h_\alpha\right)\right\rangle$ defines a tempered distribution on $S\left(\Re^n\times\Re^n\times S^1\right)$. Furthermore, the analogous of (2.3) says that the last double series converges absolutely. Now, for $\lambda \in \Re - \{0\}$, $(z,t) \in C^n \times \Re$, we can write (see, for example [Fo]), $E_\lambda\left(h_\alpha,h_\alpha\right)\left(z,t\right)$ in terms of Laguerre polynomials as

$$E_\lambda\left(h_\alpha,h_\alpha\right)\left(z,t\right) = e^{-i\lambda t} e^{-\frac{1}{4}|\lambda||z|^2} \prod_{j=1}^{n} L^0_{\alpha_j}\left(\frac{1}{2}\left|\lambda\right|\left|z_j\right|^2\right). \tag{2.6}$$

For $f \in S\left(\Re^{2n}\right)$, we set $\nu_{k,l}(f) = \nu_k(g_l(f))$, where $g_l(f)(z,t) = e^{ilt} f(z)$, $(z,t) \in C^n \times \Re$ and where we use the identification of C^n with $\Re^{2n}$ given by (1.1). Then $\nu_{k,l} \in S'\left(\Re^{2n}\right)$ if $l \in Z - \{0\}$. In particular, we have that the series

$$e^{-\frac{1}{4}|z|^2} \sum_{\|\alpha\|=k} \prod_{j=1}^{n} L^0_{\alpha_j}\left(\frac{1}{2}|z_j|^2\right) \tag{2.7}$$

defines an element in $S'\left(\Re^{2n}\right)$ and so $S_{1,k} \in S'(H_n)$.

We set, for $\mu \in S'(H_n)$, $\lambda \in \Re - \{0\}$

$$\langle \delta_\lambda \mu, f \rangle = |\lambda|^{-n-1} \langle \mu, \delta_{\lambda^{-1}} f \rangle \tag{2.8}$$

where $\delta_\lambda f(z,t) = f\left(\sqrt{|\lambda|}z, \lambda t\right)$.

Lemma 2.3. *$S_{\lambda,k} \in S'(H_n)$ for all $\lambda \in \Re - \{0\}, k \in Z$.*

Proof. $S_{\lambda,k} = \delta_\lambda(S_{1,k})$ and $S_{1,k} \in S'(H_n)$. □

Remark 2.4. Since the series (2.7) belongs to $S'\left(\Re^{2n}\right)$, the same dilation argument shows that the series $e^{-\frac{1}{4}|\lambda||z|^2} \sum\limits_{\|\alpha\|=k} \prod\limits_{j=1}^{n} L^0_{\alpha_j}\left(\frac{1}{2}|\lambda||z_j|^2\right)$ defines a tempered distribution $F_{\lambda,k}$ on $\Re^{2n}$ for $\lambda \in \Re - \{0\}, k \in Z$.

For $g \in U(p,q)$, let $S^g_{\lambda,k}$ be defined by $S^g_{\lambda,k}(f) = S_{\lambda,k}(f^g)$, where $f^g(z,t) = f(gz,t)$. We have

Lemma 2.5. *$S_{\lambda,k}$ is a $U(p,q)$ invariant distribution for all $\lambda \in \Re - \{0\}, k \in Z$.*

Proof. Let w be the metaplectic representation of $SU(p,q)$ on $L^2(\Re^n)$. Then, for $g \in SU(p,q)$, $(z,t) \in H_n$, we have that

$$\pi_\lambda(gz,t) = w(g)\,\pi_\lambda(z,t)\,w\left(g^{-1}\right). \tag{2.9}$$

Furthermore, $L^2(\Re^n) = \bigoplus\limits_{k\in Z} H_k$, where H_k is, as in Lemma 2.2, the closed subspace generated by $\{h_\alpha : \|\alpha\| = k\}$. It is known that (w, H_k) is $SU(p,q)$ irreducible (see 1.12, 2.7 and 2.8, Ch.VIII in [**B-W**]).

We denote by $I_k : H_k \to L^2(\Re^n)$ the inclusion map and by $P_k : L^2(\Re^n) \to H_k$ the orthogonal projection. We also set $T_{z,t} = P_k \pi_\lambda(z,t) I_k$. Then, for $f \in S(H_n)$, the operator $T = \int_{H_n} f(z,t)\, T_{z,t}\, dzdt$ is a trace class operator. Now, by (2.9)

$$\begin{aligned}\left\langle S^g_{\lambda,k}, f\right\rangle &= \sum_{\|\alpha\|=k} \int_{H_n} f(z,t)\,\langle \pi_\lambda(gz,t)\, h_\alpha, h_\alpha\rangle\, dzdt \\ &= \sum_{\|\alpha\|=k} \int_{H_n} f(z,t)\,\left\langle \pi_\lambda(z,t)\, w\left(g^{-1}\right) h_\alpha, w\left(g^{-1}\right) h_\alpha\right\rangle dzdt\end{aligned}$$

$$= \sum_{\beta} \langle T\theta_\beta, \theta_\beta \rangle = \langle S_{\lambda,k}, f \rangle$$

with $\theta_\beta = w\left(g^{-1}\right) h_\beta$ and where we use that $\{\theta_\beta\}_\beta$ is another orthonormal basis of H_k. Then $S_{\lambda,k}$ is $SU(p,q)$ invariant. Finally, we note also that if $g = z_0 I$, $|z_0| = 1$, I the $n \times n$ identity matrix, it is clear from (2.6) that $S^g_{\lambda,k} = S_{\lambda,k}$ and so $S_{\lambda,k}$ is a $U(p,q)$ invariant distribution. $\square$

Remark 2.6. By the inversion Plancherel formula and Lemmas (2.2), (2.3) and (2.5) we have $f = \sum\limits_{k \in Z} \int_{-\infty}^{+\infty} f * S_{\lambda,k} |\lambda|^n \, d\lambda$, $f \in S(H_n)$.

Let $F_{\lambda,k} \in S'\left(\Re^{2n}\right)$ be the distribution defined in Remark 2.4. Since $F_{\lambda,k} \bigotimes 1 = e^{i\lambda t} S_{\lambda,k}$ we have that $F_{\lambda,k}$ is $U(p,q)$ invariant. Then

$$\sum_{j=1}^{n} \left(x_j \frac{\partial}{\partial y_j} - y_j \frac{\partial}{\partial x_j} \right) F_{\lambda,k} = 0.$$

From $LS_{\lambda,k} = -|\lambda|(2k+p-q) S_{\lambda,k}$ and (1.3) we have that

$$\left(-\frac{1}{4}\lambda^2 B(z) + \square \right) F_{\lambda,k} = -|\lambda|(2k+p-q) F_{\lambda,k} \tag{2.10}$$

where $\square = \sum\limits_{j=1}^{p} \left(\frac{\partial^2}{\partial x_j{}^2} + \frac{\partial^2}{\partial y_j{}^2} \right) - \sum\limits_{j=p+1}^{n} \left(\frac{\partial^2}{\partial x_j{}^2} + \frac{\partial^2}{\partial y_j{}^2} \right)$ and $B(z) = B(z,z)$ for $z = x + iy$, $x, y \in \Re^n$.

Now, according with [**T**], the space of the $U(p,q)$ invariant tempered distributions can be described as the dual of the space of the functions in $C^\infty(\Re - \{0\})$ with some kind of singularity at the origin. In order to describe them, we introduce polar coordinates on $\Re^{2n}$ as follows. For $x, y \in \Re^n$ we set $\sigma = \sum\limits_{j=1}^{p} \left(x_j^2 + y_j^2 \right) - \sum\limits_{j=p+1}^{n} \left(x_j^2 + y_j^2 \right)$, $\rho = \sum\limits_{j=1}^{n} \left(x_j^2 + y_j^2 \right)$, $u = \left(\frac{\rho+\sigma}{2}\right)^{\frac{1}{2}} w_u$, $v = \left(\frac{\rho-\sigma}{2}\right)^{\frac{1}{2}} w_v$ where w_u belongs to the $2p-1$ dimensional sphere S^{2p-1} and $w_v \in S^{2q-1}$.

For $f \in S\left(\Re^{2n}\right)$ and for $\rho, \sigma \in \Re$, $\rho \geq \sigma$, $\rho \geq 0$, let

$$(Mf)(\rho,\sigma) = \int\limits_{S^{2p-1} \times S^{2q-1}} f\left(\left(\frac{\rho+\sigma}{2}\right)^{\frac{1}{2}} w_u, \left(\frac{\rho-\sigma}{2}\right)^{\frac{1}{2}} w_v \right) dw_u dw_v$$

and let, for $\tau \in \Re$,

$$(Nf)(\tau) = \int\limits_{\rho > |\tau|} (Mf)(\rho,\tau)(\rho+\tau)^{p-1}(\rho-\tau)^{q-1} \, d\rho. \tag{2.11}$$

We note that

$$\int_{\Re^{2n}} f(x)\,dx = \frac{1}{2^n}\int_{\Re} Nf(\sigma)\,d\sigma. \tag{2.12}$$

Let H be the Heaviside function, defined by $H(\tau) = 1$ if $\tau \geq 0$ and $H(\tau) = 0$ if $\tau < 0$. Let $\mathcal{H}_0$ the space of the functions $\varphi : \Re \to C$ such that $\varphi(\tau) = \varphi_1(\tau) + H(\tau)\varphi_2(\tau)\tau^{n-1}$, $\varphi_1, \varphi_2 \in D(\Re)$, where $D(\Re)$ denotes the space of the functions in $C^\infty(\Re)$ with compact support and let $\mathcal{H}$ be the space defined analogously, but where now we require $\varphi_1, \varphi_2 \in S(\Re)$.

If $\varphi \in \mathcal{H}$, then it is regular out of the origin and $\varphi \in C^{n-2}(\Re)$. Moreover, for each $m \geq n-1$, there exists $P_m(\varphi)$, polynomial of degree m, such that $\varphi - HP_m(\varphi) \in C^m(\Re)$. So, for $m \in N$, φ admits an expansion

$$\varphi(\tau) = \sum_{j=0}^{m} B_j(\varphi)\tau^j + H(\tau)\sum_{j=0}^{m} A_j(\varphi)\tau^j + o(\tau^m) \tag{2.13}$$

with $A_j(\varphi) = 0$ for $j < n-1$.

Remark 2.7. $\mathcal{H}_0$ and $\mathcal{H}$, with the topology given in [**T**], are Frechet spaces and $N : S(\Re^{2n}) \to \mathcal{H}$, $N : D(\Re^{2n}) \to \mathcal{H}_0$ are linear, continuous and surjective maps. Moreover, their adjoints $N' : \mathcal{H}' \to S'(\Re^{2n})^{U(p,q)}$, $N' : \mathcal{H}_0' \to D'(\Re^{2n})^{U(p,q)}$ are linear homeomorphisms. (see 2.1, 4.3, 5.1 and some remarks at the beginning of §7 in [**T**]). (We also remark that 5.1 in [**T**] holds for $U(p,q)$ instead of $SO(p,q)$ with the obvious changes.)

It is also proved in [**T**] that

$$N(\Box f) = D(Nf), f \in S(\Re^{2n}) \tag{2.14}$$

where the differential operator D is defined by

$$D = 4\left(\tau\frac{\partial^2}{\partial\tau^2} + (2-n)\frac{\partial}{\partial\tau}\right) \tag{2.15}$$

so the adjoint of D is given by $D'T = 4(\tau T'' + nT')$, $T \in \mathcal{H}'$.

We say that $T \in \mathcal{H}'$ is a solution of $D'T = 0$ if $\langle D'T, \varphi\rangle = 0$ for all $\varphi \in \mathcal{H}$. It is easy to see that $T \in \mathcal{H}'$ is a solution of

$$\frac{\lambda^2}{4}\tau T + 4(\tau T'' + nT') = -|\lambda|(2k+p-q)T \tag{2.16}$$

if and only if $N'T$ is a solution of (2.10). The same assertion is true for solutions in $\mathcal{H}_0'$.

Setting $b = -|\lambda|(2k+p-q)$, (2.16) becomes $16\tau T'' + 16nT' - (\lambda^2\tau + 4b)T = 0$. As in [**Ko**], we note that if $\beta = \pm\frac{\lambda}{4}$, $\frac{\beta}{\alpha} = -\frac{1}{2}$ and $l = \frac{4n\beta - b}{4\alpha}$ and if $w(t) = e^{\beta t}v(\alpha t)$, then w is a solution of $16\tau w'' + 16nw' -$

$\left(\lambda^2\tau+4b\right)w=0$ if and only if v is a solution of the confluent hypergeometric equation (C.H.E) $tv''+(n-t)\,v'+lv=0$.

For $T\in\mathcal{H}'$ and for $k\in Z$, $\lambda\in\Re-\{0\}$ we set

$$\langle T_{\lambda,k},\varphi\rangle=\left\langle\delta_{\frac{|\lambda|}{2}}T,\psi_\lambda(\varphi)\right\rangle,\psi_\lambda(\varphi)(t)=e^{-\frac{|\lambda|}{4}t}\varphi(t) \tag{2.17}$$

for $k\geq 0$, where $\delta_\lambda\varphi(t)=\varphi(\lambda t)$ and $\langle\delta_\lambda T,\varphi\rangle=|\lambda|^{-1}\langle T,\delta_{\lambda^{-1}}\varphi\rangle$.

We also set

$$\langle T_{\lambda,k},\varphi\rangle=\left\langle\delta_{-\frac{|\lambda|}{2}}T,\psi_\lambda(\varphi)\right\rangle,\psi_\lambda(\varphi)(t)=e^{\frac{|\lambda|}{4}t}\varphi(t) \tag{2.18}$$

if $k<0$.

We note that if $k\geq 0$ then $T\in\mathcal{H}_0'$ is a solution of the C.H.E. with parameter $l=k-q$ if and only if $T_{\lambda,k}$ is a solution in $\mathcal{H}_0'$ of (2.16). If $k<0$ then $T\in\mathcal{H}_0'$ solves the C.H.E. with parameter $l=-k-p$ if and only if $T_{\lambda,k}$ solves (2.16).

Our aim is to find all the solutions in $\mathcal{H}'$ of (2.16). We note that if S is such a solution, then $S=T_{\lambda,k}$ for some solution $T\in\mathcal{H}_0'$ of the C.H.E. with parameter $l=k-q$ if $k\geq 0$ and $l=-k-p$ if $k<0$. This leads us to determine all the solutions in $\mathcal{H}_0'$ of C.H.E. with parameter $l\geq -n+1$ such that the corresponding $T_{\lambda,k}\in\mathcal{H}'$.

3. About the confluent hypergeometric equation.

As in [**Sz**], if m, β are non negative integers, we denote by $\{L_m^\beta\}$, the Laguerre polynomials. Then $L_m^\beta(x)$ is defined as the only polynomial solution of

$$tv''+(\beta+1-t)v'+mv=0$$

and normalized by the condition

$$\int_0^\infty e^{-x}x^\beta L_m^\beta(x)L_{m'}^\beta(x)dx=\Gamma(\beta+1)\binom{m+\beta}{m}\delta_{m,m'}. \tag{3.1}$$

We have that

$$L_m^0(t)=\sum_{j=0}^m\binom{m}{j}(-1)^j\,\frac{x^j}{j!} \tag{3.2}$$

and that $\frac{d}{dt}L_m^\beta=-L_{m-1}^{\beta+1}$.

Let D_l be the differential operator on $\mathcal{H}$ given by

$$D_l\varphi(\tau)=\tau\varphi''+(2-n)\varphi'+\tau\varphi'+(l+1)\varphi. \tag{3.3}$$

Then its adjoint D_l' is $D_l'T=tT''+(n-t)T'+lT$. We recall that $A_j(\varphi)=0$ for $\varphi\in\mathcal{H}$, $j\leq n-2$. It is easy to see that if φ admits an asymptotic development

$$\sum_{j\geq 0}B_j(\varphi)\,\tau^j+H\sum_{j\geq 0}A_j(\varphi)\,\tau^j$$

then the expansion around $\tau = 0$ of $D_l\varphi$ is

$$\sum_{j\geq 0}\left[(l+1+j)B_j(\varphi)+(j+1)(j+2-n)B_{j+1}(\varphi)\right]\tau^j + H\sum_{j\geq 0}\left[(l+1+j)A_j(\varphi)+(j+1)(j+2-n)A_{j+1}(\varphi)\right]\tau^j. \tag{3.4}$$

With the natural restrictions on f, integration by parts gives

$$\int_a^b f(t)\left(D_l\varphi\right)(t)\,dt = \int_a^b \left(D_l'f\right)(t)\varphi(t)dt + R(b,\varphi) - R(a,\varphi) \tag{3.5}$$

where $-\infty \leq a < b \leq +\infty$ and

$$R(b,\varphi) = (1-n+b)f(b)\varphi(b) + bf(b)\varphi'(b) - bf'(b)\varphi(b). \tag{3.6}$$

Proposition 3.1. *For $l \geq 0$, $T = \left(L^0_{l+n-1}H\right)^{(n-1)}$ is a solution in $\mathcal{H}_0'$ of $D_l'T = 0$.*

Proof. Let $c_{j,l} = \left(L^0_{l+n-1}\right)^{(n-2-j)}(0)$, $0 \leq j \leq n-2$. Then a computation shows that

$$T = \left(L^0_{l+n-1}\right)^{(n-1)}H + \sum_{j=0}^{n-2} c_{j,l}\delta^{(j)}$$

and so $T \in \mathcal{H}'$ since every $\varphi \in \mathcal{H}$ is in $C^{n-2}(\Re)$. Also

$$\begin{aligned}\left\langle D_l'T,\varphi\right\rangle &= \langle T, D_l\varphi\rangle \\ &= \int_0^\infty \left(L^0_{l+n-1}\right)^{(n-1)}(t)\left(D_l\varphi\right)(t)\,dt + \left\langle \sum_{j=0}^{n-2} c_{j,l}\delta^{(j)}, D_l\varphi\right\rangle.\end{aligned}$$

By (3.4), (3.5) and (3.6) we have

$$\int_0^\infty \left(L^0_{l+n-1}\right)^{(n-1)}(t)\left(D_l\varphi\right)(t)\,dt = (n-1)\left(L^0_{l+n-1}\right)^{(n-1)}(0)\,B_0(\varphi)$$

and by (3.4)

$$\begin{aligned}&\left\langle \sum_{j=0}^{n-2} c_{j,l}\delta^{(j)}, D_l\varphi\right\rangle \\ &= \sum_{j=0}^{n-2} c_{j,l}(-1)^j j!B_j\left(D_l\varphi\right) \\ &= \sum_{j=0}^{n-2} c_{j,l}(-1)^j j!\left((l+1+j)B_j\varphi + (j+1)(j+2-n)B_{j+1}(\varphi)\right) \\ &= \sum_{j=0}^{n-2} d_{j,l}B_j(\varphi)\end{aligned}$$

where $d_{0,l} = (l+1)\, c_{0,l}$ and $d_{j,l} = (-1)^j\, j!\, ((l+1+j)\, c_{j,l} + (n-j-1)\, c_{j-1,l})$ if $1 \le j \le n-2$. Since $c_{j,l} = (-1)^{n-j} \binom{l+n-1}{n-j-2}$ the lemma follows. □

Now, it is proved in [**T**] that if $S \in \mathcal{H}'$ and $\operatorname{supp}(S) = \{0\}$ then there exists $m_1, m_2 \in N \cup \{0\}$ $\alpha_0, \ldots \alpha_{m_1}, \alpha'_0, \ldots \alpha'_{m_2} \in C$ such that

$$S(\varphi) = \sum_{j=0}^{m_1} \alpha_j B_j(\varphi) + \sum_{j=0}^{m_2} \alpha'_j A_j(\varphi), \quad \varphi \in \mathcal{H}.$$

We will need the following:

Lemma 3.2. *Assume $l \ge -n+1$. If $S \in \mathcal{H}'$, $\operatorname{supp} S = \{0\}$ and if*

$$D'_l S = c_{n-1} B_{n-1} + d_{n-1} A_{n-1} + \sum_{j=0}^{n-2} c_j B_j$$

with $c_0, \ldots, c_{n-1}, d_{n-1} \in C$, then $c_{n-1} = d_{n-1} = 0$.

Proof. We write $S = \sum_{j=0}^{m_1} \alpha_j B_j + \sum_{j=0}^{m_2} \alpha'_j A_j$. Suppose $c_{n-1} \ne 0$. By (3.4) the coefficient of $B_j(\varphi)$ in the expansion of $D_l(\varphi)$ is $(l+1+j)\, \alpha_j$ $+j\,(j+1-n)\, \alpha_{j-1}$ and so $c_{n-1} = (l+n)\, \alpha_{n-1}$ and $\alpha_j = -\frac{j(j+1-n)}{l+1+j} \alpha_{j-1}$ for $j \ge -l$. Then $\alpha_j \ne 0$ if $j \ge n$. Contradiction. Analogously $d_{n-1} \ne 0$ would imply $\alpha'_j \ne 0$ for $j \ge n$. □

If $l \ge 0$, a solution of the C.H.E. is the function $f_1(t) = L_l^{n-1}(t)$. Another solution $f_2 \in C^2((-\infty, 0))$ of the C.H.E., linearly independent with f_1, is obtained setting $f_2(t) = c(t) f_1(t)$ where $c(t)$ satisfy

$$t f_1(t) c''(t) + \left[2t f'_1(t) + (n-t) f_1(t)\right] c'(t) = 0.$$

Then for $t < 0$,

$$f_2(t) = f_1(t) \int_{-\infty}^{t} f_1(s)^{-2} s^{-n} e^s ds \tag{3.7}$$

is well defined since the zeros of the Laguerre's polynomials are in $(0, +\infty)$.

Also

$$\left.\begin{aligned} &f_2(t) \underset{t\to-\infty}{=} o\left(e^t\right), \\ &f'_2(t) \underset{t\to-\infty}{=} o\left(e^t\right), \\ &f_2(t) \backsim -\tfrac{1}{f_1(0)(n-1)} t^{-n+1} \quad \text{as} \quad t \to 0. \end{aligned}\right\} \tag{3.8}$$

Lemma 3.3. *Let for* $\varphi \in \mathcal{H}$,

$$\langle Pf(f_2), \varphi \rangle = \lim_{\epsilon \to 0^+} \int_{-\infty}^{-\epsilon} f_2(t) \left(\varphi(t) - \sum_{j=0}^{n-2} \frac{\varphi^{(j)}(0)}{j!} t^j \right) dt.$$

Then $Pf(f_2) \in \mathcal{H}'$ *and* $D_l' Pf(f_2) = -\frac{1}{f_1(0)} B_{n-1}(\varphi)$.

Proof. $Pf(f_2) \in \mathcal{H}'$ by Lemma 3.3 in [**T**]. On the other hand, from (3.4) it follows that if $\psi(t) = \sum_{j=0}^{n-2} \frac{\varphi^{(j)}(0)}{j!} t^j$ then $D_l \psi = \sum_{j=0}^{n-2} \frac{(D_l\varphi)^{(j)}(0)}{j!} t^j$. Thus

$$\begin{aligned} \langle D_l' Pf(f_2), \varphi \rangle &= \langle Pf(f_2), D_l' \varphi \rangle \\ &= \lim_{\epsilon \to 0^+} \int_{-\infty}^{-\epsilon} f_2(t) \left((D_l\varphi)(t) - \sum_{j=0}^{n-2} \frac{(D_l\varphi)^{(j)}(0)}{j!} t^j \right) dt \\ &= \lim_{\epsilon \to 0^+} \int_{-\infty}^{-\epsilon} f_2(t) D_l(\varphi - \psi)(t)\, dt = \lim_{\epsilon \to 0^+} R(-\epsilon, \varphi_1) \end{aligned}$$

where $\varphi_1 = \varphi - \psi$ and $R(-\epsilon, \varphi_1)$ is given by (3.6).

As by (3.8)

$$\lim_{s \to 0^-} (1 - n + s) f_2(s) \varphi_1(s) = (1 - n) \frac{1}{f_1(0)(1 - n)} B_{n-1}(\varphi),$$

$$\lim_{s \to 0^-} s f_2(s) \varphi_1'(s) = \frac{1}{f_1(0)} \lim_{s \to 0^-} \frac{s^{-n+2}}{1-n} \left((n-1) B_{n-1} s^{n-2} + \ldots . \right)$$

$$= -\frac{1}{f_1(0)} B_{n-1}$$

and

$$\lim_{s \to 0^-} s f_2'(s) \varphi_1(s) = \frac{1}{f_1(0)} B_{n-1}$$

the lemma follows. □

Proposition 3.4. *Let T be in $\mathcal{H}_0'$. Suppose that either $k \geq q$ or $k \leq -p$ and $\lambda \in \Re - \{0\}$, let $T_{\lambda,k}$ be defined as in* (2.17) *and* (2.18). *If $T_{\lambda,k}$ is a tempered solution (i.e., $T_{\lambda,k} \in \mathcal{H}'$) of* (2.16) *then T is a multiple of $\left(L_{l+n-1}^0 H \right)^{(n-1)}$ where $l = k - q$ if $k \geq q$ and $l = -k - p$ if $k \leq -p$.*

Proof. We know that there exists a basis of the solution space in $C^2(0, +\infty)$ given by $f_1(t)$ and a certain function $g(t)$ where $g(t) \backsim e^t$ as $t \to +\infty$ [**Se**]. In particular when we write T restricted to $(0, +\infty)$, as a linear combination $af_1 + bg$, the condition $T_{\lambda,k} \in \mathcal{H}'$ implies $b = 0$.

We now consider $S = T - a\left(L^0_{l+n-1}H\right)^{(n-1)}$. Then supp$S \subset (-\infty, 0]$, $D'_l S = 0$ and the corresponding $S_{\lambda,k} \in \mathcal{H}'$.

Writing S restricted to $(-\infty, 0)$ as a linear combination $\alpha f_1 + \beta f_2$ we obtain that $\alpha = 0$. Thus $S - \beta Pf\,(f_2)$ has support at $t = 0$ and by Lemma 3.3

$$D'_l\left(S - \beta Pf\,(f_2)\right) = -\beta \frac{1}{f_1\,(0)} B_{n-1}.$$

If $\beta \neq 0$, this contradicts Lemma 3.2. Thus supp$S = \{0\}$. But, from (3.4), it is easy to see that there is not nontrivial solution S supported at the origin of $D'_l S = 0$ if $l \geq 0$. So $S = 0$ and the proof is complete. □

To state a similar result for $-p < k < q$ we will need some facts about the equation

$$tv'' + (n - t)\,v' - lv, \quad l = 1, \dots, n-1. \tag{3.9}$$

Lemma 3.5. *For $l = 1, \dots, n-1$ there exists a polynomial P_{l-1} of degree $l-1$ with $P_{l-1}\,(0) = 1$ such that for all open interval $I \subset \Re - \{0\}$ (not necessarily finite) two linearly independent solutions in $C^2\,(I)$ are given by $g_1\,(t) = t^{1-n} P_{l-1}\,(t)\,e^t$ and $g_2\,(t) = t^{1-n} T_{n-2}\left(P_{l-1}\,(t)\,e^t\right)$ where $T_{n-2}\,(g)$ denotes the Taylor polynomial of degree $n-2$ around the origin for the function g.*

Proof. Following the notation of [**Se**], we can write every solution of (3.9) belonging to $C^2\,(I)$ as $\alpha.{}_1F_1\,(l, n, t) + \beta t^{1-n}.{}_1F_1\,(1 + l - n, 2 - n, t)$ where

$${}_1F_1\,(a, c, t) = \sum_{j=0}^{\infty} \frac{(a)_j}{(c)_j} \frac{t^j}{j!} \tag{3.10}$$

and $(a)_j = a\,(a + 1) \dots (a + j - 1)$.

By (3.10) ${}_1F_1\,(1 + l - n, 2 - n, t) = \sum\limits_{j=0}^{\infty} p_{l-1}\,(j)\,\frac{t^j}{j!}$ where $p_{l-1}\,(j) = \sum\limits_{k=0}^{l-1} a_k j^k$ for some $a_1, \dots, a_{k-1} \in \Re$ and $a_0 = 1$. Induction on k shows that $\sum\limits_{j=0}^{\infty} j^k \frac{t^j}{j!} = q_k\,(t)\,e^t$ with q_k a polynomial of degree k such that $q_k\,(0) = 0$ for $k > 0$. So $g_1\,(t) = t^{1-n}.{}_1F_1\,(1 + l - n, 2 - n, t)$ is a solution of the desired form.

Also

$$\begin{aligned}
&{}_1F_1\,(l, n, t)\\
&= \sum_{j=0}^{\infty} \frac{(l)_j}{(n)_j} \frac{t^j}{j!} = \frac{(n-1)!}{(l-1)!} \sum_{j=0}^{\infty} \frac{(j+1)\dots(j+l-1)}{(n+j-1)!} t^j\\
&= \frac{(n-1)!}{(l-1)!} \sum_{j=0}^{\infty} \frac{(j + (n-1) + (2-n)) \dots ((j + n - 1) + (l - n))}{(n+j-1)!} t^j
\end{aligned}$$

$$= \frac{(n-1)!}{(l-1)!} \frac{1}{t^{n-1}} \sum_{j=n-1}^{\infty} (j+2-n) \dots (j+l-n) \frac{t^j}{j!}$$
$$= \frac{(n-1)!}{(l-1)!} (2-n) \dots (l-n)$$
$$\cdot \frac{1}{t^{n-1}} \left({}_1F_1 (1+l-n, 2-n, t) - T_{n-2} \left({}_1F_1 (1+l-n, 2-n, t) \right) \right).$$

So we can take $g_2(t) = t^{1-n} T_{n-2} \left({}_1F_1 (1+l-n, 2-n, t) \right)$. □

Lemma 3.6. *For $\varphi \in \mathcal{H}$, let $Pf^-(g_1)$ and $Pf^+(g_2)$ be defined by*

$$\left\langle Pf^-(g_1), \varphi \right\rangle = \lim_{\epsilon \to 0^+} \int_{-\infty}^{-\epsilon} g_1(t) \left(\varphi(t) - \sum_{j=0}^{n-2} \frac{\varphi^{(j)}(0)}{j!} t^j \right) dt,$$
$$\left\langle Pf^+(g_2), \varphi \right\rangle = \lim_{\epsilon \to 0^+} \int_{\epsilon}^{1} g_2(t) \left(\varphi(t) - \sum_{j=0}^{n-2} \frac{\varphi^{(j)}(0)}{j!} t^j \right) dt$$
$$+ \int_1^{\infty} g_2(t) \varphi(t) \, dt.$$

Then $Pf^-(g_1)$ and $Pf^+(g_2)$ belong to $\mathcal{H}'$ and they satisfy:

(i) $D_l' (Pf^-(g_1)) = (n-1) B_{n-1}$,

(ii) $D_l' (Pf^+(g_2)) = -(n-1)(B_{n-1} + A_{n-1}) + \sum_{j=0}^{n-2} \beta_j B_j$ *for some constants* $\beta_1, \dots, \beta_{n-2}$.

Proof. The proof follows similar lines those of Lemma 3.3, but now, to prove (i) we take account of that $P_{l-1}(0) = 1$ where P_{l-1} is as in Lemma 3.5. For (ii) we observe that if $\varphi \in \mathcal{H}$ and if $\psi(t) = \sum_{j=0}^{n-2} B_j(\varphi) t^j$, we have $R(1, \varphi - \psi) - R(1, \varphi) = -(2-n)\psi(1) - \psi'(1) f_2(1) + f_2'(1) \psi(1)$.

The constants β_j are determined by $f_2(1)$ and $f_2'(1)$. □

Lemma 3.7. *For each $l = -1, -2, \dots, -n+1$, the space of the solutions $T \in \mathcal{H}_0'$ which are supported at the origin of the equation $D_l' T = 0$ is one dimensional.*

Proof. For such a T we write $T = \sum_{j=0}^{m_1} \alpha_j B_j + \sum_{j=n-1}^{m_2} \alpha_j' A_j$. From $\langle T, D_l \varphi \rangle = 0$ and (3.4) we obtain that $\alpha_j (l+1+j) + \alpha_{j-1} (j+1-n) = 0$ for all j. If $j = n-1$, this implies that $\alpha_{n-1}(l+n) = 0$ and so $\alpha_j = 0$ for all $j \geq n-1$. The same argument says that $\alpha_j' = 0$, $j \geq n-1$ and thus $T = \sum_{j=0}^{n-2} \alpha_j B_j$. Let

$j_0 = -l - 1$. Then $\alpha_{j_0 - 1} = 0$. . Since

$$\alpha_j = -\frac{j+1-n}{l+1+j}\alpha_{j-1} \tag{3.11}$$

for $j \neq j_0$ we have $\alpha_0 = \alpha_1 = \cdots = \alpha_{j_0-1} = 0$. So T is completely determined by α_{j_0}. On the other hand, it is clear that for each α_{j_0} we obtain in this way a solution supported at $\{0\}$. □

Remark 3.8. Let l, T be as in Lemma 3.7. If we write $T = \sum_{j=0}^{n-2} \gamma_{j,l}\delta^{(j)}$ instead of $\sum_{j=0}^{n-2} \alpha_j B_j$, by (3.11) we see that $\{\gamma_{j,l}\}$ satisfy

$$(l+1+j)\,\gamma_{j,l} + (n-j-1)\,\gamma_{j-1,l} = 0$$

for $0 \leq j \leq n-2$. But this is also the recurrence relation for the successive derivatives at the origin of the polynomial L^0_{l+n-1}, so we can choose a nontrivial solution as $T_0 = \sum_{j=0}^{n-2} \gamma_{j,l}\delta^{(j)}$ with $\gamma_{j,l} = \left(L^0_{l+n-1}\right)^{(n-j-2)}(0)$, $0 \leq j \leq n-2$. Now, a computation shows that $T_0 = \left(L^0_{l+n-1}H\right)^{(n-1)}$.

Proposition 3.9. *Let T be in $\mathcal{H}'_0$. Suppose $-p < k < q$, $\lambda \in \Re - \{0\}$, let $T_{\lambda,k}$ be defined as in* (2.17) *and* (2.18). *If $T_{\lambda,k}$ is a tempered solution* (*i.e., $T_{\lambda,k} \in \mathcal{H}'$*) *of* (2.16) *then T is a multiple of the distribution T_0 defined in Remark* 3.8.

Proof. We argue as in Proposition 3.4. Suppose $0 \leq k < q$. So $T_{\lambda,k}$ is given by (2.17). Now, $T_{\lambda,k} \in \mathcal{H}'$ implies that T restricted to $(0, +\infty)$ agrees with αg_2 and T restricted to $(-\infty.0)$ agrees with βg_1, for some $\alpha, \beta \in C$ and where g_1, g_2 are defined as in Lemma 3.5. So $S = T - \beta P f^-(g_1) - \alpha P f^+(g_2)$ has support at the origin and, by Lemma 3.6, it satisfies $D'_l(S) = -\beta(n-1)B_{n-1} + \alpha(n-1)(B_{n-1} + A_{n-1}) + \sum_{j=0}^{n-2}\beta_j B_j$. But, by Lemma 3.2 $\alpha = \beta = 0$ and so T has support at the origin and the lemma follows from Lemma 3.7. The case $-p < k < 0$ is analogous. □

4. Determination of $S_{\lambda,k}$ and $\wp_k$.

In this section we compute explicitly the distributions $S_{\lambda,k}$ and μ_k. Taking account of Remark 3.8 and Proposition 3.1, we consider the particular distribution T given by $T = \left(L^0_{l+n-1}H\right)^{(n-1)}$ where $l = k - q$ if $k \geq 0$ and $l = -k - p$ if $k < 0$. Let $F_{\lambda,k} \in S'\left(\Re^{2n}\right)$ be defined as in Remark 2.4. Since $F_{\lambda,k} \in S'(H_n)^{U(p,q)}$ and satisfies (2.10), the considerations in Remark 2.7 and Propositions 3.4 and 3.9 imply that $F_{\lambda,k} = c_{\lambda,k}N'(T_{\lambda,k})$ for some $c_{\lambda,k} \in C$. In order to compute $c_{\lambda,k}$ we apply both distributions to the function

(4.1)

$$f_\lambda(z) = f_\lambda(z_1, \dots z_n) = e^{-\frac{|\lambda|}{4}|z|^2} \sum_{\substack{\beta_1+\cdots+\beta_n=|k|,\\ \beta_1\geq 0,\dots\beta_n\geq 0}} \prod_{j=1}^{n} L^0_{\beta_1}\left(\frac{1}{2}|\lambda||z_j|^2\right).$$

By (3.1) we have that, if $k \geq 0$

$$\langle F_{\lambda,k}, f_\lambda\rangle = 2^n\pi^n|\lambda|^{-n} \sum_{\substack{\beta_1+\cdots+\beta_p=|k|,\\ \beta_1\geq 0,\dots\beta_p\geq 0}} 1 = 2^n\pi^n|\lambda|^{-n}\binom{p+k-1}{p-1} \tag{4.2}$$

and if $k < 0$

(4.3)

$$\langle F_{\lambda,k}, f_\lambda\rangle = 2^n\pi^n|\lambda|^{-n} \sum_{\substack{\beta_1+\cdots+\beta_q=|k|,\\ \beta_1\geq 0,\dots\beta_q\geq 0}} 1 = 2^n\pi^n|\lambda|^{-n}\binom{q-k-1}{q-1}.$$

On the other hand, by well known properties of the Laguerre polynomials,

$$f_\lambda(z) = e^{-\frac{|\lambda|}{4}|z|^2} L^{n-1}_{|k|}\left(\frac{1}{2}|\lambda||z|^2\right). \tag{4.4}$$

So, for $t \geq 0$, and taking account of that the volume of the n dimensional sphere is $2\pi^{\frac{n+1}{2}}/\Gamma\left(\frac{n+1}{2}\right)$, we have

(4.5)

$$\begin{aligned} &Nf_\lambda\left(2|\lambda|^{-1}t\right)\\ &= \frac{4\pi^{p+q}}{(p-1)!\,(q-1)!}\int_{2|\lambda|^{-1}t}^{\infty} e^{-\frac{|\lambda|}{4}\rho} L^{n-1}_{|k|}\left(\frac{|\lambda|\rho}{2}\right)\\ &\qquad\cdot\left(\rho+2|\lambda|^{-1}t\right)^{p-1}\left(\rho-2|\lambda|^{-1}t\right)^{q-1}d\rho\\ &= \frac{4\pi^{p+q}}{(p-1)!\,(q-1)!}2^{n-1}|\lambda|^{-(n-1)}\int_t^\infty e^{-\frac{s}{2}} L^{n-1}_{|k|}(s)\,(s+t)^{p-1}(s-t)^{q-1}\,ds. \end{aligned}$$

Now,

$$\langle F_{\lambda,k}, f_\lambda\rangle = c_{\lambda,k}\left\langle N'(T_{\lambda,k}), f_\lambda\right\rangle = c_{\lambda,k}\langle T_{\lambda,k}, N(f_\lambda)\rangle.$$

From (4.5), the definition of $T_{\lambda,k}$ and (4.2) we obtain that $c_{\lambda,k}$ is independent of λ. In order to compute $c_{\lambda,k}$ we consider first the case $k \geq 0$. By (2.17)

$$\langle T_{\lambda,k}, N(f_\lambda)\rangle = \left\langle 2|\lambda|^{-1}\delta_{\frac{|\lambda|}{2}}T, t\to e^{-\frac{|\lambda|}{4}t}N(f_\lambda)(t)\right\rangle$$

$$= 2\left|\lambda\right|^{-1}\left\langle T, t \to e^{-\frac{t}{2}} N\left(f_\lambda\right)\left(2\left|\lambda\right|^{-1} t\right)\right\rangle$$

thus, by (4.5), we need to evaluate $T\left(\psi_0\right)$ where $T = \left(L^0_{k-q+n-1} H\right)^{(n-1)}$ and $\psi_0\left(t\right) = e^{-\frac{t}{2}}\varphi_0\left(t\right)$ with

$$\varphi_0\left(t\right) = e^{-\frac{t}{2}} \int_0^\infty e^{-\frac{\rho}{2}} L^{n-1}_k\left(\rho + t\right)\left(\rho + 2t\right)^{p-1}\rho^{q-1} d\rho.$$

Since $k-q+n-1 = k+p-1$ and $L^{n-1}_k\left(\rho + t\right)\left(\rho + 2t\right)^{p-1}$ is a polynomial in t of degree $k+p-1$ we can use the Leibnitz formula for the derivatives of a product, the fact that every polynomial can be written as a linear combination of the Laguerre polynomials and the orthogonality relations (3.1) to obtain that

$$\begin{aligned} &T\left(\psi_0\right) \\ &= (-1)^{n-1}\int_0^\infty L^0_{k+p-1}\left(t\right)\int_0^\infty e^{-\frac{\rho}{2}}\rho^{q-1} e^{-t} L^{n-1}_k\left(\rho+t\right)\left(\rho + 2t\right)^{p-1} d\rho dt. \end{aligned}$$

Since $L^{n-1}_k\left(\rho + t\right) = \sum\limits_{m+j=k} L^{n-2}_m\left(\rho\right) L^0_j\left(t\right)$, we repeat the same argument to obtain that

$$\begin{aligned} &T\left(\psi_0\right) \\ &= 2^{p-1}(-1)^{n-1}\int_0^\infty L^0_{k+p-1}\left(t\right)\left[\int_0^\infty e^{-\frac{\rho}{2}}\rho^{q-1} L^{n-2}_0\left(0\right) d\rho\right] e^{-t} L^0_k\left(t\right) t^{p-1} dt \\ &= (-1)^{n-1} 2^{p-1}(-1)^q 2^q (q-1)! \int_0^\infty e^{-t} L^0_{k+p-1}\left(t\right)\frac{(-1)^k}{k!} t^{k+p-1} dt \\ &= (-1)^{n+q-1} 2^{n-1}(q-1)! \frac{(-1)^k}{k!}(-1)^{k+p-1}(k+p-1)! \end{aligned}$$

where we have used (3.1) and (3.2).

Finally, by (4.2), we find that

$$2^n \pi^n \frac{(p+k-1)!}{k!\,(p-1)!} = c_{\lambda,k} 2^n \frac{4\pi^n}{(p-1)!\,(q-1)!} 2^{n-1}\frac{(k+p-1)!}{k!}(q-1)!$$

and so

$$c_{\lambda,k} = \frac{1}{2^{n+1}}.$$

If $k < 0$, we can repeat the above computation, using (2.18) instead of (2.17) and replacing $L^0_{k-q+n-1}$ by $L^0_{-k-p+n-1}$. In this case we also find $c_{\lambda,k} = \frac{1}{2^{n+1}}$.

Theorem 4.1. *If* $k \geq q$, $\lambda \in \Re - \{0\}$, $f \in S\left(C^n\right)$, *then*

$$\left\langle F_{\lambda,k}, f\right\rangle = \frac{1}{2}\int\limits_{B(z)\geq 0} e^{-\frac{|\lambda|}{4}B(z)} L^{n-1}_{k-q}\left(\frac{|\lambda|}{2} B\left(z\right)\right) f\left(z\right) dz$$

$$+\frac{1}{2^n}\sum_{l=0}^{n-2}4^l\left|\lambda\right|^{-(l+1)}\sum_{j=l}^{n-2}\frac{1}{2^j}\binom{j}{l}(-1)^{n-j}\binom{n+k-q-1}{k-q+j+1}\left\langle\delta_B^l,f\right\rangle$$

where $\delta_B^l=N'\left(\delta^{(l)}\right)$.

Proof.

$$\begin{aligned}\langle F_{\lambda,k},f\rangle&=\frac{1}{2^{n+1}}\left\langle N'T_{\lambda,k},f\right\rangle=\frac{1}{2^{n+1}}\left\langle T_{\lambda,k},Nf\right\rangle\\&=\frac{1}{2^{n+1}}\left\langle T,t\to 2\left|\lambda\right|^{-1}e^{-\frac{t}{2}}Nf\left(2\left|\lambda\right|^{-1}t\right)\right\rangle.\end{aligned}$$

Now, as at the beginning of the proof of Proposition 3.1,

$$T=L_{k-q}^{n-1}H+\sum_{j=0}^{n-2}\left(L_{k-q+n-1}^{0}\right)^{(n-2-j)}(0)\,\delta^{(j)}.$$

But

$$\begin{aligned}&2\left|\lambda\right|^{-1}\int_0^\infty L_{k-q}^{n-1}(t)\,e^{-\frac{t}{2}}Nf\left(2\left|\lambda\right|^{-1}t\right)dt\\&=\int_0^\infty L_{k-q}^{n-1}\left(\frac{\left|\lambda\right|t}{2}\right)e^{-\frac{\left|\lambda\right|t}{4}}Nf(t)\,dt\\&=2^n\int_{B(z)\geq 0}e^{-\frac{\left|\lambda\right|}{4}B(z)}L_{k-q}^{n-1}\left(\frac{\left|\lambda\right|}{2}B(z)\right)f(z)\,dz\end{aligned}$$

where the last equality follows from (2.12) applied to the function

$$F(z)=L_{k-q}^{n-1}\left(\frac{\left|\lambda\right|B(z)}{2}\right)e^{-\frac{\left|\lambda\right|B(z)}{4}}f(z).$$

On the other hand, a computation shows that

$$\begin{aligned}&\left\langle\sum_{j=0}^{n-2}\left(L_{k-q+n-1}^{0}\right)^{(n-2-j)}(0)\,\delta^{(j)},t\to 2\left|\lambda\right|^{-1}e^{-\frac{t}{2}}Nf\left(2\left|\lambda\right|^{-1}t\right)\right\rangle\\&=2\sum_{l=0}^{n-2}4^l\left|\lambda\right|^{-(l+1)}\sum_{j=l}^{n-2}\binom{j}{l}\left(L_{k-q+n-1}^{0}\right)^{(n-2-j)}(0)\,\frac{1}{2^j}\left\langle\delta_B^l,f\right\rangle\end{aligned}$$

and the theorem follows. □

Remark 4.2. Theorem 4.1 remains true for $k\leq -p$, with the obvious changes in the proof, if we replace L_{k-q}^{n-1} by L_{-k-p}^{n-1}, $\binom{n+k-q-1}{k-q+j+1}$ by $\binom{n-k-p-1}{-k-p+j+1}$

and the integration region $\{z : B(z) \geq 0\}$ by $\{z : B(z) \leq 0\}$. It is also immediate to see that if $-p < k < q$, $\lambda \in \Re - \{0\}$, $f \in S(C^n)$, then

$$\langle F_{\lambda,k}, f \rangle = \frac{1}{2^n} \sum_{l=0}^{n-2} 4^l |\lambda|^{-(l+1)} \sum_{j=l}^{n-2} \frac{1}{2^j} \binom{j}{l} \gamma_{j,k} \left\langle \delta_B^l, f \right\rangle$$

with $\gamma_{j,l}$ as in Remark 3.8, i.e.,

$$\gamma_{j,k} = \left(L^0_{k-q+n-1}\right)^{(n-j-2)}(0) = (-1)^{n-j} \binom{n+k-q-1}{n-j-2}$$

for $q-k-1 \leq j \leq n-2$ and $\gamma_{j,k} = 0$ if $j < q-k-1$ and where δ_B^l is as in Theorem 4.1.

Remark 4.3. We have computed the distributions $F_{\lambda,k}$ and the constant $c_{\lambda,k}$, and so also $S_{\lambda,k} = e^{-i\lambda t} F_{\lambda,k}$.

Next, we compute μ_k. We first assume $k \geq q$. Taking account of Theorem 4.1. We recall that for $f = f(z,t) \in S'(H_n)$

$$\langle \mu_k, f \rangle = \int_{-\infty}^{\infty} \left\langle e^{-i\lambda t} F_{\lambda,k}, f \right\rangle |\lambda|^n \, d\lambda.$$

By Theorem 4.1 $|\lambda|^n e^{-i\lambda t} \langle F_{\lambda,k}, f(.,t) \rangle = J_1(f)(\lambda,t) + J_2(f)(\lambda,t)$, $t \in \Re$, where

$$J_1(f)(\lambda,t) = \frac{1}{2} |\lambda|^n e^{-i\lambda t} \int_{B(z)\geq 0} e^{-\frac{|\lambda|}{4} B(z)} L^{n-1}_{k-q} \left(\frac{|\lambda|}{2} B(z) \right) f(z,t) \, dz$$

and

$$J_2(f)(\lambda,t)$$
$$= \frac{1}{2^n} e^{-i\lambda t} \sum_{l=0}^{n-2} 4^l |\lambda|^{n-(l+1)} \sum_{j=l}^{n-2} \frac{1}{2^j} \binom{j}{l} \left(L^0_{k-q+n-1}\right)^{(n-j-2)}(0) \left\langle \delta_B^l, f(.,t) \right\rangle.$$

So, by well known properties of the Fourier transform on $S'(\Re)$,

$$\int_{\Re} \left(\int_{\Re} J_2(f)(\lambda,t) \, dt \right) d\lambda \tag{4.6}$$
$$= \frac{1}{2^n} \sum_{l=0}^{n-2} 4^l (-i)^{n-l-1} \sum_{j=l}^{n-2} \frac{1}{2^j} \binom{j}{l} \gamma_{j,k} \left\langle \nu_l, \frac{\partial^{n-l-1} f}{\partial t^{n-l-1}} \right\rangle$$

where $\nu_l = \delta_B^l \otimes pv\left(\frac{1}{t}\right)$ if $n-l-1$ is odd and $\nu_l = \delta_B^l \otimes \delta$ if $n-l-1$ is even. Let $I_1(f) = \int_{\Re} \left(\int_{\Re} J_1(f)(\lambda,t) \, dt \right) d\lambda$. The properties of the Fourier transform in $S'(\Re)$ imply that

$$(4.7)\quad I_1(f) = \int\limits_{\Re} \left\langle e^{-\lambda i t} e^{-\frac{|\lambda|}{4}B(z)} L_{k-q}^{n-1}\left(\frac{|\lambda|}{2}B(z)\right) H(B(z)), f \right\rangle |\lambda|^n \, d\lambda$$

$$= i \int\limits_{\Re} \left\langle e^{-\lambda i t} e^{-\frac{|\lambda|}{4}B(z)} L_{k-q}^{n-1}\left(\frac{|\lambda|}{2}B(z)\right) H(B(z)), h \right\rangle |\lambda|^{n-1} \, d\lambda$$

where $h(z,t) = \frac{\partial\left(pv\left(\frac{1}{t}*f\right)\right)}{\partial t}(z,t)$.

Now, following [**St**], we will compute (4.7).

Lemma 4.4. *For* $f \in S(C^n \times \Re)$ *there exists* $\int\limits_{C^n\times\Re} \frac{H(B(z))}{B(z)+it} f(z,t)\, dzdt$ *and*

$$\lim_{\epsilon\to 0} \int\limits_{C^n\times\Re} \frac{H(B(z))}{B(z)+\epsilon+it} f(z,t)\, dzdt = \int\limits_{C^n\times\Re} \frac{H(B(z))}{B(z)+it} f(z,t)\, dzdt.$$

Proof. We write

$$\frac{1}{B(z)+\epsilon+it} = P(t, B(z)+\epsilon) - iQ(t, B(z)+\epsilon)$$

where $P(t,s) = \frac{s}{s^2+t^2}$, $Q(t,s) = \frac{t}{s^2+t^2}$, $t, s \in \Re$. Thus, for $s \in \Re\, \|P(.,s)\|_{L^1(\Re)} = \pi$. So

$$\int\limits_{\Re} |P(t, B(z)+\epsilon) f(z,t)|\, dt \le \pi \|f(z,.)\|_{L^\infty(\Re)}, \quad z \in C^n.$$

Also, for $B(z) \ne 0$, we have

$$\lim_{\epsilon\to 0} (P(., B(z)+\epsilon) * f(z,.))(0) = (P(., B(z)) * f(z,.))(0).$$

Since $\sup\limits_{t\in\Re} |f(z,t)| \in L^1(C^n)$, the dominated convergence theorem implies that $P(t, B(z)) f(z,t) \in L^1(C^n \times \Re)$ and

$$\lim_{\epsilon\to 0} \int\limits_{C^n\times\Re} P(t, B(z)+\epsilon) H(B(z)) f(z,t)\, dzdt$$

$$= \int\limits_{C^n\times\Re} P(t, B(z)) H(B(z)) f(z,t)\, dzdt.$$

On the other hand, let $G_\epsilon(z) = \int\limits_{\Re} Q(t, B(z)+\epsilon) f(z,t)\, dt$. So

$$G_\epsilon(z) = \int\limits_{|t|<1} Q(t, B(z)+\epsilon)\left[f(z,t) - f(z,0)\right] dt$$

$$+\int_{|t|\geq 1} Q\left(t,B\left(z\right)+\epsilon\right)f\left(z,t\right)dt.$$

Now, for $|t|<1$

$$\left|\frac{f\left(z,t\right)-f\left(z,0\right)}{t}\right|=\left|\frac{\partial f}{\partial t}\left(z,\zeta\left(z,t\right)\right)\right|\leq\sup_{|u|<1}\left|\frac{\partial f}{\partial t}\left(z,u\right)\right|.$$

Also

$$\sup_{|t|<1}\left|tQ\left(t,B\left(z\right)+\epsilon\right)\right|\leq 1,\quad \sup_{|t|\geq 1}\left|Q\left(t,B\left(z\right)+\epsilon\right)\right|\leq 1.$$

Thus $|G_\epsilon(z)| \leq \sup_{|u|<1}\left|\frac{\partial f}{\partial t}(z,u)\right| + \|f(z,.)\|_{L^1(\Re-[-1,1])}$. So, as above, we can use the dominated convergence theorem to obtain that $Q\left(t,B\left(z\right)\right)H\left(B\left(z\right)\right)f\left(z,t\right)\in L^1\left(C^n\times\Re\right)$ and

$$\lim_{\epsilon\to 0}\int_{C^n\times\Re} Q\left(t,B\left(z\right)+\epsilon\right)H\left(B\left(z\right)\right)f\left(z,t\right)dzdt$$

$$=\int_{C^n\times\Re} Q\left(t,B\left(z\right)\right)H\left(B\left(z\right)\right)f\left(z,t\right)dzdt.$$

□

Following [**St**], we use the generatrix identity for the Laguerre polynomials

$$\sum_{s=0}^{\infty}L_s^{n-1}\left(t\right)r^s=\left(1-r\right)^{-n}e^{-\frac{r}{1-r}t} \tag{4.8}$$

to obtain, for $\epsilon>0$

$$\int_0^\infty e^{-\epsilon\lambda}e^{-i\lambda t}e^{-\frac{\lambda}{4}B(z)}L_{k-q}^{n-1}\left(\frac{\lambda}{2}B\left(z\right)\right)H\left(B\left(z\right)\right)\lambda^{n-1}d\lambda \tag{4.9}$$

$$=\alpha_k\frac{\left[B\left(z\right)-4\epsilon-4it\right]^{k-q}}{\left[B\left(z\right)+4\epsilon+4it\right]^{k+p}}H\left(B\left(z\right)\right)$$

where

$$\alpha_\kappa=4^n\left(n-1\right)!\binom{p+k-1}{k-q}\left(-1\right)^{k-q}. \tag{4.10}$$

Indeed, by (4.8), we can write, for $|r|<1$, $B\left(z\right)\geq 0, t\in\Re$, $\epsilon>0$

$$\sum_{k=q}^{\infty}r^{k-q}\int_0^\infty e^{-\epsilon\lambda}e^{-i\lambda t}e^{-\frac{\lambda}{4}B(z)}L_{k-q}^{n-1}\left(\frac{\lambda}{2}B\left(z\right)\right)\lambda^{n-1}d\lambda$$

$$=\sum_{s=0}^{\infty}r^s\int_0^\infty e^{-\epsilon\lambda}e^{-i\lambda t}e^{-\frac{\lambda}{4}B(z)}L_s^{n-1}\left(\frac{\lambda}{2}B\left(z\right)\right)\lambda^{n-1}d\lambda$$

$$= (1-r)^{-n}\int_0^\infty \exp\left(-\lambda\left(\frac{B(z)(1+r)+4(\epsilon+it)(1-r)}{4(1-r)}\right)\right)\lambda^{n-1}d\lambda$$
$$= \frac{4^n(n-1)!}{[B(z)+4\epsilon+4it+r(B(z)-4\epsilon-4it)]^n}.$$

Now, we compare the Taylor developments to obtain (4.9).

Write
$$\frac{B(z)-it}{B(z)+it} = \frac{2B(z)}{B(z)+it} - 1.$$

Now, letting $\epsilon \to 0^+$, and taking account of Lemma 4.4, we have

$$\int_0^\infty \left\langle e^{-i\lambda t}e^{-\frac{\lambda}{4}B(z)}L_{k-q}^{n-1}\left(\frac{\lambda}{2}B(z)\right)H(B(z))\lambda^{n-1}, f\right\rangle d\lambda \tag{4.11}$$
$$= \alpha_k \lim_{\epsilon\to 0}\left\langle \frac{[B(z)-4\epsilon-4it]^{k-q}}{[B(z)+4\epsilon+4it]^{k+p}}H(B(z)), f\right\rangle.$$

Now, this limit is

$$\alpha_k \lim_{\epsilon\to 0}\left\langle \left[\frac{2B(z)}{B(z)+4\epsilon+4it} - 1\right]^{k-q}\frac{H(B(z))}{[B(z)+4\epsilon+4it]^n}, f\right\rangle$$
$$= \alpha_k \lim_{\epsilon\to 0}\sum_{l=0}^{k-q}\binom{k-q}{l}(-1)^{k-q-l}2^l\left\langle \frac{B(z)^l H(B(z))}{[B(z)+4\epsilon+4it]^{n+l}}, f\right\rangle$$
$$= \alpha_k \sum_{l=0}^{k-q}\binom{k-q}{l}(-1)^{k-q-l}\frac{2^l(-4i)^{n+l-1}}{(n+l-1)!}\left\langle \frac{B(z)^l H(B(z))}{B(z)+4it}, \frac{\partial^{n+l-1}f}{\partial t^{n+l-1}}\right\rangle.$$

So

$$\int_0^\infty \left\langle e^{-i\lambda t}e^{-\frac{\lambda}{4}B(z)}L_{k-q}^{n-1}\left(\frac{\lambda}{2}B(z)\right)H(B(z))\lambda^{n-1}, f\right\rangle d\lambda \tag{4.12}$$
$$= \alpha_k \sum_{l=0}^{k-q}\beta_{k,l}\left\langle \frac{B(z)^l H(B(z))}{B(z)+4it}, \frac{\partial^{n+l-1}f}{\partial t^{n+l-1}}\right\rangle$$

where

$$\beta_{k,l} = \binom{k-q}{l}(-1)^{k-q-l}\frac{2^l(-4i)^{n+l-1}}{(n+l-1)!}. \tag{4.13}$$

From (4.11) a change of variable gives

$$\int_{-\infty}^0 \left\langle e^{-i\lambda t}e^{-\frac{|\lambda|}{4}B(z)}L_{k-q}^{n-1}\left(\frac{|\lambda|}{2}B(z)\right)H(B(z))|\lambda|^{n-1}, f\right\rangle d\lambda \tag{4.14}$$
$$= \alpha_k \sum_{l=0}^{k-q}\overline{\beta}_{k,l}\left\langle \frac{B(z)^l H(B(z))}{B(z)-4it}, \frac{\partial^{n+l-1}f}{\partial t^{n+l-1}}\right\rangle$$

where, by (4.13), $\overline{\beta}_{k,l} = (-1)^{n+l-1}\,\beta_{k,l}$. So we have:

Theorem 4.5. *For $k \geq q$ and $0 \leq l \leq k-q$, let $\alpha_k, \beta_{k,l}$ defined by* (4.10) *and* (4.13) *respectively. Then we have $\mu_k(f) = I_1(f) + I_2(f)$ where*

$$I_1(f)$$
$$= \frac{i\alpha_k}{2}\sum_{l=0}^{k-q}\beta_{k,l}\left\langle\left(\frac{B(z)^l H(B(z))}{B(z)+4it} + (-1)^{n+l-1}\frac{B(z)^l H(B(z))}{B(z)-4it}\right), \frac{\partial^{n+l}\left(pv\left(\frac{1}{t}*f\right)\right)}{\partial t^{n+l}}\right\rangle$$

and

$$I_2(f)$$
$$= \frac{1}{2^n}\sum_{l=0}^{n-2}4^l\sum_{j=l}^{n-2}(-i)^{n-l-1}\frac{1}{2^j}\binom{j}{l}\left(L^0_{k-q+n-1}\right)^{(n-j-2)}(0)\left\langle\nu_l, \frac{\partial^{n-l-1}f}{\partial t^{n-l-1}}\right\rangle$$

where $\nu_l = \delta^l_B \otimes pv\left(\frac{1}{t}\right)$ if $n-l-1$ is odd and $\nu_l = \delta^l_B \otimes \delta$ if $n-l-1$ is even.

Proof. It follows from (4.12), (4.14), (4.7) and (4.6). ☐

Remark 4.6. If $k \leq -p$, Theorem 4.5 remains true if we replace $k-q$ by $-k-p$ and $H(B(z))$ by $H(-B(z))$ with the same proof, using (2.18) instead of (2.17). If $-p < k < q$ the same arguments give us $\mu_k(f) = I_2(f)$, with

$$I_2(f) = \frac{1}{2^n}\sum_{l=0}^{n-2}4^l\sum_{j=l}^{n-2}\frac{1}{2^j}\binom{j}{l}\gamma_{j,k}\left\langle\nu_l, \frac{\partial^{n-l-1}f}{\partial t^{n-l-1}}\right\rangle$$

where $\gamma_{j,k}$ is defined as in Remark 3.8.

Remark 4.7. Let $A = \{(z,t) \in C^n \times \Re : B(z) = 0\}$. If $f \in S(H_n)$ and $\operatorname{supp}(f) \cap A = \emptyset$ thus $\operatorname{supp}\left(\frac{\partial}{\partial t}\left(p.v.\left(\frac{1}{t}\right)*f\right)\right) \cap A = \emptyset$, then from (4.7) and (4.11) and taking account of that $I_2(f) = 0$, we have

$$\begin{aligned}
\mu_k(f) &= I_1(f)\\
&= i\alpha_k \lim_{\epsilon\to 0}\left\langle\frac{[B(z)-4\epsilon-4it]^{k-q}}{[B(z)+4\epsilon+4it]^{k+p}}H(B(z)), \frac{\partial}{\partial t}\left(p.v.\left(\frac{1}{t}\right)*f\right)\right\rangle\\
&= i\alpha_k\left\langle\frac{[B(z)-4it]^{k-q}}{[B(z)+4it]^{k+p}}H(B(z)), \frac{\partial}{\partial t}\left(p.v.\left(\frac{1}{t}\right)*f\right)\right\rangle\\
&= i\alpha_k\left\langle-\frac{\partial}{\partial t}\left(\frac{[B(z)-4it]^{k-q}}{[B(z)+4it]^{k+p}}\right)H(B(z)), p.v.\left(\frac{1}{t}\right)*f\right\rangle.
\end{aligned}$$

This is an analogous expression to those obtained in [**St**], p. 362.

Remark 4.8. For $\epsilon = \pm 1$, $k \in Z$, we set $R_{k,\epsilon} = \{\epsilon\rho, \rho(2k+p-q) : \rho > 0\}$. The rays $R_{k,\epsilon}$ are closely related to the study of the kernels of the operators $L - i\alpha T$, $\alpha \in C$. In order to describe $\ker(L - i\alpha T)$, with $\alpha \in$

$2Z$ for n even and $\ker(L - i\alpha T)$, with $\alpha \in 1 + 2Z$ for n odd, we define $\wp_k^+, \wp_k^- : L^2(H_n) \to L^2(H_n)$ via the Plancherel inversion formula requiring that for $\lambda \in \Re - \{0\}$, $\pi_\lambda \wp_k^+ = \chi_{(0,\infty)}(\lambda) P_k \pi_\lambda$ and $\pi_\lambda \wp_k^- = \chi_{(-\infty,0)}(\lambda) P_k \pi_\lambda$, where P_k is define as at the beginning of the proof of Lemma 2.2. Thus $\wp_k^+, \wp_k^-$ are orthogonal projections over certain subspaces of $L^2(H_n)$. As in Lemma 2.2 we have $\wp_k^+ f = \int_0^{+\infty} f * S_{\lambda,k} |\lambda|^n d\lambda$, $f \in S(H_n)$ (and the analogous formula for $\wp_k^-$). If m has the same parity than n, we define $k_1(m) = -\frac{1}{2}(m + p - q)$ and $k_2(m) = \frac{1}{2}(m - p + q)$. Thus $k_1(m), k_2(m) \in Z$. We observe that $R\left(\wp_{k_1(m)}^+\right) \subset \ker(L - imT) \cap L^2(H_n)$, where $\ker(L - imT) = \{S \in S'(H_n) : (L - imT) S = 0\}$. In order to see this inclusion, we proceed as follows. As in Lemma 2.2 we construct $\mu_{k_1(m)}^{\pm} \in S'(H_n)$ such that $\wp_{k_1(m)}^{\pm} f = f * \mu_{k_1(m)}^{\pm}$. As there, we have $\left\langle \mu_{k_1(m)}^+, \varphi \right\rangle = \int_0^{+\infty} \langle S_{\lambda.k_1(m)}, \varphi \rangle |\lambda|^n d\lambda$, $\varphi \in S'(H_n)$. Then

$$\begin{aligned}
&\left\langle (L - imT)\left(\mu_{k_1(m)}^+\right), \varphi \right\rangle \\
&= \left\langle \mu_{k_1(m)}^+, (L + imT)(\varphi) \right\rangle \\
&= \int_0^{+\infty} \left\langle S_{\lambda.k_1(m)}, (L + imT)(\varphi) \right\rangle |\lambda|^n d\lambda \\
&= \int_0^{+\infty} \left\langle (L - imT) S_{\lambda.k_1(m)}, \varphi \right\rangle |\lambda|^n d\lambda = 0.
\end{aligned}$$

Now, since L, T commute with left translations $(L - imT)\left(f * \mu_{k_1(m)}^+\right) = f * \left((L - imT)\mu_{k_1(m)}^+\right) = 0$. So $R\left(\wp_{k_1(m)}^+\right) \subset \ker(L - imT) \cap L^2(H_n)$. Similarly, $R\left(\wp_{k_2(m)}^-\right) \subset \ker(L - imT) \cap L^2(H_n)$. So $R\left(\wp_{k_1(m)}^+\right) \oplus R\left(\wp_{k_2(m)}^-\right) \subset \ker(L - imT) \cap L^2(H_n)$. On the other hand, Plancherel theorem implies that $R\left(\wp_k^{\pm}\right) \perp R\left(\wp_s^{\pm}\right)$ if $k \neq s$ and $R\left(\wp_k^+\right) \perp R\left(\wp_k^-\right)$, $k \in \dot{Z}$. We know also that, as operator on $L^2(H_n)$, iLT^{-1} has a closed and self-adjoint extension (see [**M-R,1**], Th. 7.4) that we still denote by iLT^{-1}. We have $\ker(L - i\alpha T) \cap L^2(H_n) = \ker\left(LT^{-1} - i\alpha\right)$, $\alpha \in C$ (see [**M-R,2**], Proposition 1.4). Since iLT^{-1} is a self adjoint operator, we have $\ker\left(LT^{-1} - im\right) \perp \ker\left(LT^{-1} - i\widetilde{m}\right)$ for $m \neq \widetilde{m}$. Now, $L^2(H_n) = \bigoplus_{k \in Z} R(\wp_k)$. Thus we have the direct orthogonal sum

$$L^2(H_n) = \bigoplus_{m \in Z} \left(R\left(\wp_{k_1(m)}^+\right) \bigoplus R\left(\wp_{k_2(m)}^-\right) \right).$$

Then we conclude that

$$\ker(L - imT) \cap L^2(H_n) = R\left(\wp_{k_1(m)}^+\right) \bigoplus R\left(\wp_{k_2(m)}^-\right)$$

and that if n is even then $\ker(L - i\alpha T) \cap L^2(H_n) = 0$ if $\alpha \notin 2Z$ and that if n is odd then $\ker(L - i\alpha T) \cap L^2(H_n) = 0$ if $\alpha \notin 1 + 2Z$.

The projectors $\wp_k^{\pm}$, $k \in Z$ can be computed proceeding as in the determination of $\wp_k$. As in Lemma 2.2 we construct $\mu_k^{\pm} \in S'(H_n)$ such that $\wp_k^{\pm} f = f * \mu_k^{\pm}$, and then, with the same arguments used for μ_k, we decompose $\mu_k^{+}(f) = I_1^{+}(f) + I_2^{+}(f)$, where

$$I_1^{+}(f) = \int_0^{\infty} \left\langle e^{-\lambda i t} e^{-\frac{\lambda}{4} B(z)} L_{k-q}^{n-1}\left(\frac{\lambda}{2} B(z)\right) H(B(z)), f \right\rangle \lambda^n d\lambda$$

and

$$I_2^{+}(f) = \int_{\Re}\int_{\Re} \frac{1}{2^n} e^{-i\lambda t} H(\lambda) \sum_{l=0}^{n-2} 4^l \lambda^{-(l+1)} \cdot \sum_{j=l}^{n-2} \frac{(-1)^{n-j}}{2^j} \binom{j}{l} \binom{n+l-1}{l+j+1} \left\langle \delta_B^l, f(.,t) \right\rangle dt d\lambda$$

thus, using the properties of the Fourier transform and taking account of that $\widehat{H} = \delta - ip.v.\left(\frac{1}{t}\right)$ we can obtain explicit formulas for μ_k^{+} of similar type those given for μ_k. Since $\mu_k^{-} = \mu_k - \mu_k^{+}$ we obtain also an explicit description for μ_k^{-}.

References

[B-J-R] C. Benson, J. Jenkins and G. Ratcliff, *Bounded k-spherical function on Heisenberg groups*, J. of. Func. Analysis, **105** (1992), 409-443.

[B-W] A. Borel and N. Wallach, *Continuous cohomology, discrete subgroups, and representatios of reductive groups*, Annals of Mathematics Studies, **94**, Princeton University Press.

[F] G. Folland, *Harmonic Analysis in Phase Space*, Annals of Mathematics Studies, **122**, Princeton University Press.

[G-Sh] I. Gelfand and G. Shilov, *Les Distributions*, 1, Dunod, Paris, 1962.

[H-R] A. Hulanicki and F. Ricci, *A tauberian theorem and tangential convergence of bounded harmonic functions on balls in C^n*, Invent. Math., **62** (1980), 325-331.

[Ko] A. Koranyi, *Some applications of Gelfand pairs in classical analysis*, in 'Harmonic Analysis and Group Representations', CIME, (1980), 333-348.

[M-R,1] D. Müller and F. Ricci, *Analysis of second order differential operators on Heisenberg groups* I, Invent. Math., **101** (1990), 545-582.

[M-R,2] ———, *Analysis of second order differential operators on Heisenberg groups* II, J. of Funct. Analysis, **108**(**2**) (1992), 296-346.

[R] F. Ricci, *Harmonic Analysis on the Heisemberg Group*, Politecnico di Torino.

[Se] J. Seaborn, *Hypergeometric Functions and their applications*, Springer-Verlag.

[St] R. Strichartz, *L^p harmonic analysis and Radon transforms on the Heisenberg group*, J. of Funct. Analyisis, **96** (1991), 350-406.

[Sz] G. Szegö, *Orthogonal Polynomials*, Colloquium Publication, Vol. XXIII, A.M.S.

[T] A. Tengstrand, *Distributions invariant under an orthogonal group of arbitrary signature*, Math. Scand., **8** (1960), 201-218.

[V] V.S. Varadarajan, *Lie groups, Lie algebras and their representations*, Prentice Hall, Series in Modern Analysis.

Received August 11, 1998. The second author was partially supported by CONICET, CONICOR and SECYT-UNC.

Facultad de Matemática, Astronomía y Física
Universidad Nacional de Cordoba
Ciudad Universitaria
5000 Cordoba
Argentina
E-mail address: godoy@mate.uncor.edu

Facultad de Matemática, Astronomía y Física
Universidad Nacional de Cordoba
Ciudad Universitaria
5000 Cordoba
Argentina
E-mail address: saal@mate.uncor.edu

PACIFIC JOURNAL OF MATHEMATICS
Vol. 193, No. 2, 2000

AN ELECTROSTATICS MODEL FOR ZEROS OF GENERAL ORTHOGONAL POLYNOMIALS

Mourad E.H. Ismail

We prove that the zeros of general orthogonal polynomials, subject to certain integrability conditions on their weight functions determine the equilibrium position of movable n unit charges in an external field determined by the weight function. We compute the total energy of the system in terms of the recursion coefficients of the orthonormal polynomials and study its limiting behavior as the number of particles tends to infinity in the case of Freud exponential weights.

1. Introduction.

Stieltjes [**24**], [**25**] considered the following electrostatic model. Fix two charges $(\alpha+1)/2$ and $(\beta+1)/2$ at $x=1$ and $x=-1$, respectively, then put n movable unit charges at distinct points in $(-1,1)$. The question is to determine the equilibrium position of the movable charges when the interaction forces arise from a logarithmic potential. Stieltjes proved that the equilibrium position is attained at the zeros of the Jacobi polynomial $P_n^{(\alpha,\beta)}(x)$. For a proof see Szegő's book [**26**]. Another electrostatic problem is to have a fixed point charge $(\alpha+1)/2$ at $x=0$ and n movable unit point charges at distinct points in $[0,\infty)$. The state of equilibrium in the presence of an additional external potential $v(x)=x$ is now reached at the zeros of the Laguerre polynomial $L_n^{(\alpha)}(x)$ provided that the point charges interact according to a logarithmic potential. Stieltjes [**24**], [**25**] stated evaluations of the discriminants of the classical orthogonal polynomials of Hermite, Laguerre and Jacobi and his results explicitly give the minimum energy for the Stieltjes electrostatic model, which is the energy of the system at the equilibrium position. Later Hilbert [**11**] proved Stieltjes statements and Schur gave a very elegant proof in [**22**]. Recently Forrester and Rogers [**9**] considered similar problems on the unit circle. In a very recent work, Grünbaum [**10**] gave an electrostatic interpretation of the zeros of the Koornwinder-Krall polynomials. The latter polynomials are orthogonal with respect to a measure with an absolutely continuous component supported on $[-1,1]$ and two discrete masses at the end points ± 1.

In this paper we extend the Stieltjes models to general orthogonal polynomials. In the latter part of this section we remind the reader of some definitions, some basic facts, and recent related results. In Section 2 we shall describe the new model, state and prove our first main results concerning this model. We also give an explicit formula for the total energy at the equilibrium position of our model in terms of the recursion coefficients of orthonormal polynomials associated with the model. This is formula (2.17) and we hope it will have some applications to statistical mechanics. In another work, in collaboration with Yang Chen, formula (2.17) will be analyzed further and combined with the Coulomb fluid method of F. Dyson to discuss certain models in statistical mechanics, a continuation of [**7**]. In Section 2 we consider an explicit example worked out to independently verify our results. Section 4 contains a derivation of the limiting behavior of the total energy of the n particle system as $n \to \infty$ in cases of potentials associated with Freud type weights. In Section 3 we comment on electrostatic interpretation for the Freud weights and Selberg-type integrals. We also mention how the differential equation (1.9) gives information on the Bethe Ansatz for general systems of polynomials orthogonal with respect to an absolutely continuous measure supported on an interval. This also relates to systems of nonlinear equations for the zeros such general orthogonal polynomials and extends earlier work of Ahmed, Bruschi, Calegro, Olshantsky and Perelomov [**1**], and Mehta [**18**]. The related and later work of Ahmed and Muldoon [**2**] also contains a detailed bibiolography on the subject.

Let $\{p_n(x)\}$ be polynomials orthonormal with respect to a weight function w supported on $[a, b]$, finite or infinite and $w(x) > 0$ for $x \in (a, b)$. In other words

$$\int_a^b p_m(x)\, p_n(x)\, w(x)\, dx = \delta_{m,n}. \tag{1.1}$$

We associate with $w(x)$ an external potential $v(x)$,

$$w(x) = e^{-v(x)}, \quad x \in (a, b). \tag{1.2}$$

We shall normalize w by $\int_a^b w(x)\, dx = 1$. The initial values and three term recurrence relation of $\{p_n(x)\}$ take the form

$$p_0(x) = 1, \quad p_1(x) = (x - b_0)/a_1, \tag{1.3}$$

$$x p_n(x) = a_{n+1} p_{n+1}(x) + b_n p_n(x) + a_n p_{n-1}(x), \quad n > 0. \tag{1.4}$$

Assuming v is twice differentiable and convex function on $[a,b]$ we define $A_n(x)$ and $B_n(x)$ via

$$\begin{aligned} A_n(x) &= \frac{a_n\, w(b^-)\, p_n^2(b)}{b-x} + \frac{a_n\, w(a^+)\, p_n^2(a)}{x-a} \\ &\quad + a_n \int_a^b \frac{v'(x)-v'(y)}{x-y}\, p_n^2(y)\, w(y)\, dy, \end{aligned} \tag{1.5}$$

$$\begin{aligned} B_n(x) &= \frac{a_n\, w(a^+) p_n(a) p_{n-1}(a)}{x-a} + \frac{a_n\, w(b^-)\, p_n(b)\, p_{n-1}(b)}{b-x} \\ &\quad + a_n \int_a^b \frac{v'(x)-v'(y)}{x-y}\, p_n(y) p_{n-1}(y)\, w(y)\, dy. \end{aligned} \tag{1.6}$$

In (1.5) and (1.6) it is assumed that

$$y^n \frac{v'(x)-v'(y)}{x-y}\, w(y), \quad n = 0, 1, \dots, \tag{1.7}$$

are integrable over (a,b) and the boundary terms in (1.5) and (1.6) exist. Under the latter assumptions the orthonormal polynomials p_n's satisfy the differential recurrence relation [**3**], [**4**], [**5**],

$$p_n'(x) = A_n(x) p_{n-1}(x) - B_n(x) p_n(x), \tag{1.8}$$

and the differential second order equation

$$p_n''(x) + R_n(x) p_n'(x) + S_n(x) p_n(x) = 0, \tag{1.9}$$

where

$$R_n(x) := -\left[v'(x) + \frac{A_n'(x)}{A_n(x)} \right], \tag{1.10}$$

$$\begin{aligned} S_n(x) &:= B_n'(x) - B_n(x) \frac{A_n'(x)}{A_n(x)} - B_n(x)[v'(x) + B_n(x)] \\ &\quad + \frac{a_n}{a_{n-1}} A_n(x) A_{n-1}(x). \end{aligned} \tag{1.11}$$

The representations (1.5) and (1.6) of the A_n's and B_n's are in [**5**] but earlier versions are in [**4**] and [**3**].

It is important to note that (1.9) follows from (1.8) and, (1.3) and (1.4), so (1.9) with R_n and S_n given by (1.10) and (1.11) may hold for polynomials orthogonal with respect to a measure with discrete part. Indeed this is the case, for example, for the Koornwinder polynomials in [**14**] and [**10**]. All we need is to eliminate v' in (1.10) and (1.11) by using

$$B_n(x) + B_{n+1}(x) = \frac{x-b_n}{a_n} A_n(x) - v'(x). \tag{1.12}$$

Observe that v' when obtained via (1.12) may depend on n and as such can be thought of as a varying weight. It is clear from (1.3) and (1.4) that

$$p_n(x) = \frac{x^n}{a_1 a_2 \dots a_n} + \text{lower order terms.} \tag{1.13}$$

The discriminant D_n of a polynomial g_n,

$$g_n(x) := \gamma x^n + \text{lower order terms,} \tag{1.14}$$

is defined by

$$D_n = D(g_n) := \gamma^{2n-2} \prod_{1\le j<k\le n} (x_j - x_k)^2, \tag{1.15}$$

where $x_1, x_2, \dots, x_n$ are the zeros of g_n, see [**8**].

Although we shall not use potential theory in this work, the reader may be interested in consulting Lubinski's publications [**15**], [**16**], [**17**] and the recent book by Saff and Totik [**21**]. A noteworthy reference is the very influential paper of Nevai [**19**] on Freud's mathematical legacy.

2. The Interacting Particle Model.

We propose that a weight function $w(x)$ creates two external fields. One is a long range field whose potential at a point x is $v(x)$ of (1.2). In addition in the presence of n unit charges w produces a short range field whose potential is $\ln(A_n(x)/a_n)$. Thus the total external potential $V(x)$ is the sum of the short and long range potentials, that is

$$V(x) = v(x) + \ln(A_n(x)/a_n). \tag{2.1}$$

The potential energy at x of a point charge e located at c is $-2e\ln|x-c|$. We shall refer to this potential as a logarithmic potential. Consider the system of n movable unit charges in $[a,b]$ in the presence of the external potential $V(x)$ of (2.1). Let

$$\mathbf{x} := (x_1, x_2, \dots, x_n)\,, \tag{2.2}$$

where $x_1, \dots, x_n$ are the positions of the particles arranged in decreasing order. The total energy of the system is

$$E(\mathbf{x}) = \sum_{k=1}^{n} V(x_k) - 2 \sum_{1\le j<k\le n} \ln|x_j - x_k|\,. \tag{2.3}$$

Let

$$T(\mathbf{x}) := \exp(-E(\mathbf{x})). \tag{2.4}$$

Theorem 2.1. *Assume $w(x) > 0$, $x \in (a,b)$ and let $v(x)$ of (1.2) and $v(x) + \ln A_n(x)$ be twice continuously differentiable functions whose second derivative is nonnegative on (a,b). Then the equilibrium position of n movable unit charges in $[a,b]$ in the presence of the external potential $V(x)$ of*

(2.1) *is unique and attained at the zeros of* $p_n(x)$, *provided that the particle interaction obeys a logarithmic potential and that* $T(\mathbf{x}) \to 0$ *as* $\mathbf{x}$ *tends to any boundary point of* $[a,b]^n$, *where*

$$T(\mathbf{x}) = \left[\prod_{j=1}^{n} \frac{\exp(-v(x_j))}{A_n(x_j)/a_n}\right] \prod_{1\le l<k\le n} (x_l - x_k)^2. \tag{2.5}$$

Before proving Theorem 2.1, observe that finding the equilibrium distribution of the charges in Theorem 2.1 is equivalent to finding the maximum of $T(\mathbf{x})$ in (2.4). The reason is that at interior points of $[a,b]^n$, the gradient of T vanishes if and only if the gradient of E vanishes. Furthermore at such points of vanishing gradients the Hessians of T and E have opposite signs. There is no loss of generality in assuming

$$x_1 > x_2 > \cdots > x_n, \tag{2.6}$$

a convention we shall follow throughout this work.

Proof of Theorem 2.1. The assumption $v''(x) > 0$ ensures the positivity of $A_n(x)$. To find an equilibrium position we solve

$$\frac{\partial}{\partial x_j} \ln T(\mathbf{x}) = 0, \quad j = 1, 2, \dots, n.$$

This system is

$$-v'(x_j) - \frac{A_n'(x_j)}{A_n(x_j)} + 2 \sum_{1\le k\le n,\, k\ne j} \frac{1}{x_j - x_k} = 0, \quad j = 1, 2, \dots, n. \tag{2.7}$$

Let

$$f(x) := \prod_{j=1}^{n} (x - x_j). \tag{2.8}$$

It is clear that

$$\begin{aligned} \sum_{1\le k\le n,\, k\ne j} \frac{1}{x_j - x_k} &= \lim_{x\to x_j} \left[\frac{f'(x)}{f(x)} - \frac{1}{x - x_j}\right] \\ &= \lim_{x\to x_j} \left[\frac{(x - x_j)f'(x) - f(x)}{(x - x_j)f(x)}\right] \end{aligned}$$

and L'Hôspital's rule implies

$$2 \sum_{1\le k\le n,\, k\ne j} \frac{1}{x_j - x_k} = \frac{f''(x_j)}{f'(x_j)}. \tag{2.9}$$

Now (2.7), (2.8) and (2.9) imply

$$-v'(x_j) - \frac{A_n'(x_j)}{A_n(x_j)} + \frac{f''(x_j)}{f'(x_j)} = 0, \tag{2.10}$$

or equivalently

$$f''(x) + R_n(x) f'(x) = 0, \quad x = x_1, \dots, x_n,$$

with R_n as in (1.10). In other words

$$f''(x) + R_n(x) f'(x) + S_n(x) f(x) = 0, \quad x = x_1, \dots, x_n. \tag{2.11}$$

To check for local maxima and minima consider the Hessian matrix

$$H = (h_{ij}), \quad h_{ij} = \frac{\partial^2 \ln T(\mathbf{x})}{\partial x_i \partial x_j}. \tag{2.12}$$

It readily follows that

$$\begin{aligned} h_{ij} &= 2(x_i - x_j)^{-2}, \quad i \neq j, \\ h_{ii} &= -v''(x_i) - \frac{\partial}{\partial x_i}\left(\frac{A_n'(x_i)}{A_n(x_i)}\right) - 2 \sum_{1 \le j \le n,\, j \neq i} \frac{1}{(x_i - x_j)^2}. \end{aligned}$$

This shows that the matrix $-H$ is positive definite because it is real, symmetric, strictly diagonally dominant and its diagonal terms are positive, [**12**, Cor. 7.2.2]. Therefore $\ln T$ has no relative minima nor saddle points. Thus any solution of (2.10) will provide a local maximum of $\ln T$ or T. There cannot be more than one local maximum since $T(\mathbf{x}) \to 0$ as $\mathbf{x} \to$ any boundary point along a path in the region defined in (2.6). Thus the system (2.7) has at most one solution. On the other hand (1.9) and (2.11) show that the zeros of

$$f(x) = a_1\, a_2\, \dots,\, a_n\, p_n(x), \tag{2.13}$$

satisfy (2.7), hence the zeros of $p_n(x)$ solve (2.7). This completes the proof of Theorem 2.1.

Let

$$x_{1n} > x_{2n} > \cdots > x_{nn}, \tag{2.14}$$

be the zeros of $p_n(x)$. In [**13**] we used an idea from Schur [**22**] to prove that the discriminant of $p_n(x)$ is given by

$$D_n = \left\{ \prod_{j=1}^{n} \frac{A_n(x_{jn})}{a_n} \right\} \left[\prod_{k=1}^{n} a_k^{2k-2n+2} \right]. \tag{2.15}$$

Our next result gives a representation for the maximum value of T in terms of the recursion coefficients $\{a_n\}$.

Theorem 2.2. *Let $T_{\max}$ and E_n be the maximum value of $T(\mathbf{x})$ and the equilibrium energy of the n particle system. Then*

$$T_{\max} = \exp(-\sum_{j=1}^{n} v(x_{jn})) \prod_{k=1}^{n} a_k^{2k}, \tag{2.16}$$

$$E_n = \sum_{j=1}^{n} v(x_{jn}) - 2\sum_{j=1}^{n} j \ln a_j. \tag{2.17}$$

Proof. Since $T_{\max}$ is

$$\left[\prod_{j=1}^{n} \frac{\exp(-v(x_{jn}))}{A_n(x_{jn})/a_n}\right] \gamma^{2-2n} D_n(p_n), \tag{2.18}$$

then (2.16) follows from (1.15) and (2.15). We also used $\gamma a_1 \cdots a_n = 1$. Now (2.17) holds because E_n is $-\ln(T_{\max})$. This completes the proof.

It is important to observe that if we only know the differential recurrence relation (1.8) and the pure three term recurrence relation (1.3)–(1.4) then using (1.12) Theorem 2.1 can be recast in the following form:

Theorem 2.3. *Let $\{p_n(x)\}$ be orthonormal with respect to a positive measure supported on $[a,b]$ and assume that $A_n(x) > 0$ on (a,b). Define v', which may depend on n (up to an additive constant), through* (1.12). *If $v + \ln(A_n(x)/a_n)$ is convex then the equilibrium position described in Theorem* 2.1 *is unique and is attained at the zeros of $p_n(x)$. Furthermore the minimum energy is given by* (2.17).

Theorem 2.4. *Theorem* 2.1 *holds if instead of $v''(x) > 0$ we require $A_n(x)$ to have a fixed sign in (a,b) and $A_n(x)$ in* (2.1) *and* (2.4) *is replaced by $|A_n(x)|$.*

Proof. The assumption $v''(x) > 0$ was only used to guarantee the positivity of $A_n(x)$ in (a,b). The rest of the proof remains the same. In the case of Laguerre polynomials [**13**]

$$w(x) = \frac{x^\alpha \exp(-x)}{\Gamma(\alpha+1)}, \quad \frac{A_n(x)}{a_n} = \frac{1}{x}, \tag{2.19}$$

and Stieltjes electrostatic interpretation of the zeros of Laguerre polynomials follows from Theorem 2.1. For the Jacobi polynomials

$$w(x) = \frac{(1-x)^\alpha (1+x)^\beta \Gamma(\alpha+\beta+2)}{2^{\alpha+\beta+1}\Gamma(\alpha+1)\Gamma(\beta+1)}, \quad \frac{A_n(x)}{a_n} = \frac{\alpha+\beta+1+2n}{1-x^2}, \tag{2.20}$$

[**13**]. This gives Stieltjes result concerning the electrostatics of zeros of Jacobi polynomials. Observe that the assumption $T(\mathbf{x}) \to 0$ as $\mathbf{x}$ tends to

a boundary point of $[0, \infty]^n$ in the Laguerre case, or a boundary point of $[-1, 1]^n$ in the Jacobi case is automatically satisfied.

The quantity E_n is related to what was denoted by F_n in [**7**]. The first sum in (2.17) is twice the interaction energy while E_n is the free energy, [**7**].

It is worth pointing out that Theorems 2.2 and 2.3 give closed forms for the energy at the equilibrium position in terms of the recursion coefficients. This gives an exactly solvable model in contrast with the earlier models where the location of the moving charges is only found approximately and has no analytic expression other than being the equilibrium position (Fekete points). In a future work we will show that for large n the first two terms in the asymptotics of the energy E_n agrees with the corresponding terms of the free energy of the system [**7**] in several models.

Let us consider the example

$$w(x) = \frac{2\exp(-x^4)}{\Gamma(1/4)}, \quad \text{or} \quad v(x) = x^4 + \ln(\Gamma(1/4)) - \ln 2. \tag{2.21}$$

In the case we have the Freud nonlinear recurrences

$$n = 4a_n^2(a_{n-1}^2 + a_n^2 + a_{n+1}^2), \tag{2.22}$$

[**19**] and $A_n(x)$ is given by, [**5**],

$$A_n(x) = 4a_n\left[x^2 + a_n^2 + a_{n+1}^2\right]. \tag{2.23}$$

In (2.22) $a_0 := 0$. We now consider the case $n = 3$, so

$$a_1a_2a_3p_3(x) = x(x^2 - a_1^2 - a_2^2). \tag{2.24}$$

Since v is even then the unknowns in (2.7) are $\pm x_1$ and zero. The only information we get from (2.7) is that x_1 must satisfy

$$4x^3 + \frac{2x}{x^2 + a_3^2 + a_4^2} = \frac{3}{x}. \tag{2.25}$$

From (2.22) we get

$$a_2^2 + a_3^2 + a_4^2 = \frac{3}{4a_3^2} = \frac{3\,a_2^2}{2[1 - 2a_2^2(a_1^2 + a_2^2)]}.$$

Now (2.22) gives $a_1^2 + a_2^2 = 1/(4a_1^2)$ and (2.25) becomes

$$4x^4 + \frac{2x^2}{x^2 - \frac{1}{4a_1^2} + \frac{2a_1^2}{12a_1^4 - 1}} = \frac{3}{x},$$

and it can be factored as

$$\left(x^2 - \frac{1}{4a_1^2}\right)\left(2x^4 + \frac{4\,a_1^2\,x^2}{12a_1^4 - 1} + \frac{1}{12a_1^4 - 1} - \frac{1}{2}\right) = 0. \tag{2.26}$$

Clearly (2.24) has solutions $\pm(2a_1)^{-1}$ which are $\pm\sqrt{a_1^2 + a_2^2}$, as predicted by Theorem 2.1 and the definition of p_3 in (2.24). To see that (2.24) has no other real solutions it suffices to show $12a_1^4 - 1 < 2$ which will make the second factor on left-hand side of (2.24) strictly positive. Now $\int_{-\infty}^{\infty} p_1^2(x)w(x)dx = 1$ and $p_1(x) = x/a_1$ give $a_1^2 = \Gamma(3/4)/\Gamma(1/4)$ and all we need to show is equivalent to

$$\frac{4\,\Gamma^2(7/4)}{9\,\Gamma^2(5/4)} < 1. \tag{2.27}$$

But $\Gamma(7/4)$ $(= 0.91906)$ and $\Gamma(5/4)$ $(= 0.90640)$ are nearly equal, so (2.25) holds and we have no other real solutions of (2.24) other than $\pm\sqrt{a_1^2 + a_2^2}$.

3. Remarks.

It is known that the zeros $\{x_{jn}\}$ of the Hermite polynomials satisfy the Bethe Ansatz:

$$x_{jn} = \sum_{1\le k\le n,\, k\ne j} \frac{1}{x_{jn} - x_{kn}}. \tag{3.1}$$

A similar property also holds for the Jacobi and Laguerre polynomials, [**1**], [**2**]. We now show that this property has analogues for general polynomials orthogonal with respect to a weight function. Let $\{x_{jn}\}$ be the zeros of $p_n(x)$. One can express the sums $\sum_{1\le k\le n,\, k\ne j} (x_{jn} - x_{kn})^{-i}$ for fixed j in terms of the values of v', A_n, B_n and their derivatives at the zeros $\{x_{jn}\}$. For $i = 1$, use (2.9) and (2.10) to get

$$2\sum_{1\le k\le n,\, k\ne j} \frac{1}{x_{jn} - x_{kn}} = v'(x_{jn}) + \frac{A_n'(x_{jn})}{A_n(x_{jn})}. \tag{3.2}$$

For Hermite polynomials (3.2) reduces to (3.1) since $v(x) = x^2$ and $A_n'(x) = 0$. As an example consider the Freud weight $w(x) = \exp(-x^4)$ so that $v(x) = x^4$. Here $A_n(x)$ is $4a_n\left(x^2 + a_n^2 + a_{n+1}^2\right)$ and (3.2) becomes

$$\sum_{1\le k\le n,\, k\ne j} \frac{1}{x_{jn} - x_{kn}} = 2x_{jn}^3 + \frac{x_{jn}}{x_{jn}^2 + a_n^2 + a_{n+1}^2}. \tag{3.3}$$

To find $\sum_{1\le k\le n,\, k\ne j} (x_{jn} - x_{kn})^{-2}$ in the general case we use

$$\begin{aligned}\sum_{1\le k\le n,\, k\ne j} \frac{1}{(x_{jn} - x_{kn})^2} &= \lim_{x\to x_{jn}} \left[-\frac{d}{dx}\frac{p_n'(x)}{p_n(x)} - \frac{1}{(x - x_{jn})^2}\right] \\ &= \lim_{x\to x_{jn}} \left[\left(\frac{p_n'(x)}{p_n(x)}\right)^2 - \frac{p_n''(x)}{p_n(x)} - \frac{1}{(x - x_{jn})^2}\right].\end{aligned}$$

After some simplification we get

$$(3.4)\qquad \sum_{1\le k\le n,\, k\neq j} \frac{1}{(x_{jn}-x_{kn})^2} = \left[\frac{p_n''(x_{jn})}{2p_n'(x_{jn})}\right]^2 - \frac{p_n'''(x_{jn})}{3p_n'(x_{jn})}.$$

In terms of the coefficients R_n and S_n in the differential equation (1.9) the above sum is
(3.5)

$$\sum_{1\le k\le n,\, k\neq j} \frac{1}{(x_{jn}-x_{kn})^2} = \frac{1}{4}R_n^2(x_{jn}) - \frac{1}{3}R_n(x_{jn}) + \frac{1}{3}\left[R_n'\left(x_{jn}\right) + S_n'(x_{jn})\right].$$

Similarly we can generate sums of higher powers of differences of zeros.

In the remainder of this section we shall concentrate on the case

$$(3.6)\qquad v(x) = c_{2m}x^{2m} + \text{lower order terms}, \quad m = 2, 3, \dots,$$

and $-a = b = \infty$.

Lemma 3.1. *If v is as in* (3.6)*, and is convex then $A_n(x)$ is of degree $2m-2$ and has only complex zeros.*

The proof follows from (1.5).

Let $z_1, z_2, \dots, z_{2m-2}$ be the zeros of $A_n(x)$.

Theorem 3.2. *Under the assumptions of Theorem* 2.1 *the electrostatic system has n movable unit charges, external field*

$$(3.7)\qquad v(x) + \ln(2mc_{2m}) + V_1(x),$$

where V_1 is the potential due to $2m-2$ unit charges at $z_1, z_2, \dots, z_{2m-2}$.

Proof. Clearly (1.5) implies

$$(3.8)\qquad A_n(x)/a_n = 2mc_{2m}\prod_{j=1}^{2m-2}(z-z_j).$$

The electrostatic interpretation now follows from (1.5) and (3.8).

Selberg [**23**] proved

$$(3.9)\qquad \int_{[0,1]^n} \left\{\prod_{j=1}^{n} t_j^{x-1}(1-t_j)^{y-1}\right\} \prod_{1\le i<k\le n} |t_i-t_k|^{2z}\,dt_1\dots dt_n = \prod_{j=1}^{n} \frac{\Gamma(x+(n-j)z)\,\Gamma(y+(n-j)z)\,\Gamma(jz+1)}{\Gamma(x+y+(2n-j-1)z)\,\Gamma(z+1)},$$

for $\mathrm{Re}(x)>0$, $\mathrm{Re}(y)>0$, and $\mathrm{Re}(z) > -\min\{1/n, \mathrm{Re}(x)/n-1, \mathrm{Re}(y)/(n-1)\}$. Here $[0,1]^n$ is the unit cube in R^n. This integral is the multivariate generalization of the beta integral and is now called the Selberg integral. It is important to note that if we normalize the Jacobi polynomials to be orthogonal on $[0,1]$ then the Stieltjes-Hilbert results provide the L_∞ norm of

$$\left\{\prod_{j=1}^{n} t_j^{\alpha}(1-t_j)^{\beta}\right\} \prod_{1\le i<k\le n} |t_i - t_k|^2. \tag{3.10}$$

On the other hand the Selberg integral (3.9) essentially gives the L_p norm of the expression in (3.10). One is then led to view the Stieltjes-Hilbert results as limiting cases of the Selberg integral. Our Theorem 2.2 extends the Stieltjes-Hilbert results from Jacobi polynomials to general orthogonal polynomials, so it would be of interest to explore the analogue of the Selberg integral and evaluate the integrals

$$\int\limits_{[a,b]^n} \left(\prod_{j=1}^{n} \frac{\exp(-v(t_j))}{A_n(t_j)/a_n}\right)^{p} \prod_{1\le i<k\le n} (t_i - t_k)^{2p}\, dt_1\, \ldots\, dt_n. \tag{3.11}$$

In particular for $v(x)$ as in (2.21) the integral in (3.11) is

$$\int\limits_{R^n} \frac{\exp(-p\sum_{j=1}^{n} t_j^4)}{\prod_{j=1}^{n}(t_j^2 + a_n^2 + a_{n+1}^2)^p} \prod_{1\le i<k\le n} (t_i - t_k)^{2p}\, dt_1 \ldots dt_n. \tag{3.12}$$

The a_n's can be generated from (2.22) with the initial values

$$a_0 = 0, \quad a_1^2 = \frac{\Gamma(3/4)}{\Gamma(1/4)}, \quad a_2^2 = \frac{\Gamma(5/4)}{\Gamma(3/4)} - \frac{\Gamma(3/4)}{\Gamma(1/4)}. \tag{3.13}$$

It must be emphasized that the a_n's in (3.12) are not arbitrary but are the recursion coefficients of the Freud polynomials.

4. Energy Asymptotics.

In this section we first discuss the large n asymptotics of E_n in the potential model of the Freud weights. We use the approximation

$$\sum_{j=1}^{n} v(x_{jn}) \sim \int\limits_{x_{nn}}^{x_{1n}} v(x)\,\sigma(x)\,dx, \quad n\to\infty, \tag{4.1}$$

where $\sigma(x)$ is the density of the zeros. In (4.1) $f_n \sim g_n$ as $n \to \infty$ means $f_n/g_n \to 1$ as $n \to \infty$. In the literature on potential theory the measure $\sigma(x)dx$ is called the equilibrium measure. Consider the Freud weight function

$$w_F(x;\alpha) := \frac{\exp(-|x|^{\alpha})}{2\Gamma(1+1/\alpha)}, \quad \alpha \ge 1. \tag{4.2}$$

This case is well studied in the literature but we will use the special form of σ from [**6**] because we need the same normalization. We have

(4.3)
$$\sigma = \sigma(x,\alpha) = \frac{\alpha}{\pi}\frac{2^{1-\alpha}\Gamma(\alpha)}{[\Gamma(\alpha/2)]^2}\, b^{\alpha-2}\sqrt{b^2-x^2}\; {}_2F_1\left(1-\alpha/2, 1; 3/2; 1-(x/b)^2\right),$$

where

(4.4)
$$b := x_{1n} = -x_{nn}.$$

Furthermore from the Freud conjectures

(4.5)
$$b^{\alpha} \sim \frac{[\Gamma(\alpha/2)]^2\, 2^{\alpha-1} n}{\Gamma(\alpha)}.$$

The $A_n(x)$ is now defined by (1.5) with vanishing boundary terms. Using the beta function integral [**20**], the fact

(4.6)
$$(a)_n = \Gamma(a+n)/\Gamma(a),$$

and the Gauss sum for a hypergeometric function of unit argument, [**20**], we get

$$\begin{aligned}
&\int_{-b}^{b} \sigma(x;\alpha)\, v(x) dx \\
&= \frac{\alpha}{\pi}\frac{2^{2-\alpha}\Gamma(\alpha)}{[\Gamma(\alpha/2)]^2} b^{2\alpha} \\
&\quad \cdot \int_0^1 |x|^{\alpha}\sqrt{1-x^2}\; {}_2F_1(1-\alpha/2, 1; 3/2; 1-x^2) dx \\
&= \frac{\alpha}{\pi}\frac{2^{1-\alpha}\Gamma(\alpha)}{[\Gamma(\alpha/2)]^2} b^{2\alpha} \sum_{k=0}^{\infty} \frac{(1-\alpha/2)_k}{(3/2)_k} \int_0^1 t^{(\alpha-1)/2}(1-t)^{k+1/2}\, dt \\
&= \frac{\alpha}{\pi}\frac{2^{1-\alpha}\Gamma(\alpha)}{[\Gamma(\alpha/2)]^2} b^{2\alpha} \sum_{k=0}^{\infty} \frac{(1-\alpha/2)_k \Gamma(k+3/2)\Gamma((\alpha+1)/2)}{(3/2)_k \Gamma(k+2+\alpha/2)} \\
&= \frac{2^{1-\alpha}\Gamma(3/2)\Gamma(\alpha+1)\Gamma((\alpha+1)/2)}{\pi\Gamma^2(\alpha/2)\,\Gamma(2+\alpha/2) b^{-2\alpha}}\, {}_2F_1(1-\alpha/2, 1; 2+\alpha/2; 1) \\
&= \frac{2^{1-\alpha} b^{2\alpha}\, \Gamma(\alpha)\Gamma((\alpha+1)/2)}{\sqrt{\pi}\, \alpha\, \Gamma^3(\alpha/2)}.
\end{aligned}$$

The above calculation and the duplication formula for the gamma function and (4.5) give

$$\sum_{j=1}^{n} v(x_{jn}) \sim \int_{-b}^{b} \sigma(x;\alpha)\, v(x) dx \sim \frac{n^2}{\alpha}, \tag{4.7}$$

as $n \to \infty$. It is also known that $a_n \sim b/2$. Therefore (4.5) gives

$$\sum_{j=1}^{n} j \ln a_j \sim \frac{1}{2} n^2 \ln n, \tag{4.8}$$

and we have proved

$$E_n \sim -\frac{n^2}{\alpha} \ln n. \tag{4.9}$$

This is the same as the main term in the large n form of the free energy in [**7**]. More precise asymptotics will be developed in a future work. In some cases it is possible to express E_n entirely in terms of the recursion coefficients. This is certainly the case when $v(x) = x^{2m}$ and m is a positive integer. We illustrate this for $v(x) = x^4$. Since v is even the p_n's satisfy $p_n(-x) = (-)^n p_n(x)$, so that (1.12) imply

$$a_1 a_2 \ldots a_n p_n(x) = x^n - c_{n,1} x^{n-2} + c_{n,2} x^{n-4} + \cdots . \tag{4.10}$$

From (1.3) and (1.4) it follows that

$$c_{n+1,1} = c_{n,1} + a_n^2 \quad \text{and} \quad c_{n+1,2} = c_{n,2} + a_n^2 c_{n-1,1}, \tag{4.11}$$

hence

$$c_{n,1} = \sum_{j=1}^{n-1} a_j^2, \quad \text{and} \quad c_{n,2} = \sum_{k=3}^{n-1} a_k^2 \left(\sum_{j=1}^{k-2} a_j^2 \right). \tag{4.12}$$

Now

$$\sum_{j=1}^{n} v(x_{jn}) = \sum_{j=1}^{n} x_{jn}^4 = \left[\sum_{j=1}^{n} x_{jn}^2 \right]^2 - 2 \sum_{1 \le j < k \le n} x_{jn}^2 x_{kn}^2. \tag{4.13}$$

Formula (4.10) and $p_n(-x) = (-1)^n p_n(x)$ imply

$$c_{n,1} = \sum_{j=1}^{n} x_{jn}^2, \quad c_{n,2} = \sum_{1 \le j < k \le n} x_{jn}^2 x_{kn}^2, \tag{4.14}$$

hence

$$E_n = c_{n,1}^2 - 2c_{n,2} - 2 \sum_{k=1}^{n} k \ln a_k, \tag{4.15}$$

and $c_{n,1}$ and $c_{n,2}$ are given by (4.12).

Acknowledgments.

Thanks to Richard Askey whose enthusiastic lectures on the Stieltjes-Hilbert electrostatic models and Selberg integrals introduced me to the subject and for his comments on this manuscript. Thanks also to Walter Van Assche who kindly verified that the equilibrium position of Freud weight with $\alpha = 4$ is indeed attained at the zeros of the Freud polynomials for several n's using his wonderful computer program. I am indebted to a very careful referee who repeated the calculations and pointed out several slips. I grateful to Joseph Keller for his interest and criticisms which improved this manuscript.

References

[1] S. Ahmed, M. Bruschi, F. Calegro, M.A. Olshantsky and A.M. Perelomov, *Properties of the zeros of the classical orthogonal polynomials and of the Bessel functions*, Nuovo Cimento, **49B** (1979), 173-199.

[2] S. Ahmed and M. Muldoon, *Reciprocal power sums of differences of zeros of special functions*, SIAM J. Math. Anal., **14** (1983), 372-382.

[3] W. Bauldry, *Estimates of asymmetric Freud polynomials on the real line*, J. Approximation Theory, **63** (1990), 225-237.

[4] S.S. Bonan and D.S. Clark, *Estimates of the Hermite and the Freud polynomials*, J. Approximation Theory, **63** (1990), 210-224.

[5] Y. Chen and M.E.H. Ismail, *Ladder operators and differential equations for orthogonal polynomials*, J. Phys., **A30** (1997), 7818-7829.

[6] ———, *Hermitean matrix ensembles and orthogonal polynomials*, Studies in Appl. Math., **100** (1998), 33-52.

[7] Y. Chen, M.E.H. Ismail and W. Van Assche, *Tau-function constructions of the recurrence coefficients of orthogonal polynomials*, Advances in Appl. Math., **20** (1998), 141-168.

[8] L.E. Dickson, *New Course on the Theory of Equations*, Wiley, New York, 1939.

[9] P.J. Forrester and J.B. Rogers, *Electrostatics and the zeros of the classical orthogonal polynomials*, SIAM J. Math. Anal., **17** (1986), 461-468.

[10] F.A. Grünbaum, *Variation on a theme of Stieltjes and Heine*, J. Comp. Appl. Math., **99** (1998), 189-194.

[11] D. Hilbert, *Über die Discriminante der in Endlichen abbrechenden hypergeometrischen Reihe*, J. für die reine und angewandte Matematik, **103** (1885), 337-345.

[12] R.A. Horn and C. Johnson, *Matrix Analysis*, Cambridge University Press, Cambridge, 1992.

[13] M.E.H. Ismail, *Discriminants and functions of the second kind of orthogonal polynomials*, Results in Mathematics, **34** (1998), 132-149.

[14] T.H. Koornwinder, *Orthogonal polynomials with weight function* $(1-x)^\alpha(1+x)^\beta + M\delta(x+1) + N\delta(x-1)$, Canadian Math. Bull., **27** (1984), 205-214.

[15] D.S. Lubinsky, *A survey of general orthogonal polynomials for weights on finite and infinite intervals*, Acta Applicandae Mathematicae, **10** (1987), 237-296.

[16] ______, *Strong Asymptotics for Extremal Errors and Polynomials Associated with Erdős-Type Weights*, Pitman Research Notes in Mathematics, Vol. 202, Longman, Harlow, 1989.

[17] ______, *An update on orthogonal polynomials and weighted approximation on the real line*, Acta Applicandae Mathematicae, **33** (1993), 121-164.

[18] M.L. Mehta, *Properties of the zeros of a polynomial satisfying a second order linear partial differential equation*, Lett. Nuovo Cimento, **26** (1979), 361-362.

[19] P. Nevai, *Géza Freud, orthogonal polynomials and Christoffel functions: A case study*, J. Approx. Theory, **48** (1986), 3–167.

[20] E.D. Rainville, *Special Functions*, Chelsea, Bronx, 1971.

[21] E.B. Saff and V. Totik, *Logarithmic Potentials With External Fields*, Springer-Verlag, New York, 1997.

[22] I. Schur, *Affektlose Gleichungen in der Theorie der Laguerreschen und Hermiteschen Polynomes*, J. für die reine und angewandte Matematik, **165** (1931), 52-58.

[23] A. Selberg, Bemerkninger om et multiplet integral, Norsk Mat. Tidsskr., **26** (1944), 71-78.

[24] T.J. Stieltjes, *Sur quelques théorèmes d'algèbre*, Comptes Rendus de l'Academie des Sciences, Paris, **100** (1885), 439-440; Oeuvres Complètes, Vol. 1, 440-441.

[25] ______, *Sur les polynômes de Jacobi*, Comptes Rendus de l'Academie des Sciences, Paris, **100** (1885), 620-622; Oeuvres Complètes, Vol. 1, 442-444.

[26] G. Szegő, *Orthogonal Polynomials*, Fourth Edition, Amer. Math. Soc., Providence, 1975.

Received March 24, 1998. This research was partially supported by NSF grant DMS-9625459.

University of South Florida
Tampa, FL 33620-5700
E-mail address: ismail@math.usf.edu

PACIFIC JOURNAL OF MATHEMATICS
Vol. 193, No. 2, 2000

STEINITZ CLASS OF MORDELL–WEIL GROUPS OF ELLIPTIC CURVES WITH COMPLEX MULTIPLICATION

Tong Liu and Xianke Zhang

Let E be an elliptic curve having Complex Multiplication by the ring $\mathcal{O}_K$ of integers of $K = \mathrm{Q}(\sqrt{-D})$, let $H = K(j(E))$ be the Hilbert class field of K. Then the Mordell–Weil group $E(H)$ is an $\mathcal{O}_K$-module. Its Steinitz class $\mathrm{St}(E)$ is studied here. In particular, when D is a prime number, $\mathrm{St}(E)$ is determined: If $D \equiv 3 \pmod 4$ then $\mathrm{St}(E) = 1$; if $D \equiv 1 \pmod 4$ then $\mathrm{St}(E) = [\mathcal{P}]^t$, where $\mathcal{P}$ is any prime-ideal factor of 2 in K, $[\mathcal{P}]$ the ideal class of K represented by $\mathcal{P}$, t is a fixed integer. In addition, general structure for modules over Dedekind domain is also discussed. These results develop the results by D. Dummit and W. Miller for $D = 10$ and specific elliptic curves to more general D and general elliptic curves.

1. Introduction.

Let $K = \mathbf{Q}(\sqrt{-D})$ be an imaginary quadratic number field, $\mathcal{O}_K$ the ring of all integers of K. Let E be an elliptic curve having Complex Multiplication by the ring $\mathcal{O}_K$. Then E is defined over the field $F = \mathbf{Q}(j(E))$, where $j(E)$ denotes the j-invariant of E. So $H = K(j(E))$ is the Hilbert class field of K, [**4**], and the Mordell-Weil group $E(H)$ (i.e., all the H-rational points of E) is naturally a module over the Dedekind domain $\mathcal{O}_K$ (operation is the complex multiplication). By the structural theorem for finitely generated modules over Dedekind domain we have

$$E(H) \cong E(H)_{\mathrm{tor}} \oplus \mathcal{O}_K \oplus \cdots \oplus \mathcal{O}_K \oplus \mathcal{A} = E(H)_{\mathrm{tor}} \oplus \mathcal{O}_K^{s-1} \oplus \mathcal{A},$$

where $\mathcal{A}$ is an ideal of $\mathcal{O}_K$ which is uniquely determined by $E(H)$ up to a multiplication by a number from K. Thus $E(H)$ determines uniquely an ideal class $[\mathcal{A}]$ of K represented by $\mathcal{A}$; $[\mathcal{A}]$ is said to be the Steinitz class of E and denoted by $\mathrm{St}(E)$. (Similarly, any module M over a Dedekind domain R defines an ideal class of R, which is said to be the Steinitz class of M and denoted by $\mathrm{St}(M)$.) So the structure of the Mordell-Weil group $E(H)$, as a module over the Dedekind domain $\mathcal{O}_K$, is uniquely determined by its Steinitz class, rank s, and its torsion part. Therefore, it is important to determine the Steinitz class. D. Dummit and W. Miller, [**1**] in 1996 determined the Steinitz class of some specific elliptic curves when $D = 10$ and found some of their properties.

Since the Steinitz class $\mathrm{St}(E)$ is essentially concerned only with the free part of $E(H)$, we denote

$$E(\cdot)_f = E(\cdot)/E(\cdot)_{\mathrm{tor}},$$

that is, the quotient group of the Mordell group $E(\cdot)$ modulo its torsion part. Note that $E(\cdot)_f$ is isomorphic to the free part of $E(\cdot)$. This notation will be used also for any subgroup I of $E(\cdot)$ to define I_f. Also we can assume the Weierstrass equation of the elliptic curve E to be ([**5**])

$$E:\ y^2 = f(x) = x^3 + a_2x^2 + a_4x + a_6$$

with $a_2,\ a_4,\ a_6 \in F$.

We will first analyze the interior structure of $E(H)$, give a general theorem for the structure of modules over Dedekind domain, and then determine Steinitz classes $\mathrm{St}(E)$ for some types of elliptic curves. In particular, when $D = p$ is a prime number and $p \equiv 3 \pmod 4$, we will prove that $\mathrm{St}(E)$ is the principal class of K; And when $D = p$ is a prime number and $p \equiv 1 \pmod 4$, we will show that

$$\mathrm{St}(E) = [\mathcal{P}]^t, \quad \text{with } t = l + \log|H^1(G, E(H)_f)|,$$

where $\mathcal{P}$ is any prime factor of 2 in K, $l = \mathrm{rank}_{\mathbf{Z}}(E(F))$ is the $\mathbf{Z}$-rank of $E(F)$, $|H^1(G, E(H)_f)|$ is the order of the first cohomology group $H^1(G, E(H)_f)$, and $G = \mathrm{Gal}(H/F)$.

2. The Structure of the Mordell group $E(H)$.

Lemma 1. *The degree of the extension H/F is $[H : F] = 2$, where $F = \mathbf{Q}(j(E))$, $H = K(j(E))$, $j(E)$ is the j-invariant of E.*

Proof. Obviously we have $[H : F] \le 2$. If $[H : F] = 1$, then $K \subset F$. By a result in page 12-13 of [**2**] we know that $F = \mathbf{Q}(j(E))$ has a real embedding into the complex field $\mathbf{C}$. Since K is totally imaginary, $K \subset F$ is impossible. Thus $[H : F] = 2$. This proves the lemma. □

Based on Lemma 1, we assume throughout the Galois group of H/F to be $G = \mathrm{Gal}(H/F) = \{1,\ \sigma\}$. For any $\alpha \in \mathcal{O}_K$, let $[\alpha]$ denote the endomorphism of E corresponding to the multiplication by α. The multiplication by $\sqrt{-D}$ will be important to our following proof. Associating with $E:\ y^2 = f(x)$, we consider the following elliptic curve

$$E_D:\ -Dy^2 = f(x).$$

Note that E_D and E are isomorphic via the map

$$i:\ E_D(\mathbf{C}) \to E(\mathbf{C}), \qquad (x, y) \mapsto (x, \sqrt{-D}y).$$

Therefore we know that

$$\mathrm{End}(E_D) \cong \mathrm{End}(E).$$

So E_D also has complex multiplication by $\mathcal{O}_K$, and is defined over F. Also via the isomorphism i of E and E_D, we have

$$E_D(F) \cong I,$$

where

$$I = \{(x, \sqrt{-D}y) | (x, \sqrt{-D}y) \in E(H),\ x, y \in F\}.$$

The subgroup I of $E(H)$ defined here will be very important in the following analysis.

Lemma 2. *The map $i \circ [\sqrt{-D}]$ is an F-isogeny of E_D to E. Thus*

$$\mathrm{rank}_{\mathbf{Z}}(E_D(F)) = \mathrm{rank}_{\mathbf{Z}}(E(F)) = l.$$

Proof. By [**1**] we have

$$[\sqrt{-D}](x, y) = (a(x), y\sqrt{-D}b(x)),$$

with $a(x),\ b(x) \in F(x)$. So $i \circ [\sqrt{-D}]$ is an F-isogeny of E_D to E. □

Lemma 3. $(I_f : [\sqrt{-D}]E(F)_f)(E(F)_f : [\sqrt{-D}]I_f) = D^l$.

Proof.

$$\begin{aligned} D^l &= (E(F)_f : [D]E(F)_f) \\ &= (E(F)_f : [\sqrt{-D}]I_f)([\sqrt{-D}]I_f : [D]E(F)_f) \\ &= (E(F)_f : [\sqrt{-D}]I_f)(I_f : [\sqrt{-D}]E(F)_f). \end{aligned}$$

□

Lemma 4. $2E(H)_f \subset E(F)_f \oplus I_f \subset E(H)_f$,

$$\mathrm{rank}_{\mathbf{Z}}(E(H)) = \mathrm{rank}_{\mathbf{Z}}(E(F)) + \mathrm{rank}_{\mathbf{Z}}(E_D(F)) = 2\ \mathrm{rank}_{\mathbf{Z}}(E(F)) = 2l.$$

Proof. If $P = (x, y) \in E(F)_f$ with $P \in I_f$, then $y = 0$, which means that P is a torsion point. So $P = O$ is the point at infinity, and $E(F)_f \oplus I_f = E(F)_f + I_f \subset E(H)_f$. For any $Q \in E(H)_f$, we have $2Q = (Q + Q^\sigma) + (Q - Q^\sigma)$, where $G = \mathrm{Gal}(H/F) = \{1,\ \sigma\}$. Via the definition of $E(F)_f$ and I_f, we have

$$E(F)_f = \{P | P^\sigma = P,\ \forall P \in E(H)_f\}, \quad I_f = \{P | P^\sigma = -P,\ \forall P \in E(H)_f\}.$$

So $Q + Q^\sigma \in E(F)_f$, $Q - Q^\sigma \in I_f$, $2Q \in E(F)_f \oplus I_f$. Thus $2E(H)_f \subset E(F)_f \oplus I_f \subset E(H)_f$. This completes the proof. □

As for the index of $E(F)_f \oplus I_f$ in $E(H)_f$, we have the following theorem, which could be also deduced from the cohomology theory of cyclic groups.

Theorem 1.

$$(E(H)_f : E(F)_f \oplus I_f) = \frac{2^l}{|H^1(G, E(H)_f)|},$$

where $|H^1(G, E(H)_f)|$ *is the order of the cohomology group* $H^1(G, E(H)_f)$.

Proof. Consider the colomology group

$$H^1(G, E(H)_f) = Z^1(G, E(H)_f)/B^1(G, E(H)_f).$$

Let $T = \{P - P^\sigma | P \in E(H)_f\}$. We will prove that $Z^1(G, E(H)_f) \cong I_f$, $B^1(G, E(H)_f) \cong T$. For any cocycle $\xi \in Z^1(G, E(H)_f)$, let $\xi \xrightarrow{\phi} \xi_\sigma$, where $\mathrm{Gal}(H/F) = \{1, \sigma\}$. By the definition of cocycle we have that $0 = \xi_1 = \xi_{\sigma^2} = (\xi_\sigma)^\sigma + \xi_\sigma$, so $(\xi_\sigma)^\sigma = -\xi_\sigma$, thus $\xi_\sigma \in I_f$, and ϕ is a map of $Z^1(G, E(H)_f)$ to I_f. Via the map ϕ we could see that $Z^1(G, E(H)_f) \cong I_f$, $B^1(G, E(H)_f) \cong T$. Now consider the homomorphism $E(H)_f \overset{\psi=P-P^\sigma}{\longrightarrow} T$. Obviously $2I_f \subset T$. Since $\psi^{-1}(2I_f) = E(F)_f \oplus I_f$, so

$$\begin{aligned}(E(H)_f : E(F)_f \oplus I_f) = (T : 2I_f) &= (I_f : 2I_f)/(I_f : T)\\ &= 2^l/|H^1(G, E(H)_f)|.\end{aligned}$$

□

3. Main Results and Their Proofs.

We will first give a general theorem on a finitely-generated module over a Dedekind domain, which establishes a relationship between the Steinitz class and the index of the module in its corresponding free module. This theorem is the key to our final results about Steinitz class.

Theorem 2. *Suppose that* L *is a free* $\mathcal{O}_K$*-module, and* $M \subset L$ *is a submodule with* $(L : M) < +\infty$. *Then there is an integral* $\mathcal{O}_K$*-ideal* $\mathcal{A}$ *such that* $[\mathcal{A}]$ *is the Steinitz class of* M, *and* $N^K_{\mathbf{Q}}(\mathcal{A}) = (L : M)$, *where* $N^K_{\mathbf{Q}}(\cdot)$ *denotes the norm map of ideals from* K *to the rationals* $\mathbf{Q}$.

Proof. Let $L = \bigoplus_{i=1}^{n} \mathcal{O}_K e_i$, so $\{e_1, \ldots, e_n\}$ is an $\mathcal{O}_K$-basis for L. We will inductively prove that there are $\mathcal{O}_K$-ideals $\mathcal{B}_i$ $(i = 1, \ldots, n)$ such that $M \cong \bigoplus_{i=1}^{n} \mathcal{B}_i$, and $(L : M) = \prod_{i=1}^{n} (\mathcal{O}_K : \mathcal{B}_i)$.

When $n = 1$, everything is obvious. Assume then the statement is true for $n - 1$ and consider the module-homomorphism $\rho:\ L \to \mathcal{O}_K$, $\rho\left(\sum_{i=1}^{n} r_i e_i\right) = r_n$. Then $\mathcal{B} = \rho(M)$ is an ideal of $\mathcal{O}_K$, and the sequence

$$0 \to N \to M \xrightarrow{\rho} \mathcal{B} \to 0$$

is exact, where $N = \ker(\rho) \cap M$. Since $\mathcal{B}$ is a projective $\mathcal{O}_K$-module, there exists $\mathcal{O}_K$-module $\mathcal{C} \subset M$ such that $\mathcal{C} \cong \mathcal{B}$, $\rho(\mathcal{C}) = \mathcal{B}$, $M = N \oplus \mathcal{C} \cong N \oplus \mathcal{B}$. Thus

$$(L : M) = (L : N \oplus \mathcal{C}) = \left(L : \bigoplus_{i=1}^{n-1} \mathcal{O}_K + \mathcal{C}\right)\left(\bigoplus_{i=1}^{n-1} \mathcal{O}_K + \mathcal{C} : N \oplus \mathcal{C}\right)$$

where $\left(L : \bigoplus_{i=1}^{n-1} \mathcal{O}_K + \mathcal{C}\right) = (\rho^{-1}(\mathcal{O}_K) : \rho^{-1}(\mathcal{B})) = (\mathcal{O}_K : \mathcal{B})$.

Consider $\mathcal{C} \cap \bigoplus_{i=1}^{n-1} \mathcal{O}_K = \mathcal{C} \cap \ker(\rho)$. When restricted on $\mathcal{C}$, the map ρ is injective, so we have

$$\bigoplus_{i=1}^{n-1} \mathcal{O}_K + \mathcal{C} = \bigoplus_{i=1}^{n-1} \mathcal{O}_K \oplus \mathcal{C},$$

$$\begin{aligned}\left(\bigoplus_{i=1}^{n-1} \mathcal{O}_K + \mathcal{C} : N \oplus \mathcal{C}\right) &= \left(\bigoplus_{i=1}^{n-1} \mathcal{O}_K \oplus \mathcal{C} : N \oplus \mathcal{C}\right) \\ &= \left(\bigoplus_{i=1}^{n-1} \mathcal{O}_K : N\right).\end{aligned}$$

Note that $N \subset \bigoplus_{i=1}^{n-1} \mathcal{O}_K$. So via the hypothesis of our induction, we know that there are $\mathcal{O}_K$-ideals $\mathcal{B}_i$ $(i = 1, \ldots, n-1)$ such that $N \cong \bigoplus_{i=1}^{n-1} \mathcal{B}_i$, and $\left(\bigoplus_{i=1}^{n-1} \mathcal{O}_K : N\right) = \prod_{i=1}^{n-1} (\mathcal{O}_K : \mathcal{B}_i)$. Thus we have $M \cong \bigoplus_{i=1}^{n} \mathcal{B}_i$ and $(L : M) = \prod_{i=1}^{n} (\mathcal{O}_K : \mathcal{B}_i) = \prod_{i=1}^{n} N_{\mathbf{Q}}^{K}(\mathcal{B}_i) = N_{\mathbf{Q}}^{K}\left(\prod_{i=1}^{n} \mathcal{B}_i\right)$, where $\mathcal{B}_n = \mathcal{B}$. Now the proof is completed by the following lemma.

Lemma 5. *Assume $\mathcal{A}_1$ and $\mathcal{A}_2$ are two nonzero ideals of the Dedekind domain R, then we have isomorphism of R-modules: $\mathcal{A}_1 \oplus \mathcal{A}_2 \cong R \oplus \mathcal{A}_1\mathcal{A}_2$.*

Proof. See Lemma 13 in page 168 of [**3**]. □

We now intend to prove our main results via our Theorem 2. To use Theorem 2, we need first to find the corresponding L and M in the Mordell group $E(H)$. The corresponding L is given in Lemma 6. While the corresponding M is given in the proofs of Theorem 4 and 5, i.e., $M = [\sqrt{-D}]E(H)_f$ if $D \equiv 3 \pmod 4$; $M = [2\sqrt{-D}]E(H)_f$ if $D \equiv 1 \pmod 4$.

Lemma 6. *$L = \mathcal{O}_K \cdot E(F)_f$ is a free $\mathcal{O}_K$-module of rank l.*

Proof. Assume $P_1,\dots,P_l$ form a $\mathbf{Z}$-basis of $E(F)_f$. We will prove

$$L=\mathcal{O}_K\cdot E(F)_f=\bigoplus_{i=1}^{l}\mathcal{O}_KP_i.$$

Now we suppose that $\sum_{i=1}^{l}[\alpha_i]P_i=0$ for some $\alpha_i\in\mathcal{O}_K$ $(i=1,\dots,l)$. When $D\equiv 3 \pmod 4$, we have $\alpha_i=s_i+t_i(1+\sqrt{-D})/2$ $(s_i,t_i\in\mathbf{Z},\ i=1,\dots,l)$, then via $\sum_{i=1}^{l}[\alpha_i]P_i=0$ we have $\sum_{i=1}^{l}[2s_i+t_i]P_i=0$ and $\sum_{i=1}^{l}[\sqrt{-D}t_i]P_i=0$. Thus $t_i=0$, $s_i=0$, $\alpha_i=0$ $(i=1,\dots,l)$. This proves the theorem when $D\equiv 3 \pmod 4$. The case $D\equiv 1 \pmod 4$ goes in the same way.

To determine our corresponding M in the case $D\equiv 3 \pmod 4$, we need the following theorem.

Theorem 3. *For* $D\equiv 3 \pmod 4$*, we have* $|H^1(G,E(H)_f)|=1$*, and* $E(H)_f=\mathcal{O}_K\cdot E(F)_f+I_f$.

Proof. Let $P_1,\dots,P_l$ form a $\mathbf{Z}$-basis of $E(F)_f$, and $Q_1,\dots,Q_l$ form a $\mathbf{Z}$-basis of I_f. Put $\alpha=(1+\sqrt{-D})/2$. We need only to prove that $E(H)_f/(E(F)_f\oplus I_f)=C_1\oplus\cdots\oplus C_l$, where $C_i=(\overline{[\alpha]P_i})$ is subgroup of order 2 generated by $\overline{[\alpha]P_i}$ in the quotient group $E(H)_f/(E(F)_f\oplus I_f)$. (Here $\overline{a}$ denotes the residue class of a in this quotient group.) Obviously we have $\overline{[\alpha]P_i}\neq\overline{0}$; otherwise there would be $t_j,\ s_j\in\mathbf{Z}$ $(j=1,\dots,l)$ such that $[\alpha]P_i=\sum_{j=1}^{l}[t_j]P_j+\sum_{j=1}^{l}[s_j]Q_j$, then $[1+\sqrt{-D}]P_i=\sum_{j=1}^{l}[2t_j]P_j+\sum_{j=1}^{l}[2s_j]Q_j$, and $P_i=\sum_{j=1}^{l}[2t_j]P_j$, giving a contradiction.

Furthermore, if $\sum_{i=1}^{l}[u_i]\overline{[\alpha]P_i}=\overline{0}$ for some $u_i\in\mathbf{Z}$ $(i=1,\dots,l)$, then there are $t_i,\ s_i\in\mathbf{Z}$ $(i=1,\dots,l)$ such that $\sum_{i=1}^{l}[u_i\alpha]P_i=\sum_{i=1}^{l}[t_i]P_i+\sum_{i=1}^{l}[s_i]Q_i$, so

$$\sum_{i=1}^{l}[u_i]P_i+\sum_{i=1}^{l}[u_i\sqrt{-D}]P_i=\sum_{i=1}^{l}[2t_i]P_i+\sum_{i=1}^{l}[2s_i]Q_i.$$

Thus $\sum_{i=1}^{l}[u_i]P_i=\sum_{i=1}^{l}[2t_i]P_i$, which gives $u_i=2t_i$ $(i=1,\dots,l)$. Hence $[u_i]\overline{[\alpha]P_i}=\overline{[t_i][2\alpha]P_i}=\overline{[t_i(1+\sqrt{-D})]P_i}=\bar{0}$. This completes the proof. □

Now we can prove our main results via Theorem 2.

Theorem 4. *Suppose that* $D = p \equiv 3 \pmod 4$ *is a prime number, and* E *is an elliptic curve having complex multiplication by the full ring* $\mathcal{O}_K$ *of integers of* $K = \mathbf{Q}(\sqrt{-D})$*. Then the Steinitz class of* E *is the principal class, i.e.,* $\mathrm{St}(E) = 1$.

Proof. Let $L = \mathcal{O}_K \cdot E(F)_f$, $M = [\sqrt{-p}]E(H)_f$. Since $M \cong E(H)_f$, we need only to prove $\mathrm{St}(M)$ is the principal class.

By Theorem 3 we have $E(H)_f = \mathcal{O}_K \cdot E(F)_f + I_f$. Thus

$$M = [\sqrt{-p}]E(H)_f = E(F)_f \cdot (\sqrt{-p}\mathcal{O}_K) + [\sqrt{-p}]I_f \subset \mathcal{O}_K \cdot E(F)_f = L;$$

$$\begin{aligned}(L : M) &= (\mathcal{O}_K \cdot E(F)_f : [\sqrt{-p}]E(H)_f) \\ &= \frac{(E(H)_f : [\sqrt{-p}]E(H)_f)}{(E(H)_f : \mathcal{O}_K \cdot E(F)_f)} \\ &= \frac{p^l}{(E(H)_f : \mathcal{O}_K \cdot E(F)_f)}.\end{aligned}$$

Since p is a prime number, so $(L : M) = p^t$ for some t $(0 \le t \le l)$. By Theorem 2, the Steinitz class of M is equal to $[\mathcal{A}]$ for some $\mathcal{O}_K$-ideal $\mathcal{A}$, and $p^t = (L : M) = N_{\mathbf{Q}}^K(\mathcal{A})$. Since p is a prime number, $\mathcal{A} = (\sqrt{-p}\mathcal{O}_K)^t$ is principal. Thus $\mathrm{St}(E) = \mathrm{St}(M)$ is the principal class. □

Theorem 5. *Suppose that* $D = p \equiv 1 \pmod 4$ *is a prime number, and* E *is an elliptic curve having complex multiplication by the ring* $\mathcal{O}_K$ *of all integers of* $K = \mathbf{Q}(\sqrt{-D})$*. Then the Steinitz class of* E *is* $\mathrm{St}(E) = [\mathcal{P}]^t$*, where* $[\mathcal{P}]$ *is the ideal class of* K *represented by* $\mathcal{P}$ *the prime factor of* 2 *in* $\mathcal{O}_K$*,* $2^t = 2^l|H^1(G, E(H)_f)|$*. In particular, the parity of* t *determines* $\mathrm{St}(E)$*, since* $\mathcal{P}$ *is not principal while* $\mathcal{P}^2 = 2\mathcal{O}_K$ *is principal.*

Proof. Let $L = \mathcal{O}_K \cdot E(F)_f$, $M = [2\sqrt{-p}]E(H)_f$. Since $M \cong E(H)_f$, so $\mathrm{St}(E) = \mathrm{St}(M)$. Note that $[2\sqrt{-p}]E(H)_f \subset [\sqrt{-p}](E(F)_f \oplus I_f)$, $[\sqrt{-p}]I_f \subset E(F)_f$. Thus we have $M \subset \mathcal{O}_K \cdot E(F)_f = L$, and

$$\begin{aligned}(L : M) &= (\mathcal{O}_K \cdot E(F)_f : [2\sqrt{-p}]E(H)_f) \\ &= \frac{(E(H)_f : [2\sqrt{-p}]E(H)_f)}{(E(H)_f : \mathcal{O}_K \cdot E(F)_f)} \\ &= \frac{(4p)^l}{(E(H)_f : E(F)_f \oplus I_f)(E(F)_f \oplus I_f : \mathcal{O}_K \cdot E(F)_f)} \\ &= \frac{(4p)^l}{2^l|H^1(G, E(H)_f)|^{-1}(I_f : [\sqrt{-p}]E(F)_f)} \\ &= 2^l|H^1(G, E(H)_f)| \cdot p^l/(I_f : [\sqrt{-p}]E(F)_f).\end{aligned}$$

Thus $(L : M) = 2^t p^r$ for some $t, r \ge 0$, since p is a prime number. By Theorem 2 we know that $N_{\mathbf{Q}}^K(\mathcal{A}) = 2^t p^r$ for some $\mathcal{O}_K$-ideal $\mathcal{A}$. Therefore $\mathcal{A} = \mathcal{P}^t([\sqrt{-p}]\mathcal{O}_K)^r$, $\mathrm{St}(E) = [\mathcal{A}] = [\mathcal{P}^t]$. This proves the theorem. □

Corollary 1. *Suppose as in Theorem* 5. *If* $l = \mathrm{rank}_{\mathbf{Z}}(E(F)) = 1$, *then* $H^1(G, E(H)_f)$ *determines the Steinitz class of* E.

Now we analyze the examples of Dummit and Miller in [**1**] by utilizing the above method. For these examples, we have $K = \mathbf{Q}(\sqrt{-10})$, $D = 10$, $H = K(\sqrt{5}) = \mathbf{Q}(\sqrt{-10}, \sqrt{5})$. We consider the $\mathcal{O}_K$-module $L = \mathcal{O}_K \cdot E(F)_f$ and $M = 2[\sqrt{-10}]E(H)_f$. Then via the same idea in the proof of Theorem 5 we have similar ratiocination for $D = 10$:

$$\begin{aligned}(L : M) &= \frac{(E(H)_f : 2[\sqrt{-10}]E(H)_f)}{(E(H)_f : \mathcal{O}_K \cdot E(F)_f)} \\ &= \frac{(4 \cdot 10)^l}{(E(H)_f : E(F)_f \oplus I_f)(E(F)_f \oplus I_f : \mathcal{O}_K \cdot E(F)_f)} \\ &= \frac{(40)^l}{2^l |H^1(G, E(H)_f)|^{-1} (I_f : [\sqrt{-10}]E(F)_f)} \\ &= 2^l |H^1(G, E(H)_f)| 10^l / (I_f : [\sqrt{-10}]E(F)_f).\end{aligned}$$

Thus the Steinitz class of E is determined by the 2-exponent of

$$2^l |H^1(G, E(H)_f)| (I_f : [\sqrt{-10}]E(F)_f).$$

(DM1) Consider the following elliptic curve of Dummit and Miller in [**1**]:

$$E_1 : \quad y^2 = x^3 + (6 + 6\sqrt{5})x^2 + (7 - 3\sqrt{5}).$$

Then $l = 1$, $|H^1(G, E(H)_f)| = 2$, $(I_f : [\sqrt{-10}]E(F)_f) = 1$. Therefore we know that $2^l |H^1(G, E(H)_f)| (I_f : [\sqrt{-10}]E(F)_f) = 4$. Thus the Steinitz class of E_1 is the principal class, i.e., $\mathrm{St}(E_1) = 1$.

(DM2) Consider the following elliptic curve in [**1**]:

$$E_{1,\mathrm{isog}} : \ y^2 = x^3 - (912 + 12\sqrt{5})x^2 + (188 + 84\sqrt{5})x.$$

We have $l = 1$, $|H^1(G, E(H)_f)| = 2$, $(I_f : [\sqrt{-10}]E(F)_f) = 2$, and $2^l |H^1(G, E(H)_f)| (I_f : [\sqrt{-10}]E(F)_f) = 2^3$. Thus the Steinitz class $\mathrm{St}(E_{1,\mathrm{isog}}) = [\mathcal{P}]$, where $\mathcal{P}$ is a prime factor of 2 in $\mathcal{O}_K$.

(DM3) For $E_3 : \ y^2 = x^3 + 36x^2 + (162 - 72\sqrt{5})x$, in [**1**], we have $l = 2$, $|H^1(G, E(H)_f)| = 2$, $(I_f : [\sqrt{-10}]E(F)_f) = 1$, $2^l |H^1(G, E(H)_f)| (I_f : [\sqrt{-10}]E(F)_f) = 2^3$. Thus $\mathrm{St}(E_3) = [\mathcal{P}]$, $\mathcal{P}$ a prime factor of 2 in $\mathcal{O}_K$.

There are still many open problems about the Steinitz classes of elliptic curves. For example, we have the following conjecture.

Conjecture. *Both the cases* $\mathrm{St}(E) = 1$ *and* $\mathrm{St}(E) \neq 1$ *exist for some elliptic curves* E *having complex multiplication by* $\mathcal{O}_K$, *where* $K = \mathbf{Q}(\sqrt{-D})$ *with prime number* $D \equiv 1 \pmod 4$.

References

[1] D.S. Dummit and W.L. Miller, *The Steinitz class of the Mordell-Weil group of some CM elliptic curves*, J. Number Theory, **56** (1996), 52-78.

[2] B. Gross, *Arithmetic on Elliptic Curves with Complex Multiplication*, SLN, **776**, Springer-Verlag, Berlin, 1980.

[3] F. Keqin, *Introduction to Commutative Algebra*, Higher Education Press, Beijing, 1985.

[4] G. Shimura, *Introduction to the Arithmetic Theory of Automorphic Functions*, Princeton Univ. Press, 1971.

[5] J.H. Silverman, *The Arithmetic of Elliptic Curves*, Spring-Verlag, New York, 1982.

Received September 3, 1998 and revised February 10, 1999.

TSINGHUA UNIVERSITY
BEIJING 100084
P.R. CHINA

TSINGHUA UNIVERSITY
BEIJING 100084
P.R. CHINA
E-mail address: xianke@tsinghua.edu.cn

PACIFIC JOURNAL OF MATHEMATICS
Vol. 193, No. 2, 2000

TOROIDAL SURGERY ON PERIODIC KNOTS

Katura Miyazaki and Kimihiko Motegi

We show that r-Dehn surgery on a hyperbolic, periodic knot K with period $p > 2$ yields a hyperbolic manifold unless $p = 3$, $r = 0$ and the genus of K is one. Regarding hyperbolic, periodic knots with period 2, we show that only integral Dehn surgeries can yield toroidal manifolds.

1. Introduction.

A 3-manifold is *toroidal* if it contains an essential torus, i.e., an incompressible torus not parallel to a boundary component. A knot K in S^3 is called a *periodic knot with period* p if there is a homeomorphism $f : S^3 \to S^3$ such that $f(K) = K$, $\mathrm{Fix}(f) \cap K = \emptyset$, and $\mathrm{Fix}(f)$ is a circle. We call f a *periodic map of* K. For a knot K in a 3-manifold $M \subset S^3$ we denote by $M(K; r)$ the manifold obtained by r-Dehn surgery of M on K, where $r \in \mathbf{Q} \cup \{1/0\}$; if $M = S^3$, simply we denote $M(K; r)$ by $(K; r)$.

The hyperbolic Dehn surgery theorem of Thurston [**25**] shows that for hyperbolic knots K, $(K; r)$ is non-hyperbolic only for finitely many $r \in \mathbf{Q}$. In this paper we consider when Dehn surgery on a hyperbolic, periodic knot yields a non-hyperbolic, in particular toroidal, manifold. For example, the figure eight knot 4_1, which has period 2, has exactly 10 surgeries producing non-hyperbolic manifolds [**25**]; if $(4_1; r)$ is toroidal, then $r = 0, \pm 4$.

Theorem 1.1. *If K is a hyperbolic, periodic knot with period* 2 *and $(K; r)$ is toroidal, then r is an integer.*

Remark. Gordon and Luecke proved that the denominator of a toroidal surgery slope is at most two for hyperbolic knots [**11**], and furthermore if the denominator is two then the knot is strongly invertible [**12**]. Eudave-Muñoz [**5**] constructed an infinite family of strongly invertible hyperbolic knots having non-integral, toroidal surgeries. Theorem 1.1 shows that none of his knots has period 2.

Then, does a hyperbolic, periodic knot with period greater than 2 have a non-hyperbolic Dehn surgery? Our answer is "no except for a special case" (Corollary 1.4). Before giving the statement let us review what non-hyperbolic manifolds are like. Each of the following cases is an obstruction to a closed orientable manifold M being hyperbolic:

(1) M is reducible;
(2) M is a Seifert fibered manifold with a finite fundamental group;
(3) $\pi_1(M)$ has a subgroup isomorphic to $\mathbf{Z} \times \mathbf{Z}$.

In 1981, Thurston announced the Symmetry Theorem [**26**]: If M admits an action by a finite group G such that a fixed point set of some nontrivial element of G has dimension at least one, then M has a G-invariant geometric decomposition such that G acts on each piece by isometries. The theorem implies that (1)-(3) are the only obstructions to such M being hyperbolic. Recently, the Symmetry Theorem is proved in the case when the union of fixed point sets of nontrivial elements of G is a 1-manifold by Cooper, Hodgson and Kerckhoff [**3**], and Boileau and Porti [**1**]; this case of the theorem is what we need and referred to below as the Symmetry Theorem. On the other hand, if M is irreducible, condition (3) implies (3′) below [**7**, Corollary 8.6].

(3′) M is either toroidal or a Seifert fibered manifold with an infinite fundamental group.

If K is a hyperbolic, periodic knot, $(K; r)$ does not fall under case (1) by the Cabling Conjecture for symmetric knots (Hayashi and Shimokawa [**17**], Gordon and Luecke). Since the periodic map of K extends to a periodic map on $(K; r)$, the Symmetry Theorem applies to $(K; r)$. Regarding (2) and (3′), the authors proved that:

Theorem 1.2 ([**21**, Theorem 1.5 and Proposition 5.6])**.** *If K is a hyperbolic, periodic knot with period greater than 2, then $(K; r)$ is not Seifert fibered for any $r \in \mathbf{Q}$.* (*Without using the Symmetry Theorem we show that M is not a Seifert fibered manifold with an infinite fundamental group.*)

Without assuming the Symmetry Theorem, we shall prove:

Theorem 1.3. *Let K be a hyperbolic, periodic knot with period $p > 2$. Then $(K; r)$ is toroidal if and only if $p = 3, r = 0$, and the genus of K is one.*

Remark. The $(3, 3, 3)$ pretzel knot is an example of a genus one, hyperbolic, periodic knot with period 3.

Theorems 1.2 and 1.3 preclude the possibility of cases (2) and (3′). Then the Symmetry Theorem implies that:

Corollary 1.4. *Let K be a hyperbolic, periodic knot with period $p > 2$. Then $(K; r)$ is hyperbolic for any $r \in \mathbf{Q}$ except when $p = 3$, $r = 0$, and the genus of K is one.*

The if part of Theorem 1.3 is proved below. The only if part is proved in §§3, 4. Theorem 1.1 is proved in §5 by graph-theoretic arguments.

Proof of the if part of Theorem 1.3. If K has an incompressible Seifert surface of genus one, then $(K; 0)$ contains a non-separating torus obtained from the

Seifert surface by attaching a meridian disk of the glued solid torus. Gabai [**6**] shows that such a torus is incompressible. □

2. Preliminaries.

2.1. Dehn surgery on a factor knot.

Let K be a periodic knot, and f a periodic map of K with period p. Set $C = \mathrm{Fix}(f)$, which is a trivial knot in S^3 by the positive solution to the Smith Conjecture [**22**]. Then f induces the p-fold cyclic covering π from S^3 to the quotient space $S^3/\langle f\rangle = S^3$ branched along the trivial knot $C_f = \pi(C)$. We denote the factor knot $\pi(K)$ by K_f. Dehn surgeries on K and K_f are related as follows.

Take an f-invariant tubular neighborhood $N(K)$ of K. We can extend $f|S^3 - \mathrm{int}N(K)$ over $(K; m/n)$ periodically. Denote by $\bar{f}$ the resulting periodic map on $(K; m/n)$; the period of $\bar{f}$ is p. We may assume that $\bar{f}$ preserves the core K^* of the reglued solid torus. Note that for any $0 < i < p$, $\mathrm{Fix}(\bar{f}^i)$ is either C or $C \cup K^*$. The projection $\pi' : (K; m/n) \to (K; m/n)/\langle\bar{f}\rangle$ is a p-fold cyclic branched covering. Then $(K; m/n)/\langle\bar{f}\rangle$ is identified with $(K_f; m/(np))$ such that $\pi'(K^*)$ is a core of the reglued solid torus in $(K_f; m/(np))$. So denote $\pi'(K^*) = K_f^*$; see Diagram 2.1.

$$\begin{array}{ccccc}
K & \longrightarrow & K_f & & \\
\cap & & \cap & & \\
S^3 & \xrightarrow{\pi} & S^3/\langle f\rangle = S^3 & & \\
\downarrow & & & \searrow & \\
(K; \frac{m}{n}) & \xrightarrow{\pi'} & (K; \frac{m}{n})/\langle\bar{f}\rangle & = & (K_f; \frac{m}{np}) \\
\cup & & \cup & & \\
K^* & \longrightarrow & K_f^* & &
\end{array}$$

Diagram 2.1. The vertical and the slanted arrows mean Dehn surgeries.

Now choosing an f-invariant tubular neighborhood $N(C)$ of C, set $V = S^3 - \mathrm{int}N(C)$ and $V_f = V/\langle f\rangle = S^3 - \mathrm{int}N(C_f)$. Just as above a Dehn surgery of V on K and that of V_f on K_f are related (Diagram 2.2).

$$\begin{array}{ccccc}
V & \xrightarrow{\pi} & V/\langle f\rangle = V_f & & \\
\downarrow & & & \searrow & \\
V(K; \frac{m}{n}) & \xrightarrow{\pi'} & V(K; \frac{m}{n})/\langle\bar{f}\rangle & = & V_f(K_f; \frac{m}{np})
\end{array}$$

Diagram 2.2. The vertical and the slanted arrows mean Dehn surgeries.

Suppose $K \subset V$ is a hyperbolic knot. Since $\pi : V - K \to V_f - K_f$ is an unbranched covering, $V_f - K_f$ is neither toroidal nor Seifert fibered. Thus K_f is hyperbolic in V_f [**22**]. In the next subsection, we shall show that this hypothesis is satisfied if K is hyperbolic in S^3.

2.2. Hyperbolic, periodic knots.

Proposition 2.1. *Let $K \subset S^3$ be a hyperbolic, periodic knot. Let $C = \mathrm{Fix}(f)$, where f is a periodic map of K. Then $K \cup C$ is a hyperbolic link in S^3.*

Proof. $N(K)$ and $N(C)$ denote disjoint tubular neighborhoods of K and C which are preserved by f, respectively. Set $V = S^3 - \mathrm{int}N(C)$, an unknotted solid torus. Let $\mathcal{T}$ be a characteristic family of tori for $V - \mathrm{int}N(K)$ whose union is invariant under f [**19**, Theorem 8.6]. It suffices to prove $\mathcal{T} = \emptyset$. Note that since $K \subset S^3$ is hyperbolic, any torus in $\mathcal{T}$ is compressible in $S^3 - K$; in particular, any compressing disk meets C.

Assume for a contradiction that there is a torus in $\mathcal{T}$ which separates $\partial N(K)$ and ∂V. Among such tori let T be the one closest to ∂V. Let V' be the solid torus in V such that $\partial V' = T$. Note $f(V') = V'$, and $T = \partial V'$ is compressible in $S^3 - K$. It follows that V' is unknotted in S^3. By the equivariant loop theorem [**20**] there is a meridian disk D of $S^3 - \mathrm{int}V'$ such that $f(D) = D$ or $f(D) \cap D = \emptyset$. Since $C \cap D \neq \emptyset$, we have $f(D) = D$. Hence, D meets $C = \mathrm{Fix}(f)$ in a single point. This together with the unknottedness of C in S^3 shows that C is a core of the unknotted solid torus $S^3 - \mathrm{int}V'$. A core of V' and C then form a Hopf link, so that T and ∂V bounds $T^2 \times I$. This contradicts the minimality of $\mathcal{T}$.

Hence, if $\mathcal{T} \neq \emptyset$, each torus in $\mathcal{T}$ would not separate $\partial N(K)$ and ∂V. Let T be a torus in $\mathcal{T}$ such that the manifold $E \subset V - \mathrm{int}N(K)$ bounded by T does not contain a torus in $\mathcal{T} - \{T\}$. Then for any i either $f^i(E) = E$ or $f^i(E) \cap E = \emptyset$. Set $X = S^3 - \mathrm{int}(N(K) \cup \bigcup_{i \geq 0} f^i(E))$. Since T is compressible in $S^3 - K$, T is compressible in $S^3 - \mathrm{int}(N(K) \cup E)$ and thus in X. Let D be a compressing disk for $T \subset X$ such that $f(D) = D$ or $f(D) \cap D = \emptyset$ [**20**]. Just as above, the fact $C \cap D \neq \emptyset$ implies that D meets C in a single point. Hence C winds around the knotted solid torus $S^3 - \mathrm{int}E$ geometrically once, which contradicts that C is unknotted in S^3. □

3. Proof of Theorem 1.3: Case when $(m, p) = 1$ or p.

In this section and the next, we prove the only if part of Theorem 1.3.

Let K be a hyperbolic, periodic knot, and f a periodic map of K with period $p > 2$. We use the notation in §2.1 in what follows.

Assume that $(K; m/n)$ is toroidal. Note that $(K; m/n)$ is irreducible and not Seifert fibered (Theorem 1.2). By the equivariant torus decomposition theorem [**19**], $(K; m/n)$ contains an incompressible torus T such that for any i, $\bar{f}^i(T) = T$ or $\bar{f}^i(T) \cap T = \emptyset$. By rechoosing T, if necessary, the $\langle \bar{f} \rangle$-equivariant torus T meets $C \cup K^*$ transversely, and $N(C)$ and $N(K^*)$ in (possibly empty) meridian disks. Note $T \cap K^* \neq \emptyset$.

The proof is divided into three cases: (1) $(m, p) = 1$, (2) $(m, p) = p$, (3) $1 < (m, p) < p$, where (m, p) is the greatest common divisor of m and p.

The first two cases are dealt with in this section. Cases 1 and 3 will lead to contradictions.

Case 1. $(m,p)=1$; then $\mathrm{Fix}(\bar{f}^i)=C$ for $0<i<p$.

Claim 3.1. $T\cap C=\emptyset$.

Proof. Assume that T intersects C in $k(>0)$ points. Then $f(T)=T$. Moreover, since f fixes C pointwise, $\bar{f}$ preserves the orientation of T, and thus $T/\langle\bar{f}\rangle$ is an orientable surface. The assumption $(m,p)=1$ implies $m\neq 0$, and then any closed orientable surface in $(K;m/n)$ is separating. Thus k is even. The projection $\pi':T\to T/\langle\bar{f}\rangle=\pi'(T)$ is a p-fold cyclic branched covering along k branch points of index p. The Riemann-Hurewitz formula gives

$$0=\chi(T)=p\left(\chi(\pi'(T))-k\left(1-\frac{1}{p}\right)\right). \tag{1}$$

It follows $\chi(\pi'(T))>0$. Since $\pi'(T)$ is a closed, orientable surface, $\chi(\pi'(T))$ must be 2. Hence, $2=k(1-1/p)$. We then obtain $(p-1)(k-2)=2$. The solution sets in positive integers are $(k,p)=(3,3),(4,2)$. The former contradicts the fact that k is even; the latter does the assumption $p>2$. □

By Claim 3.1 $\pi':T\to\pi'(T)$ is an unbranched covering, thus $\pi'(T)$ is a Klein bottle or a torus. Hence, the $m/(np)$-surgery of the solid torus V_f on K_f contains a Klein bottle or a torus. Note that K_f is a hyperbolic knot in V_f, for K is hyperbolic in V (Proposition 2.1). Then, if $\pi'(T)$ is a Klein bottle, by [**11**] $|np|=1$. This contradicts $p>1$. It follows that $\pi'(T)$ is a torus. The fact that $\pi':V(K;m/n)\to V_f(K_f;m/(np))$ is an unbranched covering implies that $\pi'(T)$ is an essential torus in $V_f(K_f;m/(np))$. For hyperbolic knots in S^3, Gordon and Luecke [**11**] proved that the denominator of a toroidal surgery slope is at most 2. As pointed out in [**13**], their proof works also for hyperbolic knots in a solid torus. Hence $|np|\leq 2$, a contradiction.

Case 2. $(m,p)=p$; then $\mathrm{Fix}(\bar{f}^i)=C\cup K^*$ for $1<i<p$.

Let $k=|T\cap(C\cup K^*)|$. The projection $\pi':T\to T/\langle\bar{f}\rangle$ is a p-fold cyclic branched covering along k branch points of index p. As in Case 1 we obtain Equation (1), and the relevant solution set is $(k,p)=(3,3)$. This implies that T intersects C or K^* in an odd number of points, so T is a non-separating incompressible torus in $(K;m/n)$. By considering the first homology group we see $m=0$. [**6**, Corollary 8.3] shows if $(K;0)$ contains such a torus, then the genus of K is one as desired.

4. Proof of Theorem 1.3: Case when $1 < (m,p) < p$.

In this case, $\mathrm{Fix}(\bar{f}) = C$, $\bar{f}|K^*$ has period $p/(m,p)$, and $\bar{f}|S^3 - \mathrm{int}N(K^* \cup C)$ has period p. Note that $(K; m/n)$ and $(K_f; m/(np))$ do not contain non-separating closed surfaces because $m \neq 0$. Set $n_1 = |T \cap C|$ and $n_2 = |T \cap K^*|$. Then n_i are even numbers, and $n_2 > 0$.

Subcase 1. $T \cap C \neq \emptyset$.

Then $\bar{f}^{\frac{p}{(m,p)}}|T$ has $n_1 + n_2$ fixed points. This implies that $\pi' : T \to T/\langle \bar{f} \rangle = \pi'(T)$ is a p-fold cyclic branched covering along n_1 branch points of index p and $n_2(m,p)/p$ branch points of index (m,p). Note $n_2(m,p)/p = |\pi'(T) \cap K_f^*|$.

Claim 4.1. $n_1 = n_2 = 2$.

Proof. The Riemann-Hurewitz formula to the covering above gives:

$$(2) \quad 0 = \chi(T) = p\left(\chi(\pi'(T)) - n_1\left(1 - \frac{1}{p}\right) - \frac{n_2(m,p)}{p}\left(1 - \frac{1}{(m,p)}\right)\right).$$

As in the proof of Claim 3.1, we obtain $\chi(\pi'(T)) = 2$. It follows:

$$(3) \qquad 2 = n_1\left(1 - \frac{1}{p}\right) + \frac{n_2(m,p)}{p}\left(1 - \frac{1}{(m,p)}\right).$$

The right hand side of (3) is greater than $n_1/2$, therefore $4 > n_1$. Since $n_1(> 0)$ is even, $n_1 = 2$ as claimed.

Multiplying (3) by p and substituting $n_1 = 2$, we obtain $2p = 2p - 2 + n_2(m,p) - n_2$. Thus $2 + n_2 = n_2(m,p) \geq 2n_2$. It follows that the even number n_2 must be 2. □

Since $(m,p) < p$, we have $n_2(m,p)/p < n_2 = 2$. Hence $n_2(m,p)/p = 1$. This implies that the 2-sphere $\pi'(T)$ in $(K_f; m/(np))$ meets K_f^* in a single point, a contradiction.

Subcase 2. $T \cap C = \emptyset$.

Then the closed surface $\pi'(T)$ is contained in $V_f(K_f; m/(np))$.

Claim 4.2. (1) $\pi'(T)$ is a 2-sphere.
(2) K_f^* meets $\pi'(T)$ in 4 points.

Proof. Let i be the least positive integer such that $\bar{f}^i(T) = T$; then $\pi' : T \to T/\langle \bar{f}^i \rangle = \pi'(T)$ is a p/i-fold cyclic branched covering along $in_2(m,p)/p$ branch points of index (m,p). For simplicity set $k = in_2(m,p)/p$. We then have:

$$(4) \qquad 0 = \chi(T) = \frac{p}{i}\left(\chi(\pi'(T)) - k\left(1 - \frac{1}{(m,p)}\right)\right).$$

This shows $\chi(\pi'(T)) > 0$, so $\pi'(T)$ is $\mathbf{R}P^2$ or S^2. If the orientable manifold $V_f(K_f; m/(np))$ contains $\mathbf{R}P^2$, it has a $\mathbf{R}P^3$ factor in its prime decomposition. This is absurd because no surgery on a hyperbolic knot in a solid torus yields a reducible manifold [**23**]. Therefore $\pi'(T)$ is a 2-sphere as claimed.

Letting $\chi(\pi'(T)) = 2$ in (4), we obtain $2 = k(1-1/(m,p))$. The right hand side is smaller than k and greater than or equal to $k/2$, so that $2 < k \le 4$. Since $k = |\pi'(T) \cap K_f^*|$ is even, it must be 4. □

Claim 4.3. The 2-sphere $\pi'(T)$ in $V_f(K_f; m/(np))$ gives an essential tangle decomposition (defined below) of K_f^*.

Definition. Let K be a knot in a 3-manifold M. A separating 2-sphere $\widehat{S} \subset M$ gives an *essential tangle decomposition* of K if $\widehat{S}$ meets K in 4 points and $S = \widehat{S} - \mathrm{int}N(K)$ is incompressible in $M - \mathrm{int}N(K)$. Note that such an S is boundary-incompressible in $M - \mathrm{int}N(K)$.

Proof. By Claim 4.2 it suffices to see $S = \pi'(T) - \mathrm{int}N(K_f^*)$ is incompressible in $V_f(K_f; m/(np)) - \mathrm{int}N(K_f^*)$. Assume for a contradiction that S has a compressing disk D. Under the unbranched cyclic covering $\pi' : V(K; m/n) - \mathrm{int}N(K^*) \to V_f(K_f; m/(np)) - \mathrm{int}N(K_f^*)$, $\pi'^{-1}(D)$ consists of disks. Since T is incompressible in $V(K; m/n)$, each component of $\pi'^{-1}(\partial D) \cap T$ bounds a unique disk in T which meets K^*. Let Δ be an innermost one among such disks. Recall $\bar{f}|K^*$ has period $p/(m,p)$. Then $g = \bar{f}^{\frac{p}{(m,p)}}$ preserves Δ, and thus $g(\partial\Delta) = \partial\Delta$. This contradicts that $\bar{f}$ permutes the p components of $\pi'^{-1}(D)$ cyclically. □

The following proposition is essentially proved in Wu [**27**, Theorem 4.4]. We say that a Dehn surgery on a knot K is *integral* if the surgery slope on $\partial N(K)$ meets a meridian of K in a single point.

Proposition 4.4. *Let K be a knot in an irreducible 3-manifold M. Suppose that $K \subset M$ admits an essential tangle decomposition. Then $M - \mathrm{int}N(K)$ contains an incompressible, closed orientable surface of genus 1 or 2 which remains incompressible after any non-integral, nontrivial surgery on $K \subset M$.*

In our setting, $M = V_f(K_f; m/(np))$ is irreducible by [**23**], and $K_f^* \subset M$ admits an essential tangle decomposition. The solid torus $V_f = V_f(K_f; 1/0)$ contains no incompressible closed surface. But the 1/0-slope of $K_f \subset V_f$ does not meet a meridian of $K_f^* \subset M$ in a single point by $\begin{vmatrix} 1 & m/(m,np) \\ 0 & np/(m,np) \end{vmatrix} = np/(m,np) = np/(m,p) \neq \pm 1$. This contradicts Proposition 4.4. Hence, Subcase 2 does not occur (Theorem 1.3). □

Proof of Proposition 4.4. Let $\widehat{S}$ be a 2-sphere giving an essential tangle decomposition of $K \subset M$, and set $S = \widehat{S} - \mathrm{int}N(K)$. Let B be a 3-ball in M bounded by $\widehat{S}$, and let $B \cap K = t_1 \cup t_2$, two arcs properly embedded in B.

If $E = B - \mathrm{int}N(t_1 \cup t_2)$ contains an incompressible torus F, then it is incompressible in $M - \mathrm{int}N(K)$. Assume for a contradiction that F compresses after a non-integral, nontrivial surgery on $K \subset M$. We can apply [**4**, Theorem 2.4.4] after cutting $M - \mathrm{int}N(K)$ along F. Then we obtain an annulus $A \subset M - \mathrm{int}N(K)$ such that ∂A consists of an essential loop on F and a longitude of $\partial N(K)$. Isotop A so as to meet S transversely and to minimize $|A \cap S|$. Each component of $A \cap S$ is an arc whose ends are in the longitude. An outermost disk of the components of $A - S$ is then a boundary-compressing disk for S. This contradicts the definition of an essential tangle decomposition. Hence, we may assume that E does not contain an incompressible torus.

Claim 4.5 (Hayashi [**14**])**.** *$M - \mathrm{int}N(K)$ contains an incompressible, closed orientable surface F of genus 2 which has a compressing disk in M intersecting K in a single point.*

By applying [**4**, Lemma 2.5.3] or the arguments in [**24**] to $M - \mathrm{int}N(K)$ cut along F, it follows that F remains incompressible after any non-integral, nontrivial surgery on $K \subset M$. This completes the proof of Proposition 4.4. □

Proof of Claim 4.5. The arcs t_i $(i = 1, 2)$ are attached to $\widehat{S}$ such that $t_i \cap \widehat{S} = \partial t_i$. First surger $\widehat{S}$ along a 1-handle $N(t_i)$ attached to $\widehat{S}$; we obtain a torus meeting K in 2 points. Then surger the torus along a 1-handle $N(\overline{K - t_j})$ where $i \neq j$. Let F_i $(i = 1, 2)$ be the resulting closed surface of genus 2; see Figure 4.1. A cocore D of the 1-handle $N(\overline{K - t_j})$ is a compressing disk for $F_i \subset M$ meeting K in a single point, as desired.

The closed surface F_i splits $M - \mathrm{int}N(K)$ into two components. Let X be the one containing $\partial N(K)$, and Y the other. To prove the claim it suffices to see that either F_1 or F_2 is incompressible in both X and Y. If F_i compresses in X, the intersection of the compressing disk and D can be eliminated by a cut and paste argument, so $F_i - \partial D$ is compressible in $X - D$. This implies that S surgered along t_i is compressible in $E = B - \mathrm{int}N(t_1 \cup t_2)$. However, [**27**, Lemma 2.2] shows that for the atoroidal nontrivial tangle $(B, t_1 \cup t_2)$, S surgered along t_i is incompressible in E for $i = 1$ or 2. Hence, either F_1 or F_2 is incompressible in X.

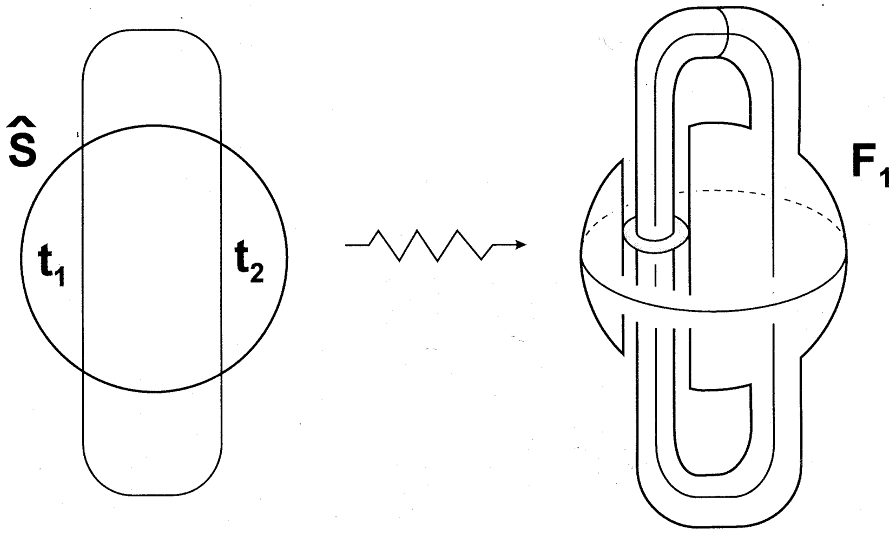


Figure 4.1.

Assume for a contradiction that there is a compressing disk Δ for $F_i \subset Y$, where $i = 1$ or 2. Let $A \subset Y$ be an annulus such that ∂A consists of meridians of the 1-handles $N(t_i)$ and $N(\overline{K - t_j})$ (the shaded annulus in Figure 4.1). By isotopy we may assume that Δ meets A transversely in arcs. Let Δ_0 be the closure of an outermost component in $\Delta - A$. If $\partial\Delta_0 \cap A$ is an arc connecting distinct components of ∂A, then S is boundary-compressible in $M - \text{int}N(K)$, a contradiction. If $\partial\Delta_0 \cap A$ is an arc connecting the same component of ∂A, then $F_i - \partial A$ is compressible in $Y - A$. This implies that S is compressible in $M - \text{int}N(K)$, a contradiction. □

5. Proof of Theorem 1.1.

Although Boyer and Zhang [**2**] showed the theorem when $(K; m/n)$ is Seifert fibered, we proceed without assuming their result.

Let K be a hyperbolic, periodic knot, and f a periodic map of K with period 2. Assume that $(K; m/n)$ is toroidal. If m is even, then [**11**] implies that $|n| = 1$ as desired. In the following we assume that m is odd.

Lemma 5.1. *There is an incompressible torus T in $(K; m/n)$ meeting $C =$ $\text{Fix}(f) = \text{Fix}(\bar{f})$ transversely such that $\bar{f}(T) = T$ or $\bar{f}(T) \cap T = \emptyset$.*

Proof. The lemma follows from the equivariant torus theorem for involutions [**18**, Corollary 4.6] unless $(K; m/n)$ is a Seifert fibered manifold over S^2 with four exceptional fibers. If $(K; m/n)$ is such a Seifert fibered manifold, first choose an $\bar{f}$-invariant Seifert fibration $p : (K; m/n) \to B = S^2$ [**19**] (see also [**21**, Lemma 5.4]). By [**21**, Proposition 5.1] C cannot be a fiber of $(K; m/n)$;

then $\bar{f}$ preserves each fiber meeting C but reverses the orientation of it. It follows that $\bar{f}$ induces an orientation reversing involution, φ, of B which fixes each point on $p(C)$. Then, φ is a reflection about the embedded circle $p(C)$. Let l be a φ-invariant circle in B which meets $p(C)$ transversely and encloses two cone points in each side (Figure 5.1). Then $p^{-1}(l)$ is an $\bar{f}$-invariant incompressible torus meeting C transversely. □

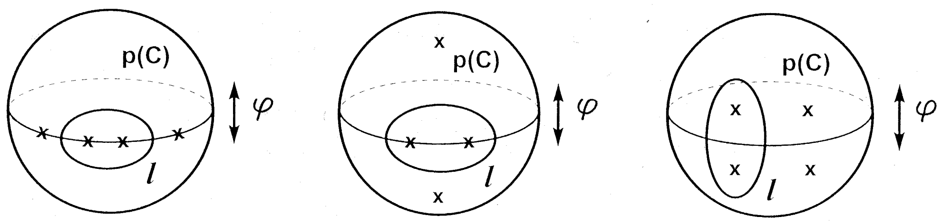


Figure 5.1.

Let T be the torus in Lemma 5.1.

Case 1. $T \cap C = \emptyset$.

The argument in the paragraph just after the proof of Claim 3.1 shows that $|n| = 1$.

Case 2. $T \cap C \neq \emptyset$.

Lemma 5.2. *If* $|n| \geq 2$, *then* $(K_f; m/(2n))$ *has two lens space summands.*

Theorem 1.1 readily follows from this lemma. If $|n| \geq 2$, then by Lemma 5.2 the non-integral surgery $(K_f; m/(2n))$ would be reducible, contradicting [**9**, Theorem 1] (Theorem 1.1). □

The rest of this section is devoted to proving Lemma 5.2 by graph-theoretic technique. The arguments are variants of those in Hayashi and Motegi [**16**, §4].

From the argument in the proof of Claim 3.1, $T/\langle \bar{f} \rangle \cong S^2$ and T meets C in four points. Consider the unbranched covering $\pi' : V(K; m/n) \to V_f(K_f; m/(2n))$. We set $S = \pi'(T - \mathrm{int}N(C))$, a 2-sphere with four open disks removed; S is properly embedded in $V_f(K_f; m/(2n))$ with components of ∂S preferred longitudes of $V_f (\subset S^3)$. Since T is separating in $(K; m/n)$, $T/\langle \bar{f} \rangle$ is separating in $(K_f; m/(2n))$ and hence S separates $V_f(K_f; m/(2n))$.

Claim 5.3. S is essential in $V_f(K_f; m/(2n))$.

Proof. (Cf. the proof of Claim 4.3.) If D is a compressing disk of S in $V_f(K_f; m/(2n))$, $\pi'^{-1}(D)$ consists of two compressing disks of $T - \mathrm{int}N(C)$ in $V(K; m/n)$. However T is incompressible, so each component of $\pi'^{-1}(\partial D)$

bounds a disk in T which meets C. Since $C = \text{Fix}(\bar{f})$, each such disk is preserved by $\bar{f}$. This contradicts that $\bar{f}$ exchanges the components of $\pi'^{-1}(D)$. □

In the following we write $M = V_f - \text{int}N(K_f)$, which is hyperbolic (Proposition 2.1).

Isotoping S so as to minimize $q_S = |S \cap K_f^*|$, we obtain an essential (i.e., incompressible and boundary-incompressible) planar surface $P_S = S \cap M$ in M. Since q_S is even and $(K; m/n) - \text{int}N(K^*)$ is atoroidal, we have $q_S \geq 2$. Let D be a meridian disk of V_f such that $q_D = |D \cap K_f|$ is minimal. Then we have an essential planar surface $P_D = D \cap M$ in M. Since K has period 2, the linking number $lk(C_f, K_f) = lk(C, K)$ is odd, so q_D is odd. If $q_D = 1$, then K is a trivial knot or a composite knot, contradicting the hyperbolicity of K. Thus $q_D \geq 3$.

We define graphs in D and S as in [**4**] and introduce the concepts of (great) x-edge cycles and $[x, x+1]$-Scharlemann cycles as in [**16**]. By an isotopy we may assume that ∂P_D and ∂P_S intersect in minimum number of points, and $P_D \cap P_S$ consists of loops and arcs which are essential in both P_D and P_S. We define Γ_D to be the graph in D such that its (fat) vertices are the disks $D \cap N(K_f)$ and its edges are the arc components e of $P_D \cap P_S$ with at least one endpoint of e in a fat vertex. Similarly, we define the graph Γ_S in S. An edge with one endpoint in ∂D or ∂S is a *boundary edge*.

Number the fat vertices of Γ_D (resp. Γ_S) $1, 2, \ldots, q_D$ (resp. $1, 2, \ldots, q_S$) in the order of appearence on K_f (resp. K_f^*). We next define a *sign* of a vertex of Γ_D to be the sign of the corresponding intersection point of K_f with D with respect to some chosen orientations of D, K_f and M. Similarly, give a sign to each vertex of Γ_S. An edge of Γ_α ($\alpha = D, S$) joining vertices of Γ_α with the same sign is a *positive edge*, and an edge joining the opposite signs is a *negative edge*.

Let p be some edge's endpoint at a fat vertex of Γ_D labelled x. Then p is in the boundary of some fat vertex of Γ_S labelled y (say). We label the edge-endpoint at the fat vertex x with y. Around each fat vertex of Γ_D the edge-endpoint labels occur in order $1, 2, \ldots, q_S, \ldots, 1, 2, \ldots, q_S$ repeated $2|n|$ times; the ordering is, without loss of generality, anticlockwise (resp. clockwise) at a positive (resp. negative) vertex. Label edge-endpoints at fat vertices of Γ_S, similarly. An edge with label x at one endpoint is an *x-edge*.

For a subgraph σ of Γ_D (resp. Γ_S), we call components of $D - \sigma$ (resp. $S - \sigma$) *faces* of σ. For a face P of a subgraph $\sigma \subset \Gamma_\alpha$ ($\alpha = D$ or S), ∂P denotes the subgraph of σ which consists of vertices and edges of σ meeting the closure of P in α. A subgraph σ of Γ_α is an *x-edge cycle* if its edges are positive x-edges, and there is a disk face P of σ such that $\sigma = \partial P$. Furthermore, if all the vertices of Γ_α in P have the same sign as the vertices of σ, then σ is a *great x-edge cycle*. A *Scharlemann cycle* is an x-edge cycle for some label x which bounds a disk face of Γ_α. In our setting Γ_α does not contain a Scharlemann cycle with only one edge. Note that a Scharlemann/x-edge cycle σ is not necessarily a "cycle", i.e., σ with its vertices regarded as points may not be homeomorphic to a circle; see Figure 5.2. The above definition of a Scharlemann cycle is a mild extension of the definition by Gordon and Luecke [**4**], but the same as in Gordon [**8**]. We orient a Scharlemann cycle $\sigma \subset \Gamma_\alpha$ anticlockwise (resp. clockwise) if the sign of the vertices of σ is positive (resp. negative). Then, if an edge of σ has a label x at its tail, then its head has the label $x+1$ (mod q_α). (Cf. Figure 5.2.) We say that σ is a *Scharlemann cycle for the interval* $[x, x+1]$, or simply $[x, x+1]$-*Scharlemann cycle.*

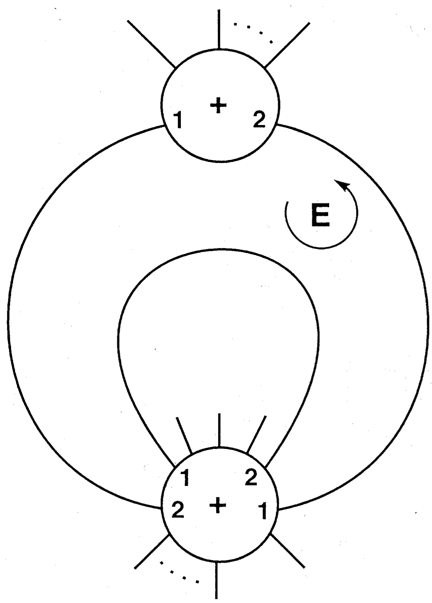


Figure 5.2.

Lemma 5.4. *The graph* Γ_S *does not contain a Scharlemann cycle.*

Proof. If Γ_S contains a Scharlemann cycle, then by [**8**, Theorem 4.1] we have a lens space summand in the solid torus V_f. This is a contradiction. □

Lemma 5.5. *If the graph* Γ_D *contains Scharlemann cycles for distinct intervals, then* $(K_f; m/(2n))$ *has at least two lens space summands.*

Proof. Let σ_i be $[x_i, y_i]$-Scharlemann cycles for $i = 1, 2$ such that $[x_1, y_1] \neq [x_2, y_2]$. Let $E_i \subset D$ be the disk face of σ_i. Then E_i is disjoint from the separating 2-sphere $\widehat{S} = T/\langle \bar{f} \rangle$ in $(K_f; m/(2n))$. There are three posibilities:

$\{x_1, y_1\} \cap \{x_2, y_2\}$ is empty, consists of one element, or two elements. (The last case occurs only when $q_S = 2$, $[x_1, y_1] = [1, 2]$ and $[x_2, y_2] = [2, 1]$.) Except for the first case, E_1 and E_2 are contained in the opposite sides of $\widehat{S}$, thus $(K_f; m/(2n))$ contains two disjoint punctured lens spaces by [**8**, Theorem 4.1].

Assume the first case happens. We consider the subgraph $\widehat{\sigma}_i$ of Γ_S consisting of two vertices of labels x_i, y_i and the edges of σ_i. Then $\widehat{S} - \widehat{\sigma}_1$ consists of open disks; $\widehat{\sigma}_2$ is contained in one of such disks because $\widehat{\sigma}_1 \cap \widehat{\sigma}_2 = \emptyset$. Hence we can choose disjoint disks D_1 and D_2 so that $\widehat{\sigma}_i$ lies in D_i. Thus there are two disjoint punctured lens spaces in $(K_f; m/(2n))$. □

Remark. Since m is assumed to be odd, $H_1((K_f; m/(2n)))$ has odd order. This implies that each Scharlemann cycle in Γ_D has an odd number of edges.

Lemma 5.6 ([**15**, Proposition 5.1])**.** *If Γ_α contains a great x-edge cycle σ, then the disk face of σ contains a Scharlemann cycle.*

Claim 5.7. Γ_S contains at most $q_D(q_S + 2)/2$ positive edges.

Proof. First we show that Γ_S contains at most $q_S + 2$ positive x-edges for every label x. Let x be an arbitrary label of fat vertices of Γ_D. Let Λ be the subgraph of Γ_S consisting of all positive x-edges and all vertices of Γ_S. (The graph Λ may have an isolated vertex.) Note that if Λ has a disk face, its boundary is a great x-edge cycle of Γ_S. Let f_d be the number of disk faces of Λ. Applying Euler's formula to the graph Λ on S, we have $q_S - k + \Sigma\chi(\text{face}) = \chi(S) = -2$, where k is the number of edges of Λ. Thus if $k \geq q_S + 3$, then $f_d \geq \Sigma\chi(\text{face}) \geq 1$, so that Γ_S contains a great x-edge cycle. Hence, Γ_S contains a Scharlemann cycle by Lemma 5.6. This contradicts Lemma 5.4.

Assume for a contradiction that Γ_S contains more than $q_D(q_S + 2)/2$ positive edges. Then the number of their endpoints is more than $q_D(q_S+2)$. By the parity rule [**4**, §2.5] every positive edge has distinct labels at its two endpoints. Since there are q_D kinds edge-endpoint labels in Γ_S, there are more than $q_S + 2$ positive x-edges for some label x. This contradicts what we show above. □

Claim 5.8. If Γ_D has at least $(q_D - 1)q_S$ positive edges, then Γ_D has Scharlemann cycles for distinct intervals.

Proof. We first show that Γ_D contains at least $q_D - 1$ Scharlmann cycles by the arguments in the proof of Claim 5.7 or [**16**, Lemmas 4.5, 4.6]. In fact, using the arguments in the second paragraph of the proof of Claim 5.7, we see that Γ_D has at least $2(q_D - 1)$ positive x-edges for some label x. Then, as in the first paragraph of the proof, apply Euler's formula to the graph Λ on D consisting of all vertices of Γ_D and all positive x-edges of Γ_D. It follows that the Euler number of the faces of Λ is at least $\chi(D) - q_D + 2(q_D - 1) = q_D - 1$.

This implies that Γ_D contains at least $q_D - 1$ great x-edge cycles bounding mutually disjoint disk faces. The claimed result then follows from Lemma 5.6.

Following the proof of [**10**, Theorem 2.3] ([**16**, Lemma 4.4]), we find Scharlemann cycles for distinct intervals. Assume for a contradiction that Γ_D contains Scharlemann cycles only for the interval (say) $[x, x+1]$. Let k be the number of Scharlemann cycles in Γ_D. As in Figure 8 in [**16**], we form a dual graph $\Lambda \subset D$ for Scharlemann cycles. First take one dual (fat) vertex in the disk face of each Scharlemann cycle in Γ_D, and then draw edges from each dual vertex to the vertices of the corresponding Scharlemann cycle. The vertices of Λ consist of q_D vertices of Γ_D and k dual vertices; the edges of Λ consist of the edges defined above. We apply Euler's formula to the graph Λ in D. The number of the vertices is q_D+k; the number of the edges is at least $3k$ by Remark after the proof of Lemma 5.5. It follows that the Euler number of the faces of Λ is at least $\chi(D)-(q_D+k)+3k = 2k+1-q_D \geq q_D-1 > 0$. This implies that there is a disk face of Λ, which contains a great x-edge cycle of Γ_D as shown in [**16**, Figure 9] and thus a Scharlemann cycle (Lemma 5.6). Hence, Γ_D contains more than k Scharlemann cycles, a contradiction. □

Proof of Lemma 5.2. Since each component of ∂S is a longitude of V_f, the graph Γ_S has at most four boundary edges. Each vertex of Γ_S has $|2n|q_D (\geq 4q_D)$ edge-endpoints; Γ_S has at most $q_D(q_S+2)/2$ positive edges (Claim 5.7). Thus, the number of endpoints of the negative edges of Γ_S is at least

$$\begin{aligned} &4q_Dq_S - 4 - q_D(q_S+2) \\ &= 3q_Dq_S - 2q_D - 4 \\ &= 2(q_D-1)q_S + (q_S-2)q_D + 2q_S - 4. \end{aligned}$$

Since $q_S \geq 2$, this number is greater than or equal to $2(q_D-1)q_S$. By the parity rule Γ_D then has at least $(q_D-1)q_S$ positive edges. Hence Γ_D contains Scharlemann cycles for distinct intervals (Claim 5.8). Lemma 5.2 now follows from Lemma 5.5. □

References

[1] M. Boileau and J. Porti, *Geometrization of* 3*-orbifolds of cyclic type*, preprint.

[2] S. Boyer and X. Zhang, *The semi-norm and Dehn filling*, Ann. Math., **148** (1998), 737-801.

[3] D. Cooper, C. Hodgson and S. Kerckhoff, *Geometric structures and symmetries of* 3*-manifolds*, Lecture series given at the Third MSJ Regional Workshop on Cone-Manifolds and Hyperbolic Geometry, July 1-10, 1998, Tokyo Institute of Technology, Tokyo, Japan.

[4] M. Culler, C. McA. Gordon, J. Luecke and P.B. Shalen, *Dehn surgery on knots*, Ann. Math., **125** (1987), 237-300.

[5] M. Eudave-Muñoz, *Non-hyperbolic manifolds obtained by Dehn surgery on a hyperbolic knot*, in 'Studies in Advanced Mathematics', **2**, part 1, (ed. W. Kazez), 1997, Amer. Math. Soc. and International Press, 35-61.

[6] D. Gabai, *Foliations and the topology of 3-manifolds*, III, J. Diff. Geom., **26** (1987), 479-536.

[7] ———, *Convergence groups are Fuchsian groups*, Ann. Math., **136** (1992), 447-510.

[8] C. McA. Gordon, *Combinatorial methods in Dehn surgery*, in 'Lectures at Knots '96, Series on knots and everything', **15** (ed. S. Suzuki), World Scientific, 263-290.

[9] C. McA. Gordon and J. Luecke, *Only integral Dehn surgery can yield reducible manifolds*, Math. Proc. Camb. Phil. Soc., **102** (1987), 97-101.

[10] ———, *Reducible manifolds and Dehn surgery*, Topology, **35** (1996), 385-409.

[11] ———, *Dehn surgeries on knots creating essential tori*, I, Comm. Anal. Geom., **4** (1995), 597-644.

[12] ———, *Dehn surgeries on knots creating essential tori*, II, Comm. Anal. Geom., to appear.

[13] ———, *Toroidal and boundary-reducing Dehn fillings*, Topol. Appl., **93** (1999), 77-90.

[14] C. Hayashi, *On tangle decompositions of super simple knots*, Master Thesis, University of Tokyo, 1992.

[15] C. Hayashi and K. Motegi, *Only single twist on unknots can produce composite knots*, Trans. Amer. Math. Soc., **349** (1997), 4465-4479.

[16] ———, *Dehn surgery on knots in solid tori creating essential annuli*, Trans. Amer. Math. Soc., **349** (1997), 4897-4930.

[17] C. Hayashi and K. Shimokawa, *Symmetric knots satisfy the cabling conjecture*, Math. Proc. Camb. Phil. Soc., **123** (1998), 501-529.

[18] W.H. Holzmann, *An equivariant torus theorem for involutions*, Trans. Amer. Math. Soc., **326** (1991), 887-906.

[19] W.H. Meeks and P. Scott, *Finite group actions on 3-manifolds*, Invent. Math., **86** (1986), 287-346.

[20] W.H. Meeks and S.-T. Yau, *Equivariant Dehn's lemma and loop theorem*, Comment. Math. Helv., **56** (1981), 225-239.

[21] K. Miyazaki and K. Motegi, *Seifert fibered manifolds and Dehn surgery*, III, Comm. Anal. Geom., **7** (1999), 551-582.

[22] J. Morgan and H. Bass (eds.), *The Smith conjecture*, Academic Press, 1984.

[23] M. Scharlemann, *Producing reducible 3-manifolds by surgery on a knot*, Topology, **29** (1990), 481-500.

[24] H. Short, *Some closed incompressible surfaces in knot complements which survive surgery*, in 'Low dimensional topology', London Math. Soc. Lect. Notes Ser., **95** (1985), Cambridge Univ. Press, 179-194.

[25] W. Thurston, *The geometry and topology of 3-manifolds*, Lecture notes, Princeton University, 1979.

[26] ———, *Three dimensional manifolds, Kleinian groups and hyperbolic geometry*, Bull. Amer. Math. Soc., **6** (1982), 357-381.

[27] Y.-Q. Wu, *Dehn surgery on arborescent knots*, J. Diff. Geom., **43** (1997), 171-197.

Received July 24, 1998. The first author was supported in part by Grant-in-Aid for Scientific Research (No. 10640091), Ministry of Education, Science and Culture. The second author was supported in part by Grant-in-Aid for Encouragement of Young Scientists (No. 09740074), Ministry of Education, Science and Culture.

Tokyo Denki University
College of Humanities & Sciences
2-2 Kanda-Nishikicho Nihon University
Tokyo 101
Japan
E-mail address: miyazaki@cck.dendai.ac.jp

Nihon University
Sakurajosui 3-25-40, Setagaya-Ku Tokyo 156
Japan
E-mail address: motegi@math.chs.nihon-u.ac.jp

PACIFIC JOURNAL OF MATHEMATICS
Vol. 193, No. 2, 2000

TOPOLOGICAL DYNAMICS ON MODULI SPACES, I

Joseph P. Previte and Eugene Z. Xia

Let M be a one-holed torus with boundary ∂M (a circle) and Γ the mapping class group of M fixing ∂M. The group Γ acts on $\mathcal{M}_{\mathcal{C}}(\mathrm{SU}(2))$ which is the space of SU(2)-gauge equivalence classes of flat SU(2)-connections on M with fixed holonomy on ∂M. We study the topological dynamics of the Γ-action and give conditions for the individual Γ-orbits to be dense in $\mathcal{M}_{\mathcal{C}}(\mathrm{SU}(2))$.

1. Introduction.

Let M be a Riemann surface of genus g with m boundary components (circles). Let

$$\{\gamma_1, \gamma_2, \dots, \gamma_m\} \subset \pi_1(M)$$

be the elements in the fundamental group corresponding to these m boundary components. The space of SU(2)-gauge equivalence classes of SU(2)-connections, $YM_2(\mathrm{SU}(2))$, is the well-known Yang-Mills two space of quantum field theory. Inside $YM_2(\mathrm{SU}(2))$ is the moduli space $\mathcal{M}(\mathrm{SU}(2))$ of flat SU(2)-connections.

The moduli space $\mathcal{M}(\mathrm{SU}(2))$ has an interpretation that relates to the representation space $\mathrm{Hom}(\pi_1(X), \mathrm{SU}(2))$ which is a real algebraic variety. The group SU(2) acts on $\mathrm{Hom}(\pi_1(M), \mathrm{SU}(2))$ by conjugation, and the resulting quotient space is precisely

$$\mathcal{M}(\mathrm{SU}(2)) = \mathrm{Hom}(\pi_1(M), \mathrm{SU}(2))/\,\mathrm{SU}(2).$$

Conceptually, the moduli space $\mathcal{M}(\mathrm{SU}(2))$ relates to the semi-classical limit of $YM_2(\mathrm{SU}(2))$.

Assign each γ_i a conjugacy class $C_i \subset \mathrm{SU}(2)$ and let

$$\mathcal{C} = \{C_1, C_2, \dots, C_m\}.$$

Definition 1.1. The relative character variety with respect to $\mathcal{C}$ is

$$\mathcal{M}_{\mathcal{C}}(\mathrm{SU}(2)) = \{[\rho] \in \mathcal{M} : \rho(\gamma_i) \in C_i, 1 \le i \le m\}.$$

The space $\mathcal{M}_{\mathcal{C}}(\mathrm{SU}(2))$ is compact, but possibly singular. The set of smooth points of $\mathcal{M}_{\mathcal{C}}(\mathrm{SU}(2))$ possesses a natural symplectic structure ω which gives rise to a finite measure μ on $\mathcal{M}_{\mathcal{C}}(\mathrm{SU}(2))$ (see [**2**, **3**, **5**]).

Let $\text{Diff}(M, \partial M)$ be the group of diffeomorphisms fixing ∂M. The mapping class group Γ is defined to be $\pi_0(\text{Diff}(M, \partial M))$. The group Γ acts on $\pi_1(M)$ fixing the γ_i's. It is known that ω (hence μ) is invariant with respect to the Γ-action. In [**2**], Goldman showed that with respect to the measure μ:

Theorem 1.2 (Goldman)**.** *The the mapping class group* Γ *acts ergodically on* $\mathcal{M}_{\mathcal{C}}(\text{SU}(2))$*.*

Since $\mathcal{M}_{\mathcal{C}}(\text{SU}(2))$ is a variety, one may also study the topological dynamics of the mapping class group action. The topological-dynamical problem is considerably more delicate. To begin, not all orbits are dense in $\mathcal{M}_{\mathcal{C}}(\text{SU}(2))$. If $\sigma \in \text{Hom}(\pi_1(M), G)$ where G is a proper closed subgroup of SU(2) and $\tau \in \Gamma$, then $\tau(\sigma) \in \text{Hom}(\pi_1(M), G)$. In other words, $\mathcal{M}_{\mathcal{C}}(G) \subset \mathcal{M}_{\mathcal{C}}(\text{SU}(2))$ is invariant with respect to the Γ-action. The main result of this paper is the following:

Theorem 1.3. *Suppose* M *is a torus with one boundary component and* $\sigma \in \text{Hom}(\pi_1(M), \text{SU}(2))$ *such that* $\sigma(\pi_1(M))$ *is dense in* SU(2)*. Then the* Γ*-orbit of the conjugacy class* $[\sigma] \in \mathcal{M}_{\mathcal{C}}(\text{SU}(2))$ *is dense in* $\mathcal{M}_{\mathcal{C}}(\text{SU}(2))$*.*

The group SU(2) is a double cover of SO(3):

$$p : \text{SU}(2) \longrightarrow \text{SO}(3).$$

The group SO(3) contains O(2), and the symmetry groups of the regular polyhedra: T' (the tetrahedron), C' (the cube), and D' (the dodecahedron). Let $\text{Pin}(2), T, C$, and D denote the groups $p^{-1}(\text{O}(2))$, $p^{-1}(T')$, $p^{-1}(C')$, and $p^{-1}(D')$, respectively. The proper closed subgroups of SU(2) consist of T, C, D, and the closed subgroups of Pin(2). In particular, $T \subset C$ and $T \subset D$. Suppose $\sigma \in \text{Hom}(\pi_1(X), \text{SU}(2))$. Denote $[\sigma]$ the corresponding SU(2)-conjugacy class in $\mathcal{M}_{\mathcal{C}}(\text{SU}(2))$. Theorem 1.3 implies that if $\sigma(\pi_1(M))$ is not contained in a group isomorphic to C, D, or Pin(2), then the Γ-orbit of the conjugacy class $[\sigma] \in \mathcal{M}_{\mathcal{C}}(\text{SU}(2))$ is dense in $\mathcal{M}_{\mathcal{C}}(\text{SU}(2))$.

A conjugacy class in SU(2) is determined by its trace. For M a torus with one boundary component, the moduli space $\mathcal{M}(\text{SU}(2))$ is a topological ball while $\mathcal{M}_{\mathcal{C}}(\text{SU}(2))$ is generically a smooth 2-sphere. The mapping class group Γ is generated by two Dehn twists τ_X, τ_Y. With a proper change of coordinates, τ_X and τ_Y act on $\mathcal{M}_{\mathcal{C}}(\text{SU}(2))$ by rotations along two axes. The angle of rotation depends on the latitude of the circle along the respective axis. The ergodicity theorem for $\mathcal{M}_{\mathcal{C}}(\text{SU}(2))$ follows since almost all the rotations are irrational multiples of 2π. However, in the context of topological dynamics, one must analyze the orbit of each class $[\rho] \in \mathcal{M}_{\mathcal{C}}(\text{SU}(2))$ upon which one or both of the τ_X, τ_Y actions are rotations of rational multiples of 2π.

The proof of Theorem 1.3 consists of two steps. The first is purely topological-dynamical in nature, concerning the case when the Γ-orbit is infinite. The second step deals with the cases where the Γ-orbits are potentially finite and involves the theory of trigonometric Diophantine equations. All in all, the proof is a delicate interplay of ideas in geometric invariant theory [**6**, **7**], topological dynamics, and Diophantine equations. Incidentally, the proof also yields the well-known result that the only proper closed subgroups SU(2) are the closed subgroups of Pin(2) and the double covers of the automorphism groups of the Platonic solids.

The following conjecture is the analogue of Theorem 1.2 in the category of topological dynamics. Theorem 1.3 is a major stepping stone in the search of a proof for this conjecture.

Conjecture 1.4. *Suppose that M is a Riemann surface with boundary and $\sigma \in \mathrm{Hom}(\pi_1(M), \mathrm{SU}(2))$ such that $\sigma(\pi_1(M))$ is dense in* SU(2)*. Then the Γ-orbit of the conjugacy class $[\sigma] \in \mathcal{M}_{\mathcal{C}}(\mathrm{SU}(2))$ is dense in $\mathcal{M}_{\mathcal{C}}(\mathrm{SU}(2))$.*

Acknowledgments. During the course of this research, Eugene Xia was with the University of Arizona. We thank Professors Michael Brin, William Goldman, Larry Grove, Kirti Joshi, David Levermore, William McCallum, Michelle Previte, and Lawrence Washington for insightful discussions during the course of this research. Eugene Xia also thanks IHÉS for their hospitality during the final phase of this research.

2. Coordinates on the Moduli Space.

For the rest of this paper, fix M to be a torus with one boundary component. We write E for $\mathcal{M}(\mathrm{SU}(2))$ and E_k for $\mathcal{M}_{\mathcal{C}}(\mathrm{SU}(2))$ such that $k = \mathrm{tr}(C)$, where C is the sole element in $\mathcal{C}$. In this section, we briefly summarize some general properties of E. Consult [**2**] for details.

The fundamental group $\pi_1(M)$ has a presentation

$$\pi_1(M) = \langle X, Y, K | K = XYX^{-1}Y^{-1} \rangle$$

where K represents the element generated by the boundary component. In particular, $\pi_1(M)$ is the free group generated by X and Y. Note

$$E = \mathrm{Hom}(\pi_1(M), \mathrm{SU}(2))/\,\mathrm{SU}(2).$$

The SU(2)-invariant polynomials [**7**] on $\mathrm{Hom}(\pi_1(M), \mathrm{SU}(2))$ are generated by the traces of the representations. In particular, a point $[\sigma] \in E$ is determined by

$$x = \mathrm{tr}(\sigma(X)), y = \mathrm{tr}(\sigma(Y)), z = \mathrm{tr}(\sigma(XY)).$$

This provides a global coordinate chart:

$$F : E \longmapsto \mathbb{R}^3$$

$$[\sigma] \overset{F}{\longmapsto} (\mathrm{tr}(\sigma(X)), \mathrm{tr}(\sigma(Y)), \mathrm{tr}(\sigma(XY))).$$

In addition, $k = \mathrm{tr}(\sigma(K))$ is given by the formula

$$k = \mathrm{tr}(\sigma(K)) = x^2 + y^2 + z^2 - xyz - 2. \tag{1}$$

The trace of every element in SU(2) is in $[-2, 2]$. In fact, one can show that

$$E = \{(x, y, z) \in [-2, 2]^3 : -2 \le k \le 2\}.$$

Let f be the map

$$f : E \longrightarrow [-2, 2]$$
$$f([\sigma]) = \mathrm{tr}(\sigma(K)).$$

The fibre $f^{-1}(k)$ is precisely E_k and is a smooth 2-sphere for each $-2 < k < 2$. The fibre $f^{-1}(2)$ is a singular sphere while $f^{-1}(-2)$ consists of one point.

The mapping class group Γ is generated by the maps τ_X and τ_Y:

$$\tau_X(X) = X \text{ and } \tau_X(Y) = YX$$

$$\tau_Y(X) = XY \text{ and } \tau_Y(Y) = Y.$$

The induced action of Γ on E preserves E_k.

With respect to the global coordinate, the actions of τ_X and τ_Y can be described explicitly:

$$\tau_X(x, y, z) = (x, z, xz - y)$$
$$\tau_Y(x, y, z) = (z, y, yz - x).$$

The action of τ_X fixes x and k, and preserves the ellipse

$$E_{x,k} = \{x\} \times \left\{ (y, z) : \frac{2 + x}{4}(y + z)^2 + \frac{2 - x}{4}(y - z)^2 = 2 + k - x^2 \right\}.$$

The topological sphere $f^{-1}(1)$ is pictured below, decomposed into ellipses.

Figure 1. The topological sphere E_1.

The change of coordinates

$$\begin{cases} \tilde{x} &= x \\ \tilde{y} &= \frac{\sqrt{2-x}+\sqrt{2+x}}{2\sqrt{2}} y + \frac{\sqrt{2-x}-\sqrt{2+x}}{2\sqrt{2}} z \\ \tilde{z} &= \frac{\sqrt{2-x}-\sqrt{2+x}}{2\sqrt{2}} y + \frac{\sqrt{2-x}+\sqrt{2+x}}{2\sqrt{2}} z \end{cases}$$

transforms $E_{x,k}$ into the circle

$$E_{x,k} = \{\tilde{x}\} \times \{(\tilde{y}, \tilde{z}) : \tilde{y}^2 + \tilde{z}^2 = 2 + k - \tilde{x}^2\}.$$

In this new coordinate system, τ_X acts as a rotation by $\cos^{-1}(x/2)$. In short, the sphere E_k is the union of circles

$$E_k = \bigcup_x E_{x,k},$$

and τ_X rotates (up to a coordinate transformation) each level set $E_{x,k}$ by an angle of $\cos^{-1}(x/2)$.

Under an analogous coordinate transformation, the action τ_Y becomes a rotation of $E_{y,k}$ by an angle of $\cos^{-1}(y/2)$.

3. The Closed Subgroup Representations.

All closed proper subgroups of SU(2) are contained in Pin(2), C, or D, where C and D are the double covers of the isometry groups of the cube and dodecahedron, respectively. In this section, we classify the Pin(2) representation classes and produce a list of global coordinates for some C, D representation classes. We shall prove later that the list is complete up to some simplifying assumptions and the variations allowed by the following proposition:

Proposition 3.1. *Suppose σ is a G representation such that $G \subset \mathrm{SU}(2)$. If (x, y, z) are the global coordinates of $[\sigma]$, then any permutation of (x, y, z) also corresponds to a G representation class. In addition, if $-I \in G$, then the triples $(-x, y, -z)$, $(x, -y, -z)$, $(-x, -y, z)$ also correspond to G representation classes.*

Proof. Suppose σ is a G representation such that $[\sigma]$ has global coordinates (x, y, z). Since G is a group, the representation σ', with

$$\sigma'(X) = \sigma(XY), \quad \sigma'(Y) = \sigma(X^{-1}),$$

is also a G representation. Moreover $[\sigma']$ has global coordinates (z, x, y). The other permutations of coordinates are handled similarly.

Suppose σ is a G representation. Since $-I \in G$, the representation σ', with

$$\sigma'(X) = -\sigma(X), \quad \sigma'(Y) = \sigma(Y),$$

is also a G representation and the global coordinates of $[\sigma']$ is $(-x, y, -z)$. The other cases are handled similarly. □

If $-I \in G$, then two classes

$$[\sigma], [\sigma'] \in \mathrm{Hom}(\pi_1(M), G)/G$$

are called S-equivalent if their global coordinates differ from one another as prescribed by Proposition 3.1. If $-I$ is not in G, then two classes $[\sigma], [\sigma']$

are called S-equivalent if their global coordinates differ from one another by a permutation.

3.1. The Pin(2) Representations. The group Pin(2) has two components, and we write

$$\mathrm{Pin}(2) = \mathrm{Spin}(2) \cup \mathrm{Spin}_{-}(2),$$

where Spin(2) is the identity component of Pin(2).

Proposition 3.2. *A representation σ is a* Spin(2) *representation if and only if* $\mathrm{tr}(\sigma(K)) = 2$.

Proof. If σ is a Spin(2) representation, then

$$\sigma(K) = \sigma(XYX^{-1}Y^{-1}) = I,$$

since Spin(2) is abelian. Hence,

$$k = \mathrm{tr}(\sigma(K)) = \mathrm{tr}(I) = 2.$$

If $k = \mathrm{tr}(\sigma(K)) = 2$, then

$$I = \sigma(K) = \sigma(X)\sigma(Y)\sigma(X)^{-1}\sigma(Y)^{-1}.$$

This implies that the image of σ is abelian, hence, is contained in Spin(2). □

Proposition 3.3. *A representation σ is a* Pin(2) *representation and not a* Spin(2) *representation if and only if $k \neq 2$ and at least two of the three global coordinates of $[\sigma]$ are zero.*

Proof. If σ is a Pin(2) representation and not a Spin(2) representation, then at least two of the following

$$\sigma(X), \sigma(Y), \sigma(XY)$$

are in $\mathrm{Spin}_{-}(2)$. Since $A \in \mathrm{Spin}_{-}(2)$ implies $\mathrm{tr}(A) = 0$, at least two of the three global coordinates of $[\sigma]$ must be zero.

Suppose two of the three global coordinates of $[\sigma]$ are zero, say $y, z = 0$. One easily finds a Pin(2) representation σ' such that the global coordinates of $[\sigma']$ are $(x, 0, 0)$:

$$\sigma'(X) = \begin{pmatrix} \cos(\theta) & \sin(\theta) \\ -\sin(\theta) & \cos(\theta) \end{pmatrix}, \sigma'(Y) = \begin{pmatrix} 0 & i \\ i & 0 \end{pmatrix},$$

where $x = 2\cos(\theta)$. If $x \neq \pm 2$, then $k \neq 2$ and σ is not a Spin(2) representation. Since global coordinates are unique,

$$[\sigma] = [\sigma'] \in \mathrm{Hom}(\pi_1(M), \mathrm{SU}(2))/\mathrm{SU}(2).$$

Thus, σ is a Pin(2) representation but not a Spin(2) representation. The proofs for the other cases are similar. □

This provides a complete characterization of the Pin(2) representation classes.

Corollary 3.4. *The space of* Spin(2) *representation classes consists precisely of* E_2 *(or* ∂E*). The* Pin(2) *representation classes consist of* E_2 *and the intersections of the three axes with* E*. For* $-2 < k < 2$*, there are exactly six points corresponding to* Pin(2) *representation classes in* E_k*.*

3.2. The C and D Representations. Since C (respectively, D) is finite, the Γ-orbit of a C (respectively, D) representation class is finite. One also notes that $-I$ is in C and D.

We introduce the quaternionic model for SU(2), namely set $1, \mathrm{i}, \mathrm{j}, \mathrm{k}$ as

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad \begin{bmatrix} i & 0 \\ 0 & -i \end{bmatrix}, \quad \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}, \quad \begin{bmatrix} 0 & i \\ i & 0 \end{bmatrix},$$

respectively. Then $\mathrm{SU}(2) = \{x + y\mathrm{i} + z\mathrm{j} + w\mathrm{k} : x^2 + y^2 + z^2 + w^2 = 1\}$, with the usual quaternionic multiplication.

For the rest of the paper, we fix the constants

$$\begin{cases} r = \frac{\sqrt{5}+1}{4} \\ s = \frac{\sqrt{5}-1}{4}. \end{cases}$$

Let

$$\begin{cases} T = \frac{\sqrt{2}}{2} + \frac{\sqrt{2}}{2}\mathrm{i} \\ U = \frac{\sqrt{2}}{2} + \frac{\sqrt{2}}{2}\mathrm{j}, \end{cases}$$

$$\begin{cases} A = r + s\mathrm{i} + \frac{1}{2}\mathrm{k} \\ B = -s + \frac{1}{2}\mathrm{i} - r\mathrm{k}. \end{cases}$$

Then

$$\langle T, U \rangle \cong C \text{ and } \langle A, B \rangle \cong D.$$

X	Y	S-equivalence classes (x,y,z)	k
T	$(TU)^{-1}$	$(\sqrt{2},1,\sqrt{2})$	1
A^3B^8	$(AABA)^{-1}$	$(1,1,1)$	0
B^{-1}	AAB	$(-2s,2s,2s)$	$\frac{1-\sqrt{5}}{2}$
ABA^3B^2	B	$(-2s,-2s,1)$	$\frac{1-\sqrt{5}}{2}$
A^{-1}	ABB	$(2r,-2r,-2r)$	$\frac{1+\sqrt{5}}{2}$
$ABAAB^6$	$AABA$	$(2r,1,2r)$	$\frac{1+\sqrt{5}}{2}$
$ABAA$	A	$(1,2r,1)$	1
AB^{-6}	B^{-6}	$(1,2s,-1)$	1
$ABAA$	A^{-1}	$(1,2r,2s)$	1

Table 1. Global coordinates for some C and D representation classes.

Table 1 is a list of S-equivalence classes that come from either $\langle T,U\rangle$ (cube) or $\langle A,B\rangle$ (dodecahedron) representation classes. We shall prove in Section 6 the following statement. Suppose (x,y,z) is a C or D representation class, but not a Pin(2) representation class. Then there exists $\gamma\in\Gamma$ such that $\gamma(x,y,z)$ is in one of the S-equivalence classes in Table 1.

4. The Irrational Rotations and Infinite Orbits.

The Dehn twist τ_Y acts on the (transformed) subsets $E_{y,k}$ via a rotation of angle $\cos^{-1}(y/2)$. The y-coordinates that yield finite orbits under τ_Y create a filtration as follows: Let $Y_n\subset(-2,2)$ such that $y\in Y_n$ implies that if $(x,y,z)\in E$ are global coordinates of a representation class, then the τ_Y-orbit of (x,y,z) is periodic with period greater than one but less than or equal to n. This gives a filtration

$$\emptyset = Y_2\subset Y_3\subset\ldots\subset Y_n\subset\ldots.$$

For example, $Y_2=\emptyset$, $Y_3=\{-1\}$, $Y_4=\{-1,0\}$, etc. In particular, Y_n is a finite set for every n. By symmetry, there exists a similar filtration X_n, with $X_n=Y_n$ as sets.

Fix $-2<k<2$ and consider the two-dimensional sphere E_k. The global coordinates provide an embedding of E_k in $\mathbb{R}^3$ as a submanifold. Hence E_k inherits a metric from the flat Riemannian metric on $\mathbb{R}^3$. This provides a distance function (metric) d on E_k. The metric d generates the usual topology on E_k. Note that there are two points on E_k that are fixed by τ_Y. These points correspond to Pin(2) representation classes.

Definition 4.1. For $\epsilon > 0$, a set U is ϵ-dense if for each $p \in U \subset E_k$, there exists a point $q \in U$ such that $0 < d(p,q) < \epsilon$.

Let $\epsilon > 0$. Since E_k is compact, there is an $M(\epsilon)$ such that $n \geq M(\epsilon)$ implies that every τ_Y-orbit $\mathcal{O}_y$ in $E_{y,k}$ is ϵ-dense for any $y \notin Y_{M(\epsilon)}$. Let $N(\epsilon)$ be the cardinality of $Y_{M(\epsilon)}$.

Proposition 4.2. *Let $(x_0, y_0, z_0) \in E_k$. Suppose that one of* $\cos^{-1}(x_0/2)$ *or* $\cos^{-1}(y_0/2)$ *is an irrational multiple of* π*. Then the* Γ*-orbit of* (x_0, y_0, z_0) *is dense in* E_k*.*

Proof. Suppose that $\cos^{-1}(x_0/2)$ is an irrational multiple of π. Let $\epsilon > 0$ and $(x_*, y_*, z_*) \in E_k$ which does not correspond to a Pin(2) representation class. Since $\cos^{-1}(x_0/2)$ is an irrational multiple of π, τ_X acts on the (transformed) subset of $E_{x_0,k}$, by an irrational rotation. By the compactness of E_k, there exists a y-value, $y_1 \neq y_*$ and $\delta > 0$ such that $E_{y_*,k}$ is in the ϵ-neighborhood of $E_{y_1,k}$ and $0 < \delta < d(E_{y_1,k}, E_{y_*,k})$. We first consider the special case where there exists an integer J such that the y-coordinate of $\tau_X^J(x_0, y_0, z_0)$ is strictly between y_1 and y_*. Since $\cos^{-1}(x_0/2)$ is an irrational multiple of π, there are infinitely many integers J_i such that the y-coordinate of $\tau_X^{J_i}(x_0, y_0, z_0)$ is strictly between y_1 and y_*. Choose J_i such that the y-coordinate of $(\tau_X)^{J_i}(x_0, y_0, z_0)$ is not in $Y_{M(\epsilon)}$. By the triangle inequality, there is some point on the τ_Y-orbit of $(\tau_X)^{J_i}(x_0, y_0, z_0)$ that is at most 2ϵ from (x_*, y_*, z_*).

We now prove the proposition in general. Since (x_*, y_*, z_*) satisfies Equation (1) and is not a Pin(2) representation class, $E_{y_*,k}$ is a circle. Hence, we must have that $(0, y_*, z') \in E_{y_*,k}$ for some $z' \neq 0$. Therefore, there exists $x_2 \neq 0$ such that $E_{x_2,k}$ intersects $E_{y_*,k}$.

Choose $\epsilon' = \frac{\delta}{N(\epsilon)+2}$. By the filtration X_n, the set $X_{M(\epsilon')}$ contains all x-values that have the following properties:

1) $E_{x,k}$ intersects $E_{y_*,k}$.
2) There is a point in $E_{x,k}$ whose τ_X-orbit has at most $N(\epsilon)$ points with distinct y-coordinates between y_1 and y_*.

Note that the x-coordinate of $\tau_Y(\tau_X)^J(x_0, y_0, z_0)$ is the z-coordinate of $(\tau_X)^J(x_0, y_0, z_0)$. Since $\cos^{-1}(x_0/2)$ is an irrational multiple of π, there is an infinite sequence of numbers J_i such that $|x_{J_i}| < |x_2|$, where x_{J_i} is the the x-coordinate of $\tau_Y(\tau_X)^{J_i}(x_0, y_0, z_0)$. This forces $E_{x_{J_i},k}$ to intersect $E_{y_*,k}$. Of these, choose J such that x_J is not in $X_{M(\epsilon')}$. Thus, the τ_X-orbit of $\tau_Y(\tau_X)^J(x_0, y_0, z_0)$ has at least $N(\epsilon) + 1$ points with distinct y-coordinates between y_1 and y_*.

Now at most $N(\epsilon)$ values of y yield τ_Y-orbits that are not ϵ-dense. Thus, there exists a point $(\hat{x}, \hat{y}, \hat{z})$ on the τ_X-orbit of $\tau_Y(\tau_X)^J(x_0, y_0, z_0)$ such that $\hat{y}$ is between y_1 and y_*, moreover, the τ_Y-orbit of $(\hat{x}, \hat{y}, \hat{z})$ is ϵ-dense. Since the

ϵ-neighborhood of $E_{\hat{y},k}$ covers $E_{y_*,k}$, some point in the τ_Y-orbit of $(\hat{x}, \hat{y}, \hat{z})$ comes within 2ϵ of (x_*, y_*, z_*). Finally, by Corollary 3.4, the set of Pin(2) representation classes in E_k consists of six discrete points. This implies there is no loss of generality in assuming that (x_*, y_*, z_*) does not correspond to a Pin(2) representation class. A symmetric argument holds if $\cos^{-1}(y_0/2)$ is an irrational multiple of π. □

Proposition 4.3. *Suppose the* Γ*-orbit of* $(x_0, y_0, z_0) \in E_k$ *is infinite. Then the* Γ*-orbit of* (x_0, y_0, z_0) *is dense in* E_k*.*

Proof. Let $(x_0, y_0, z_0) \in E_k$ have infinite Γ-orbit and $(x_*, y_*, z_*) \in E_k$ which does not correspond to a Pin(2) representation class.

There are two cases. One possibility is that the Γ-orbit $\mathcal{O}$ has an infinite number of points on some circle $E_{y,k}$ (respectively, $E_{x,k}$). Hence, there is an infinite number of points on $\mathcal{O} \cap E_{y,k}$ that have distinct x-coordinates. However, a priori, these points may not be dense in $E_{y,k}$. One uses the infinite number of points on $\mathcal{O} \cap E_{y,k}$ with distinct x-coordinates as in the proof of Proposition 4.2.

The other possible case is that no circle $E_{x,k}$ (or $E_{y,k}$) has an infinite number of points on $\mathcal{O}$. As in the previous case, there are an infinite number of points on the Γ-orbit of (x_0, y_0, z_0) having distinct x-coordinates. The proof again follows similarly to the proof of Proposition 4.2. □

5. Trigonometric Diophantine Equations.

It remains for us to classify the finite Γ-orbits. These orbits exist and can be constructed by taking G-representation classes where G is a closed proper subgroup of SU(2) as described in Section 3.

The problem amounts to considering cases where the rotations generated by τ_X and τ_Y are both rational. In such cases, an additional iteration is made as follows. By assumption, both $\cos^{-1}(x/2)$ and $\cos^{-1}(y/2)$ are rational multiples of π. Also, $\cos^{-1}(z/2)$ is a rational multiple of π since the x-coordinate of $\tau_Y(x, y, z)$ is z. Since $\tau_X(x, y, z) = (x, z, xz-y)$, in order for the orbit to be finite, $\cos^{-1}(\frac{xz-y}{2})$ must be a rational multiple of π. In particular, $x = 2\cos(\theta_x)$, $y = 2\cos(\theta_y)$, $z = 2\cos(\theta_z)$ and $xz - y = 2\cos(\theta_{xz-y})$, where all angles are rational multiples of π in $[0, \pi]$. Hence,

$$2\cos(\theta_x)\cos(\theta_z) - \cos(\theta_y) = \cos(\theta_{xz-y}),$$

or

$$\cos(\theta_x + \theta_z) + \cos(\theta_x - \theta_z) - \cos(\theta_y) = \cos(\theta_{xz-y}), \tag{2}$$

where all angles are rational multiples of π, $0 \leq \theta_x + \theta_z \leq 2\pi$, and $-\pi \leq \theta_x - \theta_z \leq \pi$. Similarly, the action of τ_Y gives

$$\cos(\theta_y + \theta_z) + \cos(\theta_y - \theta_z) - \cos(\theta_x) = \cos(\theta_{yz-x}), \tag{3}$$

where all angles are rational multiples of π, $0 \leq \theta_y + \theta_z \leq 2\pi$, and $-\pi \leq \theta_y - \theta_z \leq \pi$. Equations (2) and (3) are referred to as the τ_X-equation and τ_Y-equation at (x, y, z) (or at $(\theta_x, \theta_y, \theta_z)$), respectively.

Proposition 5.1. *Let $(x, y, z) \in E_k$ with x, y, z all nonzero. Suppose that two terms appearing in Equation* (2) *cancel one another. Then $k = 2$.*

Proof. Suppose $\cos(\theta_y) = -\cos(\theta_{xz-y})$. Then $xz - y = -y$ which implies that $x = 0$ or $z = 0$, a contradiction to the assumption that x, y, z are all nonzero.

Suppose $\cos(\theta_y) = \cos(\theta_x + \theta_z)$. Recall that (x, y, z) satisfies Equation (1). Thus,

$$\begin{aligned}
k &= 4\cos(\theta_x)^2 + 4\cos(\theta_x + \theta_z)^2 + 4\cos(\theta_z)^2 \\
&\quad -8\cos(\theta_x)\cos(\theta_x + \theta_z)\cos(\theta_z) - 2 \\
&= 4\cos(\theta_x)^2 + 4\cos(\theta_z)^2 - 4\cos(\theta_x - \theta_z)\cos(\theta_x + \theta_z) \\
&= 2\cos(2\theta_x) + 2 + 2\cos(2\theta_z) - 2\cos(2\theta_x) - 2\cos(2\theta_z) = 2.
\end{aligned}$$

A similar argument applies in the case $\cos(\theta_y) = \cos(\theta_x - \theta_z)$. □

A symmetric argument shows that if two terms appearing in Equation (3) cancel one another, then $k = 2$. For the remainder of this paper, we assume that $(x, y, z) \in E$ does not correspond to a Pin(2) representation class. This implies that $k \neq 2$. In addition, we may assume that all coordinates of $(x, y, z) \in E_k$ are nonzero: For if $x = 0$, then y, z must both be nonzero by Proposition 3.3. The point $\tau_Y(0, y, z) = (z, y, yz)$ has all nonzero entries.

Equation (2) is an at most four-term Diophantine equation, the solutions to which are few as shown by Conway and Jones.

Theorem 5.2 (Conway, Jones, [**1**]). *Suppose that we have at most four distinct rational multiples of π lying strictly between 0 and $\pi/2$ for which some linear combination of their cosines is rational, but no proper subset has this property. That is,*

$$A\cos(a) + B\cos(b) + C\cos(c) + D\cos(d) = E,$$

for A, B, C, D, E rational and $a, b, c, d \in (0, \pi/2)$ rational multiples of π. Then the appropriate linear combination is proportional to one from the *following list:*

$$\begin{aligned}\cos(\pi/3) &= 1/2\\ \cos(t+\pi/3)+\cos(\pi/3-t)-\cos(t) &= 0 \quad (0<t<\pi/6)\\ \cos(\pi/5)-\cos(2\pi/5) &= 1/2\\ \cos(\pi/7)-\cos(2\pi/7)+\cos(3\pi/7) &= 1/2\\ \cos(\pi/5)-\cos(\pi/15)+\cos(4\pi/15) &= 1/2\\ -\cos(2\pi/5)+\cos(2\pi/15)-\cos(7\pi/15) &= 1/2\\ \cos(\pi/7)+\cos(3\pi/7)-\cos(\pi/21)+\cos(8\pi/21) &= 1/2\\ \cos(\pi/7)-\cos(2\pi/7)+\cos(2\pi/21)-\cos(5\pi/21) &= 1/2\\ -\cos(2\pi/7)+\cos(3\pi/7)+\cos(4\pi/21)+\cos(10\pi/21) &= 1/2\\ -\cos(\pi/15)+\cos(2\pi/15)+\cos(4\pi/15)-\cos(7\pi/15) &= 1/2.\end{aligned}$$

The nonzero cosine terms in Equation (2) are not necessarily in $(0,\pi/2)$. By applying the identities $\cos(\pi/2-t)=-\cos(\pi/2+t)$ and $\cos(\pi-t)=\cos(\pi+t)$, we derive from Equation (2) a new four-term cosine equation whose arguments are in $[0,\pi/2]$. That is, by a possible change of sign, each term in Equation (2) may be rewritten with angle in $[0,\pi/2]$.

Equation (2) cannot correspond to the last four equations appearing in Theorem 5.2 since these equations have five nonzero terms. By Proposition 5.1, we may assume that the angles appearing in Equation (2) are all distinct. For if two or more angles are the same, then after combining terms, the resulting equation must be proportional to the first equation in Theorem 5.2. This leads to $k=2$ by Proposition 5.1.

Proposition 5.3. *Suppose $(x,y,z)\in E_k$ are not the global coordinates of a* Pin(2) *representation class with x,y,z all nonzero. Suppose that the Γ-orbit of (x,y,z) is finite and that some angle in Equation* (2) *or Equation* (3) *is an integer multiple of π. Then (x,y,z) is S-equivalent to a triple appearing in Table* 1.

Proof. By the assumption $k\neq 2$, the only way that Equation (2) can have angles equal to an integer multiple of π is if $\theta_x-\theta_z=0$ or $\theta_x+\theta_z=\pi$, i.e., $x=z$ or $x=-z$. Note that both cannot happen simultaneously.

Suppose $\theta_x+\theta_z=\pi$. Then Equation (2) becomes

$$\cos(\theta_x-\theta_z)-\cos(\theta_y)-\cos(\theta_{xz-y})=1. \tag{4}$$

Theorem 5.2 and the assumption $y\neq 0$ lead to the following cases:

(A) Two terms in Equation (4) correspond to the first equation of Theorem 5.2, with one term equal to zero (angle $\pi/2$).

Since $y\neq 0$, we must have that $\cos(\theta_y)=-\frac{1}{2}$, so $y=-1$. Now either $\cos(\theta_{xz-y})=0$ or $-\frac{1}{2}$. If $\cos(\theta_{xz-y})=0$, then $xz=-1$, hence $x^2=1$ which yields the triples $(1,-1,-1)$ and $(-1,-1,1)$. If $\cos(\theta_{xz-y})=\frac{1}{2}$, then

$xz = -2$, or $x = \pm\sqrt{2}$. The resulting triples (x, y, z) are $(\sqrt{2}, -1, -\sqrt{2})$ and $(-\sqrt{2}, -1, \sqrt{2})$. Note that all of the above triples belong to an S-equivalence class appearing in Table 1.

(B) Two terms in Equation (4) correspond to the third equation of Theorem 5.2 while the remaining term corresponds to the first equation in Theorem 5.2 . The resulting triples are:

$$(-2s, 2s, 2s), (2s, 2s, -2s), (-2s, -1, 2s), (2s, -1, -2s),$$
$$(-2r, -2r, 2r), (2r, -2r, -2r), (-2r, -1, 2r), (2r, -1, -2r),$$
$$(-1, -2r, 1), (1, -2r, -1), (-1, 2s, 1), (1, 2s, -1).$$

Note that all above triples belong to an S-equivalence class appearing in Table 1.

Suppose $\theta_x - \theta_z = 0$. Then Equation (2) becomes

$$\cos(\theta_x + \theta_z) - \cos(\theta_y) - \cos(\theta_{xz-y}) = -1. \tag{5}$$

Theorem 5.2 and the assumption $y \neq 0$ lead to the following cases.

(A) Two terms in Equation (5) correspond to the first equation of Theorem 5.2, with one term equal to zero (angle $\pi/2$). The various possibilities lead to the triples:

$$(1, 1, 1), (-1, 1, -1), (\sqrt{2}, 1, \sqrt{2}), (-\sqrt{2}, 1, -\sqrt{2}).$$

(B) Two terms in Equation (5) correspond to the third equation of Theorem 5.2 while the other corresponds to the first equation. The various possibilities lead to the triples:

$$(2r, 2r, 2r), (-2r, 2r, -2r), (2r, 1, 2r), (-2r, 1, -2r),$$
$$(1, 2r, 1), (-1, 2r, -1), (1, -2s, 1), (-1, -2s, -1),$$
$$(2s, -2s, 2s), (-2s, -2s, -2s), (2s, 1, 2s), (-2s, 1, -2s).$$

Again, the S-equivalence classes of these triples appear in Table 1. A similar argument holds if some angle in Equation (3) is an integer multiple of π, i.e., $y = \pm z$. □

Henceforth, we assume that all angles in nonzero cosine terms appearing in Equation (2) are distinct and not integer multiples of π. Under these assumptions, Equation (2) can be rewritten as an equation that satisfies the hypotheses of the following which is a special case of Theorem 5.2:

Theorem 5.4 (Conway, Jones, [**1**]). *Suppose that we have at most four distinct rational multiples of π lying strictly between 0 and $\pi/2$ for which some linear combination of their cosines is zero, but no proper subset has this property. That is,*

$$A\cos(a) + B\cos(b) + C\cos(c) + D\cos(d) = 0,$$

for A, B, C, D rational and $a, b, c, d \in (0, \pi/2)$ rational multiples of π. Then the linear combination is proportional to one from the following list:

$$\cos(t + \pi/3) + \cos(\pi/3 - t) - \cos(t) = 0 \quad (0 < t < \pi/6)$$
$$\cos(\pi/5) - \cos(2\pi/5) - \cos(\pi/3) = 0$$
$$\cos(\pi/7) - \cos(2\pi/7) + \cos(3\pi/7) - \cos(\pi/3) = 0$$
$$\cos(\pi/5) - \cos(\pi/15) + \cos(4\pi/15) - \cos(\pi/3) = 0$$
$$-\cos(2\pi/5) + \cos(2\pi/15) - \cos(7\pi/15) - \cos(\pi/3) = 0.$$

6. Proof of Theorem 1.3.

In this section we prove the following proposition which in turn proves Theorem 1.3:

Proposition 6.1. *Let $(x, y, z) \in E_k$ with x, y, z nonzero. Then the Γ-orbit of (x, y, z) is infinite or there is $\gamma \in \Gamma$ such that $\gamma(x, y, z)$ is S-equivalent to a triple in Table* 1.

An immediate consequence of Proposition 6.1 is:

Corollary 6.2. *Suppose (x, y, z) is a C or D representation class, but not a* Pin(2) *representation class. Then there exists $\gamma \in \Gamma$ such that $\gamma(x, y, z)$ is in one of the S-equivalence classes in Table* 1.

The proof of Proposition 6.1 presented here is lengthy and highly computational. We begin by outlining the overall strategy. Consider all triples (x, y, z) that arise from the solutions of Equation (2) provided by Theorem 5.4. For a triple (x, y, z) to have a finite Γ-orbit, the four-term trigonometric equations that arise from repeated applications of τ_X or τ_Y must have solutions provided by Theorem 5.4 or violate the hypotheses of Theorem 5.4. We prove Proposition 6.1 by showing that all triples (x, y, z) that arise from the solutions of Equation (2) provided by Theorem 5.4 fall into one of the following three categories:

1) (x, y, z) has $k \geq 2$ or one of the global coordinates (x, y, z) is zero.
2) (x, y, z) belongs to one of the S-equivalence classes in Table 1. Hence, corresponds to a C or a D representation class.
3) (x, y, z) has infinite Γ-orbit with $-2 < k < 2$. Hence, has dense Γ-orbit by Proposition 4.3.

Definition 6.3. The pairs of angles (θ_z, θ_x), $(\pi - \theta_x, \pi - \theta_z)$, and $(\pi - \theta_z, \pi - \theta_x)$ are called the symmetric, dual, and dual-symmetric pairs of (θ_x, θ_z), respectively.

Recall that by a possible change of sign, each term in Equation (2) may be rewritten with angle in $[0, \pi/2]$. Therefore, for fixed $a, b \in [0, \pi/2]$, we obtain the following eight systems of equations

$$\begin{cases} \cos(a) &= \pm\cos(\theta_x+\theta_z) \\ \cos(b) &= \pm\cos(\theta_x-\theta_z) \end{cases} \qquad \begin{cases} \cos(b) &= \pm\cos(\theta_x+\theta_z) \\ \cos(a) &= \pm\cos(\theta_x-\theta_z), \end{cases}$$

for $\theta_x, \theta_z \in (0,\pi)$. The following, together with their dual, symmetric, and dual-symmetric pairs, are all possible pairs (θ_x, θ_z) that satisfy one of the above eight systems of equations:

$$(\tfrac{a+b}{2}, |\tfrac{a-b}{2}|), (\pi - \tfrac{a+b}{2}, |\tfrac{a-b}{2}|),$$
$$(\pi/2 - \tfrac{a+b}{2}, \pi/2 + \tfrac{a-b}{2}), (\pi/2 - \tfrac{a+b}{2}, \pi/2 + \tfrac{b-a}{2}).$$

We prove in detail the cases in which Equation (2) corresponds to Equation 2 or 3 in Theorem 5.4. The argument for Equation 3 is the simplest and exemplifies the primary techniques used in the other cases. The case of Equation 2 involves a free parameter t, hence, is somewhat more involved than the others. For the other cases, we simply enumerate all the possible solutions and categorize them according to the above mentioned categories.

We first consider the case of

$$\cos(\pi/7) - \cos(2\pi/7) + \cos(3\pi/7) - \cos(\pi/3) = 0$$

in detail. Consider $a = \pi/7$ and $b = 2\pi/7$. As given above, the possibilities for (θ_x, θ_z) are

$$(3\pi/14, \pi/14), (11\pi/14, \pi/14), (2\pi/7, 3\pi/7), (2\pi/7, 4\pi/7),$$

along with their symmetric, dual, and dual-symmetric pairs. Suppose first that $\theta_y = \pi/3$ (respectively, $2\pi/3$). Then the angles $\theta_y \pm \theta_z$ are rationally related to π by reduced rationals with denominators 21 or 42. Thus, for (x, y, z) to have a finite Γ-orbit, the τ_Y-equation at $(\theta_x, \theta_y, \theta_z)$ must correspond to Equation 1. Since $x \neq 0$ this equation can be rewritten as Equation 1 only if $yz - x = 0$, or $y = \frac{x}{z}$. Then $y \neq \frac{x}{z}$ for $\theta_y = \pi/3$ (or $2\pi/3$), and x, z as given by the four pairs listed above (as well as their dual, symmetric, and dual-symmetric pairs). Thus the τ_Y-equation at $(\theta_x, \theta_y, \theta_z)$ must violate the hypotheses of Theorem 5.4.

If $\theta_{xz-y} = \pi/3$ or $2\pi/3$ then, consider the point $\tau_X(\theta_x, \theta_y, \theta_z) = (\theta_x, \theta_z, \theta_{xz-y})$ (in angle notation). The τ_Y-equation at $(\theta_x, \theta_z, \theta_{xz-y})$ is

$$\cos(\theta_z + \theta_{xz-y}) + \cos(\theta_z - \theta_{xz-y}) - \cos(\theta_x) = \cos(\theta_{z(xz-y)-x}),$$

but here $xz - y$ is playing the role of y in the argument given above.

This type of argument works for the other five cases (e.g., $a = \pi/7, \quad b = 3\pi/7$, etc.).

Now we cover the case

$$\cos(t + \pi/3) + \cos(\pi/3 - t) - \cos(t) = 0 \quad (0 < t < \pi/6).$$

Recall that we are under the standing assumption $x, y, z \neq 0$. There are six cases.

Case 1: For $a = t + \pi/3$, and $b = \pi/3 - t$, the possibilities are:

$$(\pi/3, t), (2\pi/3, t), (\pi/6, \pi/2 + t), (\pi/6, \pi/2 - t),$$

and their dual, symmetric, and dual-symmetric pairs. Since $y \neq 0$, θ_y corresponds to t and $xz - y = 0$.

The first two pairs, together with their duals $(\pi/3, \pi - t), (2\pi/3, \pi - t)$, yield $y = \pm z$. By Proposition 5.3, these triples are either S-equivalent to those appearing in Table 1 or have infinite Γ-orbits.

In the above symmetric (dual-symmetric) pairs, we have $y = \pm x$ and $z = 1$. The τ_X preimage of $(x, \pm x, 1)$ is $(x, \pm x^2 - 1, \pm x)$ (signs taken together). Since $\pm x^2 - 1 \neq 0$, we may apply Proposition 5.3 to this triple as above. The same holds for $z = -1$.

Now consider the last two pairs. Note that the angle $\theta_x = \pi/6$ does not appear in Equations 1-5. Therefore, any triple associated with these pairs will have an infinite $\Gamma-$orbit. For the symmetric pairs, i.e., $\theta_z = \pi/6$ (respectively, $5\pi/6$), note that $\tau_X(\theta_x, \theta_y, \theta_z) = (\theta_x, \theta_z, \theta_{xz-y})$. The τ_Y-equation at $(\theta_x, \theta_z, \theta_{xz-y})$ cannot be put into the form of Equations 1-5, as above. This argument also applies to the duals of all such pairs.

Case 2: For $a = t + \pi/3$, and $b = t$, the possibilities are:

$$(t + \pi/6, \pi/6), (5\pi/6 - t, \pi/6), (\pi/3 - t, 2\pi/3), (\pi/3 - t, \pi/3),$$

and their dual, symmetric, and dual-symmetric pairs. Since $y \neq 0$, θ_y corresponds to $\pi/3 - t$ and $xz - y = 0$.

As in the previous case, the last two pairs together with their dual, symmetric, and dual-symmetric pairs yield triples (or triples whose τ_X preimage) are either S-equivalent to those appearing in Table 1 or have infinite Γ-orbits. The argument for the first two pairs is also similar to that in the previous case.

Case 3: For $a = \pi/3 - t$, and $b = t$, the possibilities are:

$$(\pi/6, \pi/6 - t), (5\pi/6, \pi/6 - t), (\pi/3, 2\pi/3 - t), (\pi/3, t + \pi/3),$$

together with their dual, symmetric, and dual-symmetric pairs. An argument similar to the one given in case 1 holds.

Case 4: For $a = \pi/3 + t$, and $b = \pi/2$, the possibilities are:

$$(5\pi/12 + t/2, \pi/12 - t/2), (7\pi/12 - t/2, \pi/12 - t/2),$$
$$(\pi/12 - t/2, 5\pi/12 + t/2), (\pi/12 - t/2, 7\pi/12 - t/2),$$

and their duals. Note that the last two pairs are the symmetric pairs of the first two. As in previous arguments, it is enough to consider $\theta_y = t$ or $\pi - t$.

We first handle the triple (in angle notation) $(5\pi/12 + t/2, t, \pi/12 - t/2)$. Note that three of the angles in the τ_Y-equation of this point (all angles rewritten in $[0, \pi/2]$) are: $5\pi/12+t/2$, $\pi/12+t/2$, and $|\pi/12-3t/2|$. However, both $|\pi/12-3t/2|$ and $\pi/12+t/2$ correspond to angles in $[0, \pi/6)$. If $|\pi/12-3t/2|$ is nonzero, then the τ_Y-equation of this point cannot correspond to Equations 1-5 since no equation in Theorem 5.4 has two angles in $[0, \pi/6)$. If $|\pi/12 - 3t/2| = 0$, then Proposition 5.3 applies. For $(5\pi/12 + t/2, \pi - t, \pi/12-t/2)$, the angles in the τ_Y-equation of this point (all angles rewritten in $[0, \pi/2]$) are again: $5\pi/12 + t/2$, $|-\pi/12 + 3t/2|$, and $\pi/12 + t/2$.

The dual of the pair $(5\pi/12+t/2, \pi/12-t/2)$ is $(7\pi/12-t/2, 11\pi/12+t/2)$. Three of the angles in the τ_Y-equation of $(7\pi/12 - t/2, t, 11\pi/12 + t/2)$ are: $7\pi/12 - t/2$, $11\pi/12 + 3t/2$, and $11\pi/12 - t/2$, which, when rewritten in $[0, \pi/2]$, become: $5\pi/12 + t/2$, $|\pi/12 - 3t/2|$, and $\pi/12 + t/2$. These are the same angles as before, thus the same argument applies. The same argument also holds for $(7\pi/12 - t/2, \pi - t, 11\pi/12 + t/2)$.

Consider the symmetric pair $(\pi/12 - t/2, 5\pi/12 + t/2)$. For the triple $(\pi/12 - t/2, t, 5\pi/12 + t/2)$, three of the angles in the τ_Y-equation of this point are: $\pi/12 - t/2$, $5\pi/12 + 3t/2$, and $5\pi/12 - t/2$. Note that the angle $5\pi/12 + 3t/2$, rewritten in $[0, \pi/2]$, is:

$$\begin{cases} 5\pi/12 + 3t/2 & t \in (0, \pi/18] \\ 7\pi/12 - 3t/2 & t \in [\pi/18, \pi/6). \end{cases}$$

This angle is in $(\pi/3, \pi/2]$. Further note that the angle $5\pi/12 - t/2$ is in $(\pi/3, \pi/2)$. A calculation rules out Equation 1. Now of the remaining equations, the angle $\pi/12-t/2$ can only correspond to the angle $\pi/15$ in Equation 4. However, Equation 4 does not have any angles inside $(\pi/3, \pi/2)$. The same argument holds for the triple $(\pi/12 - t/2, \pi - t, 5\pi/12 + t/2)$.

The dual of the pair $(\pi/12 - t/2, 5\pi/12 + t/2)$ is $(11\pi/12 + t/2, 7\pi/12 - t/2)$ and the τ_Y-equation of the triples $(11\pi/12 + t/2, t, 7\pi/12 - t/2)$ and $(11\pi/12 + t/2, \pi - t, 7\pi/12 - t/2)$ have the same three angles (in $[0, \pi/2]$) as above.

A similar argument holds for for the triples associated with the pairs $(7\pi/12 - t/2, \pi/12 - t/2)$ and $(\pi/12 - t/2, 7\pi/12 - t/2)$.

Case 5: For $a = \pi/3 - t$, and $b = \pi/2$, the possibilities are:

$$(5\pi/12 - t/2, \pi/12 + t/2), (7\pi/12 + t/2, \pi/12 + t/2),$$
$$(\pi/12 + t/2, 5\pi/12 - t/2), (\pi/12 + t/2, 7\pi/12 + t/2),$$

and their duals. Note that the two last pairs are the symmetric pairs of the first two. As in previous arguments, it is enough to consider $\theta_y = t$ or $\pi - t$.

We first handle the triple (in angle notation) $(5\pi/12 - t/2, t, \pi/12 + t/2)$. Note that three of the angles in the τ_Y-equation of this point are: $5\pi/12-t/2$, $\pi/12 + 3t/2$, and $\pi/12 - t/2$. A calculation shows that Equation 1 is ruled out. Thus, $\pi/12 - t/2$ can only correspond to the angle $\pi/15$ in Equation 4, while $5\pi/12 - t/2$ is in $(\pi/3, \pi/2)$ which rules out Equation 4. The same holds for the triple $(5\pi/12-t/2, \pi-t, \pi/12+t/2)$ and those associated with the dual of $(5\pi/12-t/2, \pi/12+t/2)$.

We now consider the triples associated with the symmetric pair $(\pi/12+t/2, 5\pi/12-t/2)$. Consider the triple $(\pi/12+t/2, t, 5\pi/12-t/2)$. Three of the angles in the τ_Y-equation of this point are: $\pi/12+t/2$, $5\pi/12-3t/2$, and $5\pi/12+t/2$. One can check that Equation 1 is ruled out. Now $5\pi/12+t/2$ can only correspond to the angles $7\pi/15$ and $3\pi/7$ in Equations 3 and 5. That is, $t=\pi/10$ or $\pi/42$. But in either case, the angle $\pi/12+t/2$ will not appear in any of Equations 2-5. The same holds for the triple $(\pi/12+t/2, \pi-t, 5\pi/12-t/2)$ and those associated with the dual of $(\pi/12+t/2, 5\pi/12-t/2)$.

Similar arguments hold for triples associated with the pairs $(7\pi/12+t/2, \pi/12+t/2)$ and $(\pi/12+t/2, 7\pi/12+t/2)$.

Case 6: For $a=t$, and $b=\pi/2$, the possibilities are:

$$(\pi/4+t/2, \pi/4-t/2), (3\pi/4-t/2, \pi/4-t/2),$$
$$(\pi/4-t/2, \pi/4+t/2), (\pi/4-t/2, 3\pi/4-t/2),$$

and their duals. Note that the two last pairs are the symmetric pairs of the first two. As before, it is enough to consider $\theta_y=\pi/3-t$ or $2\pi/3+t$.

We first handle the triple (in angle notation) $(\pi/4+t/2, \pi/3-t, \pi/4-t/2)$. Note that three of the angles in the τ_Y-equation are: $\pi/4+t/2$, $7\pi/12-3t/2$, and $\pi/12-t/2$. It is clear that Equation 1 is ruled out. The angle $\pi/12-t/2$ can only correspond to the angle $\pi/15$ in Equation 4, while $7\pi/12-3t/2$, when rewritten in $[0, \pi/2]$, corresponds to an angle in $(\pi/3, \pi/2]$, which rules out Equation 4. The same holds for the triple $(\pi/4+t/2, 2\pi/3+t, \pi/4-t/2)$ and those associated with the dual of $(\pi/4+t/2, \pi/4-t/2)$.

Consider the symmetric pair $(\pi/4-t/2, \pi/4+t/2)$. For the triple $(\pi/4-t/2, \pi/3+t, \pi/4+t/2)$, three of the angles (rewritten in $[0\pi/2]$) in the τ_Y-equation are: $\pi/4-t/2$, $\pi/12+t/2$, and $5\pi/12-3t/2$. Note that both $\pi/4-t/2$ and $\pi/12+t/2$ are less than $\pi/4$, leaving only Equations 1 and 4. However, a calculation rules out Equation 4, while Equation 1 is clearly ruled out. The same holds for the triple $(\pi/4-t/2, 2\pi/3+t, \pi/4+t/2)$ and those associated with the dual of $(\pi/4-t/2, \pi/4+t/2)$.

Similar arguments hold for triples associated with the pairs $(3\pi/4-t/2, \pi/4-t/2)$ and $(\pi/4-t/2, 3\pi/4-t/2)$.

For the other three equations, we simply list the solutions $(\theta_x, \theta_y, \theta_z)$ in their respective category. For simplicity, we do not list the solutions correspond to the Pin(2) representations and those with one global coordinate equal to zero. Note that the complete set of solutions include the dual, symmetric, and dual-symmetric solutions in the first and third coordinates to those listed below.

$$\cos(\pi/5)-\cos(2\pi/5)-\cos(\pi/3)=0.$$

1) $(k\geq 2)$ $(\pi/5, \pi/3, 3\pi/5)$, $(\pi/5, 2\pi/3, 2\pi/5)$.

2) Triples appearing in Table 1: $(\pi/5, 2\pi/3, 3\pi/5)$, $(\pi/5, \pi/3, 2\pi/5)$.
3) Triples with an angle not appearing in Theorem 5.4:
$(3\pi/10, \pi/3, \pi/10)$, $(3\pi/10, 2\pi/3, \pi/10)$, $(7\pi/10, \pi/3, \pi/10)$, $(7\pi/10, 2\pi/3, \pi/10)$, $(7\pi/30, 2\pi/5, 13\pi/30)$, $(7\pi/30, 3\pi/5, 13\pi/30)$, $(7\pi/30, 2\pi/5, 17\pi/30)$, $(7\pi/30, 3\pi/5, 17\pi/30)$, $(7\pi/20, 2\pi/5, 3\pi/20)$, $(7\pi/20, 3\pi/5, 3\pi/20)$, $(7\pi/20, \pi/3, 3\pi/20)$, $(7\pi/20, 2\pi/3, 3\pi/20)$, $(3\pi/20, 2\pi/5, 13\pi/20)$, $(3\pi/20, 3\pi/5, 13\pi/20)$, $(3\pi/20, \pi/3, 13\pi/20)$, $(3\pi/20, 2\pi/3, 13\pi/20)$, $(11\pi/30, \pi/5, \pi/30)$, $(11\pi/30, 4\pi/5, \pi/30)$, $(19\pi/30, \pi/5, \pi/30)$, $(19\pi/30, 4\pi/5, \pi/30)$, $(9\pi/20, \pi/5, \pi/20)$, $(9\pi/20, 4\pi/5, \pi/20)$, $(9\pi/20, \pi/3, \pi/20)$, $(9\pi/20, 2\pi/3, \pi/20)$, $(11\pi/20, \pi/5, \pi/20)$, $(11\pi/20, 4\pi/5, \pi/20)$, $(11\pi/20, \pi/3, \pi/20)$, $(11\pi/20, 2\pi/3, \pi/20)$, $(5\pi/12, \pi/5, \pi/12)$, $(5\pi/12, 4\pi/5, \pi/12)$, $(5\pi/12, 2\pi/5, \pi/12)$, $(5\pi/12, 3\pi/5, \pi/12)$, $(7\pi/12, \pi/5, \pi/12)$, $(7\pi/12, 4\pi/5, \pi/12)$, $(7\pi/12, 2\pi/5, \pi/12)$, $(7\pi/12, 3\pi/5, \pi/12)$.
Triples (x, y, z) whose τ_Y-equation does not correspond to any equation in Theorem 5.4:
$(4\pi/15, 2\pi/5, \pi/15)$, $(4\pi/15, 3\pi/5, \pi/15)$, $(11\pi/15, 2\pi/5, \pi/15)$, $(11\pi/15, 3\pi/5, \pi/15)$, $(2\pi/15, \pi/5, 8\pi/15)$, $(2\pi/15, 4\pi/5, 8\pi/15)$, $(2\pi/15, \pi/5, 7\pi/15)$, $(2\pi/15, 4\pi/5, 7\pi/15)$.

Note that the categories above are not mutually exclusive.

$$\cos(\pi/5) - \cos(\pi/15) + \cos(4\pi/15) - \cos(\pi/3) = 0.$$

1) $(k \geq 2)$ $(\pi/3, 4\pi/5, 2\pi/5)$, $(3\pi/5, \pi/5, \pi/3)$.
2) Triples appearing in Table 1: $(\pi/3, \pi/5, 2\pi/5)$, $(\pi/3, \pi/3, 2\pi/5)$, $(\pi/3, 2\pi/3, 2\pi/5)$, $(3\pi/5, 4\pi/5, \pi/3)$, $(3\pi/5, \pi/3, \pi/3)$.
3) Triples with an angle not appearing in Theorem 5.4:
$(11\pi/30, 4\pi/15, 17\pi/30)$, $(11\pi/30, 11\pi/15, 17\pi/30)$, $(11\pi/30, \pi/3, 17\pi/30)$, $(11\pi/30, 2\pi/3, 17\pi/30)$, $(11\pi/30, 4\pi/15, 13\pi/30)$, $(11\pi/30, 11\pi/15, 13\pi/30)$, $(11\pi/30, \pi/3, 13\pi/30)$, $(11\pi/30, 2\pi/3, 13\pi/30)$, $(7\pi/30, \pi/15, \pi/30)$, $(7\pi/30, 14\pi/15, \pi/30)$, $(7\pi/30, \pi/3, \pi/30)$, $(7\pi/30, 2\pi/3, \pi/30)$, $(23\pi/30, \pi/15, \pi/30)$, $(23\pi/30, 14\pi/15, \pi/30)$, $(23\pi/30, \pi/3, \pi/30)$, $(23\pi/30, 2\pi/3, \pi/30)$, $(7\pi/30, \pi/15, 13\pi/30)$, $(7\pi/30, 14\pi/15, 13\pi/30)$, $(7\pi/30, 4\pi/15, 13\pi/30)$, $(7\pi/30, 11\pi/15, 13\pi/30)$, $(7\pi/30, \pi/15, 17\pi/30)$, $(7\pi/30, 14\pi/15, 17\pi/30)$, $(7\pi/30, 4\pi/15, 17\pi/30)$, $(7\pi/30, 11\pi/15, 17\pi/30)$, $(\pi/6, \pi/5, \pi/10)$, $(\pi/6, 4\pi/5, \pi/10)$, $(\pi/6, \pi/3, \pi/10)$, $(\pi/6, 2\pi/3, \pi/10)$, $(5\pi/6, \pi/5, \pi/10)$, $(5\pi/6, 4\pi/5, \pi/10)$, $(5\pi/6, \pi/3, \pi/10)$, $(5\pi/6, 2\pi/3, \pi/10)$, $(3\pi/10, \pi/5, 11\pi/30)$, $(3\pi/10, 4\pi/5, 11\pi/30)$, $(3\pi/10, 4\pi/15, 11\pi/30)$, $(3\pi/10, 11\pi/15, 11\pi/30)$, $(3\pi/10, \pi/5, 19\pi/30)$, $(3\pi/10, 4\pi/5, 19\pi/30)$, $(3\pi/10, 4\pi/15, 19\pi/30)$,

$(3\pi/10, 11\pi/15, 19\pi/30)$, $(3\pi/10, \pi/5, \pi/30)$, $(3\pi/10, 4\pi/5, \pi/30)$, $(3\pi/10, \pi/15, \pi/30)$, $(3\pi/10, 14\pi/15, \pi/30)$, $(7\pi/10, \pi/5, \pi/30)$, $(7\pi/10, 4\pi/5, \pi/30)$, $(7\pi/10, \pi/15, \pi/30)$, $(7\pi/10, 14\pi/15, \pi/30)$.
Triples (x, y, z) whose τ_Y-equation does not correspond to any equation in Theorem 5.4:
$(2\pi/15, 4\pi/15, \pi/15)$, $(2\pi/15, 11\pi/15, \pi/15)$, $(2\pi/15, \pi/3, \pi/15)$, $(2\pi/15, 2\pi/3, \pi/15)$, $(13\pi/15, 4\pi/15, \pi/15)$, $(13\pi/15, 11\pi/15, \pi/15)$, $(13\pi/15, \pi/3, \pi/15)$, $(13\pi/15, 2\pi/3, \pi/15)$, $(4\pi/15, \pi/15, 7\pi/15)$, $(4\pi/15, 14\pi/15, 7\pi/15)$, $(4\pi/15, \pi/3, 7\pi/15)$, $(4\pi/15, 2\pi/3, 7\pi/15)$, $(4\pi/15, \pi/15, 8\pi/15)$, $(4\pi/15, 14\pi/15, 8\pi/15)$, $(4\pi/15, \pi/3, 8\pi/15)$, $(4\pi/15, 2\pi/3, 8\pi/15)$, $(\pi/5, \pi/15, 7\pi/15)$, $(\pi/5, 14\pi/15, 7\pi/15)$, $(\pi/5, 4\pi/15, 2\pi/15)$, $(\pi/5, 11\pi/15, 2\pi/15)$, $(8\pi/15, \pi/15, \pi/5)$, $(4\pi/5, 11\pi/15, 2\pi/15)$, $(4\pi/5, 4\pi/15, 2\pi/15)$, $(8\pi/15, 14\pi/15, \pi/5)$.
Triples with infinite orbits, by Proposition 5.3:
$(\pi/5, \pi/5, 2\pi/15)$, $(\pi/5, 4\pi/5, 2\pi/15)$, $(4\pi/5, \pi/5, 2\pi/15)$, $(4\pi/5, 4\pi/5, 2\pi/15)$, $(3\pi/5, 2\pi/3, \pi/3)$, $(4\pi/15, \pi/15, \pi/15)$, $(4\pi/15, 14\pi/15, \pi/15)$, $(4\pi/15, 4\pi/15, \pi/15)$, $(4\pi/15, 11\pi/15, \pi/15)$, $(11\pi/15, \pi/15, \pi/15)$, $(11\pi/15, 14\pi/15, \pi/15)$, $(11\pi/15, 4\pi/15, \pi/15)$, $(11\pi/15, 11\pi/15, \pi/15)$, $(\pi/5, \pi/5, 7\pi/15)$, $(\pi/5, 4\pi/5, 7\pi/15)$, $(8\pi/15, \pi/5, \pi/5)$, $(8\pi/15, 4\pi/5, \pi/5)$.

$$-\cos(2\pi/5) + \cos(2\pi/15) - \cos(7\pi/15) - \cos(\pi/3) = 0.$$

1) $(k \geq 2)$ $(\pi/5, 3\pi/5, \pi/3)$, $(\pi/5, 2\pi/3, \pi/3)$, $(2\pi/3, 2\pi/5, \pi/5)$, $(2\pi/3, \pi/3, \pi/5)$.
2) Triples appearing in Table 1:
$(\pi/5, 2\pi/5, \pi/3)$, $(2\pi/3, 2\pi/3, \pi/5)$ $(\pi/5, \pi/3, \pi/3)$, $(2\pi/3, 3\pi/5, \pi/5)$.
3) Triples with an angle not appearing in Theorem 5.4:
$(7\pi/30, 7\pi/15, 19\pi/30)$, $(7\pi/30, \pi/3, 19\pi/30)$, $(7\pi/30, 8\pi/15, 19\pi/30)$, $(7\pi/30, 2\pi/3, 19\pi/30)$, $(7\pi/30, 7\pi/15, 11\pi/30)$, $(7\pi/30, \pi/3, 11\pi/30)$, $(7\pi/30, 8\pi/15, 11\pi/30)$, $(7\pi/30, 2\pi/3, 11\pi/30)$, $(13\pi/30, 2\pi/15, \pi/30)$, $(13\pi/30, \pi/3, \pi/30)$, $(13\pi/30, 13\pi/15, \pi/30)$, $(13\pi/30, 2\pi/3, \pi/30)$, $(17\pi/30, 2\pi/15, \pi/30)$, $(17\pi/30, \pi/3, \pi/30)$, $(17\pi/30, 13\pi/15, \pi/30)$, $(17\pi/30, 2\pi/3, \pi/30)$, $(11\pi/30, 2\pi/15, \pi/30)$, $(11\pi/30, 7\pi/15, \pi/30)$, $(11\pi/30, 13\pi/15, \pi/30)$, $(11\pi/30, 8\pi/15, \pi/30)$, $(19\pi/30, 2\pi/15, \pi/30)$, $(19\pi/30, 7\pi/15, \pi/30)$, $(19\pi/30, 13\pi/15, \pi/30)$, $(19\pi/30, 8\pi/15, \pi/30)$, $(3\pi/10, 2\pi/5, \pi/6)$, $(3\pi/10, \pi/3, \pi/6)$, $(3\pi/10, 3\pi/5, \pi/6)$, $(3\pi/10, 2\pi/3, \pi/6)$, $(7\pi/10, 2\pi/5, \pi/6)$, $(7\pi/10, \pi/3, \pi/6)$, $(7\pi/10, 3\pi/5, \pi/6)$, $(7\pi/10, 2\pi/3, \pi/6)$, $(7\pi/30, 2\pi/5, \pi/10)$, $(7\pi/30, 7\pi/15, \pi/10)$, $(7\pi/30, 3\pi/5, \pi/10)$, $(7\pi/30, 8\pi/15, \pi/10)$, $(23\pi/30, 2\pi/5, \pi/10)$, $(23\pi/30, 7\pi/15, \pi/10)$, $(23\pi/30, 3\pi/5, \pi/10)$, $(23\pi/30, 8\pi/15, \pi/10)$, $(\pi/10, 2\pi/5, 17\pi/30)$, $(\pi/10, 2\pi/15, 17\pi/30)$, $(\pi/10, 3\pi/5, 17\pi/30)$, $(\pi/10, 13\pi/15, 17\pi/30)$, $(\pi/10, 2\pi/5, 13\pi/30)$, $(\pi/10, 2\pi/15, 13\pi/30)$, $(\pi/10, 3\pi/5, 13\pi/30)$, $(\pi/10, 13\pi/15, 13\pi/30)$.

Triples (x, y, z) whose τ_Y-equation does not correspond to any equation in Theorem 5.4:
$(4\pi/15, 7\pi/15, 2\pi/15)$, $(4\pi/15, \pi/3, 2\pi/15)$, $(4\pi/15, 8\pi/15, 2\pi/15)$, $(4\pi/15, 2\pi/3, 2\pi/15)$, $(11\pi/15, 7\pi/15, 2\pi/15)$, $(11\pi/15, \pi/3, 2\pi/15)$, $(11\pi/15, 8\pi/15, 2\pi/15)$, $(11\pi/15, 2\pi/3, 2\pi/15)$, $(\pi/15, 2\pi/15, 7\pi/15)$, $(\pi/15, \pi/3, 7\pi/15)$, $(\pi/15, 13\pi/15, 7\pi/15)$, $(\pi/15, 2\pi/3, 7\pi/15)$, $(\pi/15, 2\pi/15, 8\pi/15)$, $(\pi/15, \pi/3, 8\pi/15)$, $(\pi/15, 13\pi/15, 8\pi/15)$, $(\pi/15, 2\pi/3, 8\pi/15)$, $(4\pi/15, 7\pi/15, 2\pi/5)$, $(4\pi/15, 8\pi/15, 2\pi/5)$, $(4\pi/15, 7\pi/15, 3\pi/5)$, $(4\pi/15, 8\pi/15, 3\pi/5)$, $(2\pi/5, 2\pi/15, \pi/15)$, $(2\pi/5, 13\pi/15, \pi/15)$, $(3\pi/5, 2\pi/15, \pi/15)$, $(3\pi/5, 13\pi/15, \pi/15)$.

Triples with infinite orbits, by Proposition 5.3:
$(4\pi/15, 3\pi/5, 3\pi/5)$, $(4\pi/15, 2\pi/5, 3\pi/5)$, $(3\pi/5, 3\pi/5, \pi/15)$, $(3\pi/5, 2\pi/5, \pi/15)$, $(2\pi/5, 3\pi/5, \pi/15)$, $(2\pi/5, 2\pi/5, \pi/15)$, $(2\pi/15, 13\pi/15, 8\pi/15)$, $(2\pi/15, 13\pi/15, 7\pi/15)$, $(4\pi/15, 3\pi/5, 2\pi/5)$, $(4\pi/15, 2\pi/5, 2\pi/5)$, $(2\pi/15, 2\pi/15, 7\pi/15)$, $(2\pi/15, 2\pi/15, 8\pi/15)$, $(2\pi/15, 7\pi/15, 7\pi/15)$, $(2\pi/15, 8\pi/15, 7\pi/15)$, $(2\pi/15, 8\pi/15, 8\pi/15)$, $(2\pi/15, 7\pi/15, 8\pi/15)$.

References

[1] J.H. Conway and A.J. Jones, *Trigonometric diophantine equations* (*on vanishing sums of roots of unity*), Acta Arithmetica, **XXX** (1976), 229-240.

[2] W.M. Goldman, *Ergodic theory on moduli spaces*, Ann. of Math., **146** (1997), 475-507.

[3] ______, *The symplectic nature of fundamental groups of surfaces*, Adv. Math., **54** (1984), 200-225.

[4] B. Hasselblatt and A. Katok, *Introduction to the Modern Theory of Dynamical Systems*, Cambridge University Press, 1995.

[5] J. Huebschmann, *Symplectic and Poisson Structures on Certain Moduli Spaces*, preprint, hep-th/9312112, Pub. IRMA, Lille, 1993.

[6] D. Mumford and J. Fogarty, *Geometric Invariant Theory*, 2nd ed., Springer, Berlin, 1982.

[7] P.E. Newstead, *Introductions to Moduli Problems and Orbit Spaces*, Springer-Verlag, 1978.

Received September 2, 1998 and revised January 28, 1999.

Penn State Erie-The Behrend College
Erie, PA 16563
E-mail address: jpp@vortex.bd.psu.edu

University of Massachusetts
Amherst, MA 01003-4515
E-mail address: xia@math.umass.edu

PACIFIC JOURNAL OF MATHEMATICS
Vol. 193, No. 2, 2000

ON THE ACTION SPECTRUM FOR CLOSED SYMPLECTICALLY ASPHERICAL MANIFOLDS

Matthias Schwarz

Symplectic homology is studied on closed symplectic manifolds where the class of the symplectic form and the first Chern class vanish on the second homotopy group. Critical values of the action functional are associated to cohomology classes of the manifold. Those lead to continuous sections in the action spectrum bundle. The action of the cohomology ring via the cap-action and the pants-product on the set of critical values is studied and a bi-invariant metric on the group of Hamiltonian symplectomorphisms is defined and analyzed. Finally, a relative symplectic capacity is defined which is bounded below by the π_1-sensitive Hofer-Zehnder capacity. As an application it is proven that a Hamiltonian automorphism whose support has finite such capacity has infinitely many nontrivial geometrically distinct periodic points.

1. Overview of the Results.

It is a well-known problem in symplectic geometry and Hamiltonian dynamics to study the existence of fixed points of Hamiltonian diffeomorphisms and to relate them to invariants from symplectic topology. The aim of this paper is to study the existence of "homologically visible" critical values of the action functional and their dependence on the Hamiltonian automorphism in the case of symplectically aspherical closed manifolds. The methods are provided by the theory of Floer homology. The initial aim of this paper is to consider a version of Floer homology refined by a filtration via the action functional. This version has been introduced by Floer and Hofer as so-called symplectic homology for open subsets of $\mathbb{R}^{2n}$, [**4**]. Here we study its generalization for closed symplectic manifolds which satisfy the property that

$$\text{(A)} \qquad \omega_{|\pi_2(M)} = 0 \quad \text{and} \quad c_{1|\pi_2} = 0.$$

This condition forms the simplest case for which Floer homology was studied initially. Note that, very recently, examples have been constructed of closed manifolds satisfying (A) but having nontrivial second homotopy group, [**6**], [**9**]. Observe that a closed symplectic manifold satisfying (A) is necessarily non-simply connected and $\pi_1(M)$ contains elements of infinite order. From now on we call a symplectic manifold (M,ω) satisfying (A) **symplectically aspherical**.

In terms of Floer theory, condition (A) implies that for any given Hamiltonian $H\colon [0,1]\times M \to \mathbb{R}$, the associated action functional $\mathcal{A}_H$ on the space of free contractible loops is real-valued. Having in mind that the full Floer homology can already be uniquely associated to the time-1-map ϕ_H^1 generated by the Hamiltonian H one naturally asks how the action spectrum depends on the choice of H. By the **action spectrum** one denotes the set of critical values of the action functional. In fact, it is easy to prove that any two Hamiltonian functions H and K generating the same automorphism ϕ and homotopic to each other with respect to this property have the same action associated to the fixed points provided that they are normalized as follows

$$\text{(norm)} \qquad \int_M H(t,\cdot)\omega^n = 0 \quad \text{for all } t\,.$$

However, in general, two different homotopy classes of Hamiltonians generating the same time-1-map might have action spectrum differing by a quantity $I([H][K]^{-1})$ associated to the difference of the homotopy classes. In fact, one can define a group homomorphism

$$I\colon \pi_1(\mathrm{Ham}(M,\omega)) \to \mathbb{R}$$

describing this obstruction. Here, $\mathrm{Ham}(M,\omega)$ is the group of Hamiltonian automorphisms. For a definition of I one considers $g \in \pi_1(\mathrm{Ham})$ and glues the trivial symplectic fibre bundles $D^2\times M$ with reversed orientations of the disk along their boundaries using a loop in $\mathrm{Ham}(M,\omega)$ representing g. The resulting symplectic fibre bundle E over S^2 carries the so-called coupling form ω_E defined by ω on $D^2\times M$ and one defines

$$I(g) = \int_{S^2} s^*\omega_E$$

for any section s of E. It follows from Floer theory that such a section exists and (A) implies that $I(g)\in\mathbb{R}$ does not depend on the choice of s. As an immediate consequence of a theorem by P. Seidel, [**22**], we observe:

Theorem 1.1. *The obstruction homomorphism* $I\colon \pi_1(\mathrm{Ham}(M,\omega)) \to \mathbb{R}$ *vanishes if* (M,ω) *is a closed symplectically aspherical manifold.*

We obtain for each $\phi\in\mathrm{Ham}(M,\omega)$ a well-defined action spectrum Σ_ϕ which as a whole provides the **action spectrum bundle** $\Sigma\to\mathrm{Ham}(M,\omega)$. The main result of this paper is a construction of continuous sections of this bundle associated to cohomology classes of M. Here the topology on $\mathrm{Ham}(M,\omega)$ is given by the bi-invariant Hofer-metric

$$(1) \qquad d_H\colon \mathrm{Ham}\times\mathrm{Ham}\to[0,\infty), \quad d_H(\phi,\mathrm{id}) = \inf\{\,\|H\|\mid \phi=\phi_H^1\,\}$$

where $\|H\| = \int_0^1 \operatorname{osc}_M H(t,\cdot)dt$.

The result of this paper was predominantly motivated by a remark by M. Bialy and L. Polterovich in [**1**], cf. 1.5.B. They already suggested a generalization of Hofer's minimax selector γ from the theory on $\mathbb{R}^{2n}$ to more general manifolds using Floer theory and an idea by C. Viterbo. An intrinsic motivation was the problem of infinitely many periodic points of a Hamiltonian automorphism, which is partly analyzed by Theorem 1.4 below.

Theorem 1.2. *Let (M,ω) be symplectically aspherical. Then for each nonzero cohomology class $\alpha \in H^*(M)$ there exists a section $c(\alpha)$ of the action spectrum bundle $\Sigma \to \operatorname{Ham}(M,\omega)$ which is continuous with respect to the metric from* (1). *These sections satisfy:*

1) $c(\lambda\alpha) = c(\alpha)$ *for all* $\lambda \in \mathbb{R}$ *and* $\alpha \in H^*(M)$ *with* $\lambda\alpha \neq 0$.
2) $c([M]) \leq c(\alpha) \leq c(1)$ *for all* $0 \neq \alpha \in H^*(M)$ *where* $[M] = [\omega^n] \in H^{2n}(M)$ *and* $1 \in H^0(M)$ *are the canonical generators. If* $\alpha \in H^k(M)$ *for* $0 < k < 2n$ *then we have strict inequality over the regular automorphisms.*
3) $c(1;\phi) - c([M];\phi) \leq d_H(\phi, \mathrm{id}) = \inf\{\, \|H\| \mid \phi = \phi_H^1 \,\}$, $\phi \in \operatorname{Ham}(M,\omega)$.
4) *If* $\alpha \cup \beta \neq 0$ *then* $c(\alpha \cup \beta; \psi \circ \phi) \leq c(\alpha;\phi) + c(\beta;\psi)$ *for all* $\phi, \psi \in \operatorname{Ham}(M,\omega)$.
5) $c([M];\phi) = -c(1;\phi^{-1})$ *for all* $\phi \in \operatorname{Ham}(M,\omega)$.

Note that the obstruction homomorphism I can be viewed as the *monodromy map* of the action spectrum bundle with respect to any of the sections $c(\alpha)$. This observation should be relevant for a generalization to the non-aspherical situation which will be studied in a separate paper.

The construction of the sections $c(\alpha)$ is based on Viterbo's idea for generating functions defined in the context of a cotangent bundle $M = T^*P$, cf. [**23**]. There, $c(\alpha)$ is a critical value of a generating function. In our case of closed symplectically aspherical manifolds we use the construction of an explicit isomorphism between Floer homology and standard cohomology of M which was introduced in [**16**]. A detailed description of this isomorphism will appear in [**18**]. The critical values of the Hamiltonian action functional then are defined as the infimum of all action levels below which the specified Floer homology class is still nontrivial.

The problem of finding a nontrivial continuous section in the action spectrum bundle was first introduced and treated by Hofer and Zehnder [**7**] in the context of open subsets of $\mathbb{R}^{2n}$ and a capacity for such symplectic manifolds. This so-called Hofer-Zehnder capacity, denoted by c_{HZ} below, is defined in terms of periodic solutions for compactly supported Hamiltonians on $\mathbb{R}^{2n}$. Using a variational minimax method for the associated action functional they constructed a so-called selector $\gamma\colon \operatorname{Ham}_{\mathrm{cpt}} \to \mathbb{R}$ which is continuous in the Hofer metric. For a detailed treatment see [**8**]. It is this continuous selector γ in the context of compactly supported Hamiltonians on $\mathbb{R}^{2n}$ which Theorem 1.2 replaces by the family of continuous sections $c(\alpha)$ in the case of closed symplectically aspherical manifolds. In the case of the open symplectic manifold given by a cotangent bundle, the result corresponding to Theorem 1.2 has been obtained by Y.-G. Oh in [**14**], [**15**]. There the finite-dimensional concept of C. Viterbo from [**23**] finding critical levels of generating functions has been applied in Floer homological context replacing the action functional for the generating function.

As a consequence of Theorem 1.2 we can consider the difference between $c(1;\phi)$ and $c([M];\phi)$ which is a continuous function of $\phi \in \mathrm{Ham}(M,\omega)$.

Theorem 1.3. *Given a closed symplectically aspherical manifold* (M,ω) *there exists a unique nonnegative function*

$$\gamma\colon \mathrm{Ham}(M,\omega) \to \mathbb{R}_+, \quad \textit{such that } \gamma(\phi) = c(1;\phi) - c([M];\phi)$$

for all $\phi \in \mathrm{Ham}(M,\omega)$ *which is continuous in Hofer's metric. Moreover,* $d_\gamma(\phi,\psi) = \gamma(\phi\psi^{-1})$ *is a bi-invariant metric on* $\mathrm{Ham}(M,\omega)$ *bounded above by Hofer's metric.*

Since γ measures the maximal action difference of homologically visible 1-periodic solutions of the Hamilton equation, this difference can be related to the oscillation of H if there exists no non-constant 1-periodic solutions. We show in Theorem 5.11 below that γ coincides with the Hofer distance $\|\cdot\|$ for quasi-autonomous Hamiltonians which are admissible in the sense that they admit no non-constant contractible periodic solutions which are "fast", i.e., of period $T \le 1$. This shows the close relationship between γ and the Hofer-Zehnder capacity defined via the maximal oscillation of an autonomous Hamiltonian admitting no non-constant periodic solutions, cf. [**8**]. Here, this capacity has to be refined as a π_1-sensitive capacity c^o_{HZ} with respect to the nontrivial fundamental group. We consider the larger set of admissible Hamiltonians which admit no non-constant *contractible* 1-periodic solutions. Based on γ we can define a relative capacity for subsets of (M,ω) which is monotone and invariant under global automorphisms of (M,ω). Defining a γ-capacity for subsets of M by

$$\begin{aligned} c_\gamma(A) = \sup\{\, \gamma(\phi) \mid \exists H \text{ s.t. } \phi = \phi^1_H,& \\ \operatorname{supp} X_H(t,\cdot) \subset A \text{ for all } t \in [0,1]\,\},& \end{aligned}$$

we have the estimate

$$c_{\mathrm{HZ}}(U) \le c^o_{\mathrm{HZ}}(U) \le c_\gamma(U) \le 2e(U)$$

for all open subsets $U \subset M$. Here, $e(U)$ is the displacement energy as considered in [**11**]. The idea of using symplectic homology in order to construct symplectic capacities was first carried out in the series of papers by Floer and Hofer et al., cf. [**4**], [**5**].

In the context of such a capacity based on the Floer homological approach we prove the following conditional existence result for infinitely many periodic points:

Theorem 1.4. *Assume that $\phi \in \mathrm{Ham}(M,\omega)$ admits a Hamiltonian function H such that $\phi = \phi_H^1 \neq \mathrm{id}$ and there exists a uniform bound on the γ-capacity of the support for all ϕ_H^t, $t \in [0,1]$,*

$$c_\gamma(\operatorname{supp}\phi_H^t) \leq m < \infty, \quad \textit{for all } t \in [0,1].$$

Then ϕ has infinitely many geometrically distinct non-constant periodic points corresponding to contractible solutions.

There are clearly examples for such a bounded capacity of the support of ϕ_H^t, for example if the support can be separated from itself by a Hamiltonian isotopy. However, in general, in view of Theorem 1.3 it is obvious that such a uniform bound cannot exist if $\gamma(\phi^n) \to \infty$ for $n \to \infty$. Hence our method of finding infinitely many periodic points via the action spectrum is closely related to the question of the diameter of $\mathrm{Ham}(M,\omega)$ in the Hofer-metric. It is in fact conjectured that this diameter should always be infinite. Using the methods developed in this paper we are also able to reproduce several known examples of such infinite diameter.

It is clearly an interesting question how far one obtains similar results if $\omega_{|\pi_2} \neq 0$. This will be studied in a sequel.

Organization of the paper. In Section 2 we present the construction of symplectic homology on a closed symplectically aspherical manifold for the purpose of this paper. For a more thorough exposition in the case of $\mathbb{R}^{2n}$ see [**4**]. We associate critical levels of the Hamiltonian action functional to given cohomology classes of M using the explicit isomorphism between standard cohomology of M and Floer homology for a given Hamiltonian function H. It is shown that these homologically visible critical levels depend continuously on H in the C^0-norm.

In Section 3 we analyze more closely the action spectrum, i.e., the critical levels of the action functional for a given Hamiltonian and we show that these levels are already uniquely associated to the Hamiltonian equivalence class $\phi = [H] \in \widetilde{\mathrm{Ham}}(M,\omega)$ and do not depend on the specific Hamiltonian provided that we use a suitable normalization. We consider a more intrinsic version of Floer homology. Namely, Floer homology is already canonically associated to the automorphism $\phi \in \widetilde{\mathrm{Ham}}_{\mathrm{reg}}(M,\omega)$ such that we still have the filtration by the action functional. This requires methods of J-holomorphic sections in symplectic fibre bundles over Riemann surface which are used by P. Seidel in [**22**], where a monodromy action of $\pi_1(\mathrm{Ham}(M,\omega))$ on this intrinsically defined Floer homology is analyzed.

Using these methods, we obtain in Section 4 a concise proof for the "sharp energy estimate" for the pair-of-pants multiplication on Floer homology.

This estimate expresses the compatibility of symplectic homology with this multiplication provided that it is viewed as a product

$$HF_*(\phi) \times HF_*(\psi) \to HF_*(\phi\psi)\,.$$

The result of this estimate is the nontrivial fact that the nonnegative function $\gamma\colon \mathrm{Ham}(M,\omega) \to \mathbb{R}$ satisfies the triangle inequality. As a consequence of the pair-of-pants multiplication we also obtain the continuity of the sections $c(\alpha)$ with respect to Hofer's metric.

In the last section we draw conclusions from the constructed metric γ. We define the γ-capacity, compare it to the π_1-sensitive Hofer-Zehnder capacity and deduce the results on the infinite number of periodic points.

Acknowledgments. This preprint was largely written during the author's stay at the Max-Planck Institute for Mathematics in the Sciences, Leipzig, and the Forschungsinstitut für Mathematik, ETH Zürich. He thanks them for their warm hospitality. The author also thanks Leonid Polterovich for many helpful discussions and Dusa McDuff for pointing out that the vanishing of the monodromy I follows from Seidel's theorem. Moreover, he thanks Y.-G. Oh for his comments.

2. Symplectic homology and critical levels for the Hamiltonian action.

2.1. Symplectic homology. Let (M,ω) be a closed symplectically aspherical manifold, i.e., satisfying condition (A). Throughout this section we assume $H\colon S^1 \times M \to \mathbb{R}$ to be a regular Hamiltonian, i.e., every fixed point $x \in \mathrm{Fix}\,\phi_H^1$ is non-degenerate, where $\phi_H^1\colon \mathbb{R} \times M \to M$ is the time-1-map for the flow of the non-autonomous Hamiltonian vector field X_H defined by

$$\omega(X_H,\cdot) = -dH\,.$$

Denoting the space of contractible free loops by $\Omega^o(M) \subset C^\infty(\mathbb{R}/\mathbb{Z}, M)$ we define the set of 1-periodic contractible solutions of the Hamilton equation by

$$\mathcal{P}_1(H) = \{\, x \in \Omega^o(M) \,|\, \dot{x}(t) = X_H(t,x)\,\}.$$

We have $\mathcal{P}_1(H) = \mathrm{Crit}\,\mathcal{A}_H$ for the action functional

$$\mathcal{A}_H(x) = \int_{D^2} \bar{x}^*\omega - \int_{S^1} H(t,x)dt$$

where $\bar{x}\colon D^2 \to M$ is any extension of x to the unit disk. Note that $\mathcal{A}_H$ is real-valued due to assumption (A). In view of the same condition we have the integral grading by the Conley-Zehnder index $\mu\colon \mathcal{P}_1(H) \to \mathbb{Z}$, where we choose the normalization such that for an autonomous C^2-small Morse function H we have $\mu(x) = n - \mu_{\mathrm{Morse}}(x)$ for stationary $x \in \mathcal{P}_1(H) = \mathrm{Crit}\,H$, cf. [**17**].

Given a generic ω-compatible S^1-dependent almost complex structure $J(t,p)$, $(t,p) \in S^1 \times M$, i.e., $\omega \circ (\mathrm{id} \times J) = g_J$ is a Riemannian metric on $TM \to S^1 \times M$, we obtain the moduli spaces of Floer trajectories with the component-wise structure of a smooth manifold

$$\mathcal{M}_{y,x}(J,H) = \Big\{\, u\colon \mathbb{R} \times S^1 \to M \,|\partial_s u + J(t,u)\big(\partial_t u - X_H(t,u)\big) = 0, \\ u(-\infty,\cdot) = y,\ u(+\infty,\cdot) = x \,\Big\}$$

where $u(\pm\infty,\cdot)$ denotes the uniform limit in $C^0(S^1,M)$ as $s \to \pm\infty$. A standard computation in Floer theory is:

Lemma 2.1. *The dimension satisfies* $\dim \mathcal{M}_{y,x}(J,H) = \mu(x) - \mu(y)$ *and for* $u \in \mathcal{M}_{y,x}(J,H)$ *we have*

$$E(u) \overset{\mathsf{def}}{=} \int_{-\infty}^{\infty} \int_0^1 |\partial_s u|^2 ds dt \;=\; \mathcal{A}_H(x) - \mathcal{A}_H(y) \geq 0.$$

In particular, if the flow energy $E(u)$ *vanishes then* $x \equiv y$ *and* $u(s,t) = x(t)$ *for all* $(s,t) \in \mathbb{R} \times S^1$.

Heuristically, in view of the chosen sign conventions, the Floer trajectories correspond to the positive gradient flow of $\mathcal{A}_H$.

The full Floer homology is based on the chain complex obtained from the $\mathbb{Z}$-module $C_*(H) = \mathrm{Crit}_* \mathcal{A}_H \otimes \mathbb{Z}$ which is graded by μ. The boundary operator is defined as

$$\partial\colon C_k(H) \to C_{k-1}(H),$$
$$\partial x = \sum_{\mu(y)=k-1} \#_{\mathsf{alg}} \widehat{\mathcal{M}}_{y,x}(J,H)\, y,$$

where $\#_{\mathsf{alg}}$ denotes counting the finitely many unparameterized Floer trajectories with sign determined by a coherent orientation, cf. [**3**]. Here, $\widehat{\mathcal{M}}_{y,x} = \mathcal{M}_{y,x}/\mathbb{R}$ is the space of unparameterized trajectories after dividing out the free $\mathbb{R}$-action from shifting the trajectories in the s-variable. Floer's central theorem is that $\partial \circ \partial = 0$ and that the thus defined homology

$$HF_*(J,H) = H_*(C_*(H), \partial(J,H))$$

is canonically isomorphic to the singular cohomology of M.

Symplectic homology was developed by Floer and Hofer, cf. [**4**] using the additional filtration of the complex $C_*(H)$ by the action functional.

Definition 2.2. For any $a \in \mathbb{R}$ we define the the $\mathbb{Z}$-module

$$C_k^a(H) = \left\{ \sum_{\mu(x)=k} a_x\, x \in C_k(H) \,|a_x = 0 \quad \text{for } \; \mathcal{A}_H(x) > a \right\}.$$

In view of Lemma 2.1, $\big(C_*^a(H),\partial\big)$ is obviously a sub-complex of $(C_*(H),\partial)$. Denoting $C_*^{(a,b]}(H) = C_*^b(H)/C_*^a(H)$ for $a<b$ we have the short exact sequence of chain complexes

$$0 \to C_*^a \xrightarrow{\ i\ } C_*^b \xrightarrow{\ j\ } C_*^{(a,b]} \to 0, \quad a<b\le\infty.$$

We use the convention that $C_*^{\infty}(H) = C_*(H)$ and $C_*^{(-\infty,a]} = C_*^a$. Consequently we obtain for the homology groups

$$HF_*^{(a,b]} = \big(C_*^{(a,b]}(H),\partial(J,H)\big)$$

the long exact sequence associated to $-\infty\le a<b<c\le\infty$,

(2)
$$\ldots \to HF_{k+1}^{(b,c]} \xrightarrow{\partial_*} HF_k^{(a,b]} \xrightarrow{i_*} HF_k^{(a,c]} \xrightarrow{j_*} HF_k^{(b,c]} \xrightarrow{\partial_*} HF_{k-1}^{(a,b]} \to \ldots .$$

Lemma 2.3. *It holds that $HF_*^{(-\infty,a]}(H) = \{0\}$ for $a\in\mathbb{R}$ small enough and $HF_*^{(b,\infty]}(H) = \{0\}$ for $b\in\mathbb{R}$ large enough.*

Proof. This follows immediately from the regularity of H which implies that $\mathcal{P}_1(H)$ is finite. So, choose $a<\min\{\,\mathcal{A}_H(x)\,|\,x\in\mathcal{P}_1(H)\,\}$ and $b>\max\{\,\mathcal{A}_H(x)\,|\,x\in\mathcal{P}_1(H)\,\}$. □

Let us now analyze within the context of symplectic homology the concrete realization of the canonical isomorphism between Floer homology and singular cohomology which was introduced in [**16**]. We first define the isomorphism

$$\Phi\colon H^k(M) \xrightarrow{\cong} HF_{n-k}^{(-\infty,\infty]}(H).$$

We represent the standard cohomology of M in terms of Morse cohomology, see [**19**]. That is, let f be an auxiliary Morse function on M and g a generic Riemannian metric. We choose a smooth cut-off function

$$\beta^-(s) = \begin{cases} 1, & s\le -1,\\ 0, & s\ge -\frac{1}{2},\end{cases}$$

and define the solution spaces of mixed type associated to $x\in\operatorname{Crit}_k f = \{x\in\operatorname{Crit} f|\mu_{\text{Morse}}(x)=k\}$ and $y\in\mathcal{P}_1(H)$,

$$\begin{aligned}\mathcal{M}^-_{y;x}(H,J;f) = \Big\{(u,\gamma)\,\Big|\, & u\colon\mathbb{R}\times S^1\to M,\ \gamma\colon[0,\infty)\to M,\\ & \partial_s u + J\big(\partial_t u - \beta^-(s)X_H(u)\big) = 0,\\ & \int_{-\infty}^{\infty}|\partial_s u|^2 ds dt<\infty,\ \dot\gamma + \nabla_g f\circ\gamma = 0,\\ & u(-\infty)=y,\quad u(+\infty)=\gamma(0),\quad \gamma(+\infty)=x\Big\}.\end{aligned}$$

Note that the condition of finite flow energy of u together with the cut-off function β^- in the perturbed Cauchy-Riemann equation implies by removable singularities that u has a continuous extension to $u(+\infty)$. For regularity, we have to allow even explicit dependence of J on the variable $s \in \mathbb{R}$ in order that for a generic J, f and g, $\mathcal{M}^-_{y;x}$ is again component-wise a manifold of dimension

$$\dim \mathcal{M}^-_{y;x}(H, J; f, g) = n - \mu_{\text{Morse}}(x) - \mu(y)\,.$$

(See also [**20**].) Moreover, in view of (A) and standard bubbling-off analysis these solution spaces are compact in dimension 0, i.e., finite and the following $\mathbb{Z}$-module homomorphism is well-defined,

$$\begin{aligned} &\Phi_\bullet\colon C^k(f) \to C_{n-k}(H),\\ &\Phi_\bullet(x) = \sum_{\mu(y)+\mu_{\text{Morse}}(x)=n} \#_{\mathsf{alg}}\mathcal{M}^-_{y;x}(H, J; f, g)\, y. \end{aligned} \tag{3}$$

A theorem analogous to Floer's central theorem states that $\Phi_\bullet$ commutes with the boundary operator $\partial(H, J)$ on $C_*(H)$ and the coboundary operator of the Morse cochain complex associated to f and g. Hence we obtain the induced homomorphism

$$\Phi\colon H^k(M) \to HF_{n-k}(H, J).$$

Details are carried out in [**18**]. There it is shown that Φ is indeed an isomorphism. We have an analogous representation of Φ^{-1}. Define in the reversed order for $x \in \text{Crit}\, f$ and $y \in \mathcal{P}_1(H)$

$$\begin{aligned} \mathcal{M}^+_{x;y}(H, J; f) = \Big\{ (\gamma, u) \,\Big|\, & u\colon \mathbb{R} \times S^1 \to M,\ \gamma\colon (-\infty, 0] \to M,\\ & \partial_s u + J\big(\partial_t u - \beta^-(-s) X_H(u)\big) = 0,\\ & \int_{-\infty}^{\infty} |\partial_s u|^2 ds dt < \infty,\ \dot\gamma + \nabla_g f \circ \gamma = 0,\\ & \gamma(-\infty) = x, \quad \gamma(0) = u(-\infty), \quad u(+\infty) = y. \end{aligned}$$

This implies for generic J, f and g manifolds of dimension $\mu(y)+\mu_{\text{Morse}}(x)-n$, so that we obtain

$$\begin{aligned} &\Psi_\bullet\colon C_k(H) \to C^{n-k}(f),\\ &\Psi_\bullet(y) = \sum_{\mu(y)+\mu_{\text{Morse}}(x)=n} \#_{\mathsf{alg}}\mathcal{M}^+_{x;y}(H, J; f, g)\, x. \end{aligned} \tag{4}$$

In [**16**] it is shown that the induced homomorphism

$$\Psi\colon HF_k(H) \to H^{n-k}(M)$$

equals Φ^{-1}.

2.2. Critical levels of the action functional. Following an idea of Viterbo, which in [**23**] was used to define symplectic capacities via generating functions, we will associate critical levels $c(\alpha; H)$ for the action functional $\mathcal{A}_H$ associated to given cohomology classes α in M. The same idea was also employed in [**14**].

Definition 2.4. Given the Hamiltonian $H\colon S^1 \times M \to \mathbb{R}$ we define

$$\begin{aligned} E_+(H) &= -\int_0^1 \inf_M H(t,\cdot)dt, \\ E_-(H) &= -\int_0^1 \sup_M H(t,\cdot)dt, \\ \text{and} \quad \|H\| &= E_+(H) - E_-(H). \end{aligned}$$

Note that $\|H\|$ is a semi-norm with $\|H-K\| \le \|H-K\|_{C^0(M\times S^1)}$. Consider now $x \in \mathrm{Crit}_k(f)$ and $\Phi_\bullet(x) = \sum_{\mu(y)=n-k} a_y\, y$. It is a straightforward computation (compare the energy estimate, Lemma 4.1 in [**21**] and the proof of Lemma 2.12) to show that

$$\mathcal{A}_H(y) \le E_+(H) \quad \text{for } a_y \ne 0. \tag{5}$$

Analogously, we have

$$\mathcal{A}_H(y) \ge E_-(H) \quad \text{whenever } \Psi_\bullet(y) \ne 0. \tag{6}$$

This shows:

Lemma 2.5. *The isomorphisms Φ and Ψ factorize as*

$$\Phi\colon H^k(M) \xrightarrow{\tilde\Phi} HF_{n-k}^{(-\infty,E_+(H)]}(H) \xrightarrow{i_*} HF_{n-k}(H),$$

$$\Psi\colon HF_k(H) \xrightarrow{j_*} HF_k^{(E_-(H)-\epsilon,\infty]} \xrightarrow{\tilde\Psi} H^{n-k}(M), \textit{ for all } \epsilon > 0.$$

Consider the diagram for given $a \in \mathbb{R}$

$$\begin{array}{ccccc} & & H^k(M) & & \\ & & \Phi\big\downarrow & & \\ HF_{n-k}^{(-\infty,a]}(H) & \xrightarrow{i_*^a} & HF_{n-k}^{(-\infty,\infty]}(H) & \xrightarrow{j_*^a} & HF_{n-k}^{(a,\infty]}(H) \\ & & \Psi\big\downarrow & & \\ & & H^k(M). & & \end{array}$$

In other terms, Lemma 2.5 says that

$$\begin{cases} j_*^a \circ \Phi = 0, \text{ i.e., } \operatorname{im}\Phi \subset \operatorname{im} i_*^a, & \text{if } a \ge E_+(H), \\ \Psi \circ i_*^a = 0, \text{ i.e., } \operatorname{im}\Phi \cap \operatorname{im} i_*^a = \{0\} & \text{if } a < E_-(H). \end{cases} \tag{7}$$

This lends itself to the following definition:

Definition 2.6. Given any nonzero cohomology class $\alpha \in H^*(M)$ we define

$$c(\alpha; H) = \inf\{\, a \in \mathbb{R} \mid j^a_*(\Phi(\alpha)) = 0\}.$$

In view of Lemma 2.5 and (7) the $c(\alpha; H)$ are finite real numbers which are obviously critical values of $\mathcal{A}_H$. They satisfy

$$E_-(H) \le c(\alpha; H) \le E_+(H) \quad \text{for all } 0 \ne \alpha \in H^*(M). \tag{8}$$

In the following we discuss the behaviour of $c(\alpha; H)$ with respect to variations of H leading to first continuity properties, and with respect to variations of α, in terms of the cohomology ring structure on $H^*(M)$.

2.3. The cap-action of $H^*(M)$. Let us recall from [**2**] the definition of the cap-action

$$\cap\colon H^l(M) \times HF_k(H) \to HF_{k-l}(H). \tag{9}$$

Given a generic Morse function f and a Riemannian metric g as above we denote for a critical point $x \in \operatorname{Crit} f$

$$W^s_x(f) = \{\, \gamma\colon [0,\infty) \to M \mid \dot\gamma + \nabla_g f \circ \gamma = 0,\ \gamma(+\infty) = x \,\}.$$

Recall the definition of the Floer trajectory space $\mathcal{M}_{z,y}(J,H)$ from above. Given $y, z \in \mathcal{P}_1(H)$ we set

$$\mathcal{M}_{z;x;y}(H, J; f, g) = \{\, (u,\gamma) \in \mathcal{M}_{z,y} \times W^s_x \mid u(0,0) = \gamma(0) \,\}.$$

As before, for generic J, f and g, this is a manifold of dimension

$$\dim \mathcal{M}_{z;x;y} = \mu(y) - \mu(z) - \mu_{\text{Morse}}(x),$$

compact in dimension 0. Thus, we can define

$$\cap\colon C^l(f) \times C_k(H) \to C_{k-l}(H),$$
$$x \cap y = \sum_{\mu(z) = \mu(y) - \mu_{\text{Morse}}} \#_{\mathsf{alg}} \mathcal{M}_{z;x;y}(H; f)\, z.$$

Again, $\cap$ commutes with the boundary operators in the standard way so that we obtain the cap-action (9) of the standard cohomology ring $H^*(M)$ on Floer homology. The first of the following crucial relations is essentially due to Floer and has been studied in full details in [**13**].

Proposition 2.7.

1) *The operation $\cap$ is a ring operation, i.e.,*

$$(u \cap \alpha) \cap \beta = u \cap (\alpha \cup \beta),$$

for all $\alpha, \beta \in H^*(M)$, $u \in HF_*(H)$.

2) *It is compatible with the isomorphism $\Phi\colon H^*(M) \to HF_*(H)$,*

$$\Phi(\alpha) \cap \beta = \Phi(\alpha \cup \beta).$$

The proof of the second property is carried out in [**18**].

It should be no surprise that this cap-action is also compatible with the refined structure as symplectic homology. We have:

Lemma 2.8. *For generic J, f and g and $\mu_{Morse}(x) > 0$, a non-empty solution space $\mathcal{M}_{z;x;y}(H, J; f, g)$ implies*

$$\mathcal{A}_H(z) \leq \mathcal{A}_H(y) - \epsilon(H)$$

with $0 < \epsilon(H) \stackrel{\mathsf{def}}{=} \min\{0 < \mathcal{A}_H(y) - \mathcal{A}_H(z) | y, z \in \mathcal{P}_1(H)\}$.

Proof. Suppose the assertion is not true. By definition of $\epsilon(H)$ we have $\mathcal{A}_H(y) = \mathcal{A}_H(z)$. Hence, due to Lemma 2.1, $u \in \mathcal{M}_{z,y}$ must be constant, i.e., $u(0,0) = y(0) = z(0) \in \operatorname{Fix} \phi_H^1$. However, $u(0,0) \in W_x^s(f,g)$, but for generic f and g, W_x^s is a manifold of codimension at least 1 and does not intersect the finite set $\operatorname{Fix} \phi_H^1$. □

We obtain:

Proposition 2.9. *Given $\alpha, \beta \in H^*(M)$ with $\alpha \cup \beta \neq 0$ and $\beta \neq 1$ we have the estimate*

$$c(\alpha \cup \beta; H) \leq c(\alpha; H) - \epsilon(H),$$

in particular, $c(\alpha \cup \beta; H) < c(\alpha; H)$.

Proof. In view of the definition of $c(\cdot; H)$ suppose that $j_*^a(\Phi(\alpha)) = 0$. Due to Lemma 2.8 it follows that $\Phi(\alpha) \cap \beta \in \operatorname{im} i_*^{a-\epsilon(H)}$. Hence by Proposition 2.7 we have $j_*^{a-\epsilon(H)}\big(\Phi(\alpha \cup \beta)\big)$. □

We have the following a priori estimate for the critical levels $c(\alpha; H)$, compare (8).

Proposition 2.10. *Let $1 \in H^0(M)$ and $[M] \in H^{2n}(M)$ be the two canonical classes in the ring $H^*(M)$. We have*

$$E_-(H) \leq c([M]; H) \; < \; c(1; H) \leq E_+(H),$$

in particular,

$$c([M]; H) \; < \; c(\alpha; H) \; < \; c(1; H) \quad \textit{for any } \alpha \in H^k(M),\, 0 < k < 2n.$$

The upper estimate follows directly from Proposition 2.9, the lower estimate is due to the Poincaré duality which will be discussed below.

Following Viterbo in [**23**] we can make the following:

Definition 2.11. Given any regular Hamiltonian H we define the positive number $\gamma(H)$,

$$0 < \epsilon(H) \leq \gamma(H) \stackrel{\mathsf{def}}{=} c(1; H) - c([M]; H) \leq \|H\|.$$

2.4. First continuity properties. Let H and K be any two Hamiltonians and consider the associated canonical isomorphism between the associated Floer homologies (cf. [**17**], [**20**])

$$\Phi_{KH}\colon HF_*(H) \xrightarrow{\cong} HF_*(K).$$

Namely, we choose a monotone increasing smooth cut-off function $\beta(s)$, say $\beta(s) = 0$ for $s \leq -1$ and $\beta(s) = 1$ for $s \geq 1$, and we define the homotopy of Hamiltonians

$$G_s(t,x) = K(t,x) + \beta(s)\big(H(t,x) - K(t,x)\big).$$

Given $y \in \mathcal{P}_1(K)$ and $x \in \mathcal{P}_1(H)$ we recall the associated moduli space of homotopy trajectories

$$\mathcal{M}_{y,x}(G_s, J) = \{\, u \,|\, \partial_s u + J(\partial_t u - X_{G_s}(u)) = 0,\ u(-\infty) = y,\ u(+\infty) = x \,\}.$$

We have the following energy estimate.

Lemma 2.12. *If* $u \in \mathcal{M}_{y,x}(G_s)$ *then*

$$\mathcal{A}_K(y) \leq \mathcal{A}_H(x) + E^+(H-K).$$

Recall the definition of $E^+(f) = -E^-(-f)$ in 2.4.

Proof. A simple computation shows

$$\begin{aligned} 0 \leq \int_{-\infty}^{\infty}\int_0^1 |\partial_s u|^2 ds dt &= \iint_{\mathbb{R}\times S^1} u^*\omega - \int_{-\infty}^{\infty}\int_0^1 \omega(u_s, X_{G_s}) ds dt \\ &= \mathcal{A}_H(x) - \mathcal{A}_K(y) + \int_{-\infty}^{\infty} \beta'(s) \int_0^1 \big(H(t,u(s,t)) - K(t,u(s,t))\big) dt ds \\ &\leq \mathcal{A}_H(x) - \mathcal{A}_K(y) + \int_0^1 \sup_p \big(H(t,p) - K(t,p)\big) dt. \end{aligned}$$

□

Thus it follows that the canonical homomorphism is already defined as

$$\Phi_{KH}\colon HF_*^{(-\infty,a]}(H) \to HF_*^{(-\infty,a+E^+(H-K)]}(K).$$

What is more, it is compatible with the long exact sequence of symplectic homology (2), i.e., we have the commutative diagram with $e^+ = E^+(H-K)$,

(10)

$$\begin{array}{ccccccc} HF^{(-\infty,a]}(H) & \xrightarrow{i_*^a} & HF^{(-\infty,\infty)}(H) & \xrightarrow{j_*^a} & HF^{(a,\infty]}(H) & \xrightarrow{\partial_*} & \\ \downarrow \Phi_{KH} & & \cong\downarrow \Phi_{KH} & & \downarrow \Phi_{KH} & & \\ HF^{(-\infty,a+e^+]}(K) & \xrightarrow{i_*^{a+e^+}} & HF^{(-\infty,\infty)}(K) & \xrightarrow{j_*^{a+e^+}} & HF^{(a+e^+,\infty]}(K) & \xrightarrow{\partial_*} & . \end{array}$$

Moreover, recall from the explicit isomorphism between Morse homology and Floer homology that for $\Phi_H\colon H^*(M) \xrightarrow{\cong} HF_{n-*}(H)$ we have

$$\Phi_K = \Phi_{KH} \circ \Phi_H. \tag{11}$$

Thus, if $j_*^a\big(\Phi_H(\alpha)\big) = 0$ for some $\alpha \in H^*(M)$ and $a \in \mathbb{R}$ it follows by (10) and (11) that also $j_*^{a+e^+}\big(\Phi_K(\alpha)\big) = 0$. We obtain:

Lemma 2.13. *For any two Hamiltonians H and K and any nonzero cohomology class $\alpha \in H^*(M)$ we have the estimate*

$$c(\alpha; K) \leq c(\alpha; H) + E^+(H - K).$$

This immediately implies the property that the map $H \mapsto c(\alpha; H) \in \sigma(\phi_H^1) \subset \mathbb{R}$ is continuous with respect to the semi-norm

$$\|H\| = \int_0^1 \operatorname{osc}_x H(t,x)dt \;=\; E_+(H) - E_-(H),$$

namely

$$\big|c(\alpha; H) - c(\alpha; K)\big| \leq \|H - K\| \tag{12}$$

for each $0 \neq \alpha \in H^*(M)$. By this first continuity result we see that $c(\alpha, H)$ is in fact well-defined for all Hamiltonians. Denoting by $\mathcal{H}$ the set of all Hamiltonians and by $\mathcal{H}_{reg}$ the dense subset of regular ones we have:

Proposition 2.14. *We have a well-defined function*

$$c\colon H^*(M) \setminus \{0\} \times \mathcal{H} \to \mathbb{R}, \quad (\alpha, H) \mapsto c(\alpha, H),$$

which is continuous in H with respect to the semi-norm $\|H\|$.

Proof. This follows immediately from (12) recalling from [**17**] that $\mathcal{H}_{\mathrm{reg}} \subset \mathcal{H}$ is C^0-dense □

From now on we do not assume regularity of H without further notice.

3. The action spectrum.

The fact known for $(\mathbb{R}^{2n}, \omega_o)$ that the action spectrum is well-defined already for a Hamiltonian automorphism with compact support regardless of the chosen Hamiltonian will now be transfered to the case of a symplectically aspherical closed manifold. However, a priori we have to lift the analysis to the universal covering of $\mathrm{Ham}(M, \omega)$. Choosing a suitable normalization for the generating Hamiltonian functions H we observe that the above functions $c(\alpha, \cdot)$ are uniquely defined on this covering group and we will finally show that they are continuous sections in the so-called action spectrum bundle over $\widetilde{\mathrm{Ham}}(M,\omega)$ with respect to Hofer's bi-invariant metric. Recall the definition of this metric on $\mathrm{Ham}(M,\omega)$,

$$d_H(\phi,\psi) = d_H(\phi\psi^{-1},\mathrm{id}), \quad d_H(\phi,\mathrm{id}) = \inf\{\, \|H\| \mid \phi = \phi_H^1 \,\},$$
$$\|H\| = \int_0^1 \operatorname*{osc}_{x\in M} H(t,x)\,dt.$$

From now on we assume that the symplectic manifold M is connected and that $\big(\mathrm{Ham}(M,\omega), d_H\big)$ is the topological group with Hofer's bi-invariant metric.

3.1. The action spectrum bundle. Let H and K be two Hamiltonians which generate the same symplectomorphism,

$$\phi_H^1 = \phi_K^1 \overset{\mathsf{def}}{=} \phi.$$

Without loss of generality we can assume that $H(1,\cdot) = K(1,\cdot) = 0$. Namely, replace H by $H_\alpha(t,\cdot) = \alpha'(t)H(\alpha(t),\cdot)$ where α is any monotone map $\alpha\colon [0,1] \to [0,1]$ satisfying $\alpha(0) = 0$, $\alpha(1) = 1$ and α' has compact support in $(0,1)$. Consider the loop in $\mathrm{Ham}(M,\omega)$ based at the identity

$$g_t = \begin{cases} \phi_H^t, & 0 \le t \le 1, \\ \phi_K^{2-t}, & 1 \le t \le 2\,. \end{cases} \tag{13}$$

Clearly, g_t is the flow associated to the Hamiltonian

$$\varphi^t = \phi_G^t, \quad G(t,\cdot) = \begin{cases} H(t,\cdot), & 0 \le t \le 1, \\ -K(2-t,\cdot), & 1 \le t \le 2\,. \end{cases}$$

Let $\chi\colon M \to \Omega(M)$ be the induced continuous map into the free loop space $\chi(p) = (g_t(p))_{t\in[0,2]}$. Thus, $\chi(p_o)$ is contractible exactly if all $\chi(p)$ are contractible. From Floer theory we know that for any k-periodic Hamiltonian function $G(t+k,p) = G(t,p)$ there exists at least one contractible k-periodic solution of the associated Hamiltonian equation. Therefore the map χ has its image in the component of contractible loops and we obtain:

Proposition 3.1. *Given any two Hamiltonians H and K generating the same symplectomorphism $\phi_H^1 = \phi_K^1$, there is a canonical identification of the contractible 1-periodic solutions*

$$\mathcal{P}_1(H) \cong \mathcal{P}_1(K) \neq \emptyset.$$

Hence, we can define for $\phi \in \mathrm{Ham}(M,\omega)$

$$\mathrm{Fix}^o(\phi) = \big\{\, x(0) \mid x \in \mathcal{P}_1(H),\ \phi = \phi_H^1 \,\big\}. \tag{14}$$

In the following we use the same notation for $x \in \mathrm{Fix}^o(\phi_H^1)$ and $x \in \mathcal{P}_1(H)$.

In order to associate a well-defined action already to the fixed point $x \in \mathrm{Fix}^o(\phi)$ we consider the following function $I_g\colon M \to \mathbb{R}$. Let $g\colon [0,1] \to \mathrm{Ham}(M,\omega)$ be a closed loop based at the identity

$$g_0 = g_1, \quad g_t = \phi_G^t, \quad G(t+1,p) = G(t,p).$$

Given any $p \in M$ denote by $\widetilde{g_t(p)}$ an extension of the contractible loop to the unit disk. Since $\omega_{|\pi_2} = 0$, the function

$$I(g,p) = \int_{D^2} \widetilde{g_t(p)}^* \omega - \int_0^1 G(t, g_t(p))dt$$

is well-defined.

Definition 3.2. We call a Hamiltonian function $H\colon \mathbb{R} \times M \to \mathbb{R}$ **normalized** if

$$\int_M H(t,\cdot)\omega^n = 0 \quad \text{for all } t \in \mathbb{R}.$$

Obviously, for any Hamiltonian isotopy ϕ_H^t, H can be normalized,

$$H_{\mathrm{norm}}(t,x) = H(t,x) - \frac{1}{\int_M \omega^n} \int_M H(t,\cdot)\omega^n, \tag{15}$$

and the action changes by a constant,

$$\mathcal{A}_H = \mathcal{A}_{H_{\mathrm{norm}}} - \int_0^1 \int_M H(t,x)\omega^n dt.$$

Lemma 3.3. *The function $I(g,p)$ does not depend on $p \in M$ and $I(g)$ is invariant under homotopies of the loop g if the Hamiltonian G is homotoped through normalized functions.*

Proof. Let $s \mapsto p(s)$, $s \in (-\epsilon,\epsilon)$ be any differentiable arc with $p'(0) = \xi$, then

$$\begin{aligned} \frac{d}{ds}\big(I(g,p(s))\big)\big|_{s=0} &= \int_0^1 \omega\big(Dg_t(p)\xi, X_G(t,g_t(p))\big)\,dt \\ &\quad - \int_0^1 dG(t,g_t(p)) \circ Dg_t(p)\xi\,dt = 0, \end{aligned}$$

hence $I(g,p) = I(g)$ for all $p \in M$. Let $g_{s,t} = \phi_{G_s}^t$, $G(s,t+1,p) = G(s,t,p)$, $s \in (-\epsilon,\epsilon)$, be a 1-parameter family of Hamiltonian loops based at id_M. Then a straightforward computation shows

$$\frac{d}{ds}\big(I(g(s),p)\big) = -\int_0^1 \partial_s G(s,t,g_{s,t}(p))\,dt \quad \text{f.a.}\, p \in M.$$

Integration over M with respect to ω^n shows that $\frac{d}{ds} I(g(s)) = 0$. □

Clearly, we obtain group homomorphisms

$$I\colon \pi_1(\mathrm{Ham}(M,\omega)) \to \mathbb{R} \tag{16}$$

describing the obstruction to a well-defined action spectrum for each $\phi \in \mathrm{Ham}(M,\omega)$.

Remark 3.4. If every element of $\pi_1(\mathrm{Ham}(M,\omega))$ is of finite order, then $I \equiv 0$. In general, if $\omega_{|\pi_2} \neq 0$, we only have $I\colon \pi_1(\mathrm{Ham}) \to \mathbb{R}/\omega(\pi_2(M))$ and this argument fails.

Given any point $p_o \in M$ consider the evaluation map $\mathrm{ev}_{p_o}\colon \mathrm{Ham}(M,\omega) \to M$, $\phi \mapsto \phi(p_o)$. Like in the proof of Proposition 3.1 we know that the induced homomorphism $\pi_1(\mathrm{Ham},\mathrm{id}) \to \pi_1(M,p_o)$ vanishes, $\mathrm{ev}_* = 0$. Moreover, ev_{p_o} is a locally trivial fibration with typical fibre $\mathrm{Ham}_{p_o} = \{\, \phi \in \mathrm{Ham} \mid \phi(p_o) = p_o \,\}$. We obtain the exact homotopy sequence

$$\ldots \to \pi_2(M,p_o) \to \pi_1(\mathrm{Ham}_{p_o},\mathrm{id}) \to \pi_1(\mathrm{Ham},\mathrm{id}) \to 0.$$

We conclude:

Remark 3.5. For each $\{g\} \in \pi_1(\mathrm{Ham},\mathrm{id})$ and $p_o \in M$ there exists a normalized generating Hamiltonian $G(t,p)$ such that $\nabla G(t,p_o) = 0$ for all $t \in S^1$ and thus

$$I(\{g\}) = -\int_0^1 G(t,p_o)dt.$$

Example 3.6. Consider (S^2,ω_o) with $\omega_o(S^2) = 1$ and $\{g\} \in \mathrm{Ham}(S^2,\omega_o) = \mathbb{Z}_2$ generated by the normalized autonomous Hamiltonian G generating the rotation around the axis through the poles. Its values at the poles are $\pm 1/2$. Hence for the homomorphism $I\colon \mathbb{Z}_2 \to \mathbb{R}/\mathbb{Z}$ we have $I(\{g\}) = I(1) = 1/2$ (mod 1).

Let the Hamiltonians H and K generate the same automorphism $\phi_H^1 = \phi_K^1$. We call them *equivalent*, $H \sim K$ if there exists a homotopy $(G_s)_{s\in[0,1]}$ of Hamiltonians such that each G_s generates the same automorphism, $\phi_{G_s}^1 = \phi_H^1 = \phi_K^1$ for all $s \in [0,1]$. It is easy to see that the group of equivalence classes is the universal covering of $\mathrm{Ham}(M,\omega)$,

$$\widetilde{\mathrm{Ham}}(M,\omega) = \{\, [H] \,|\, H \sim K,\ H,K \text{ normalized} \,\}.$$

From the previous analysis it follows that we obtain a well-defined action spectrum bundle over the group $\widetilde{\mathrm{Ham}}$,

$$\Sigma = \bigcup_{\bar\phi \in \widetilde{\mathrm{Ham}}} \{\bar\phi\} \times \Sigma_{\bar\phi}, \quad \Sigma_{\bar\phi} = \{\mathcal{A}_H(x) | x \in \mathrm{Fix}^o(\phi_H^1),\ \bar\phi = [H]\}.$$

The action $\pi_1(\mathrm{Ham}) \times \widetilde{\mathrm{Ham}} \to \widetilde{\mathrm{Ham}}$, $(\gamma,\bar\phi) \mapsto \gamma\bar\phi$ gives

$$\mathcal{A}_{\gamma\bar\phi}(x) = \mathcal{A}_{\bar\phi}(x) + I(\gamma), \quad \text{for all } x \in \mathrm{Fix}^o(\phi),\ \phi = \pi(\bar\phi).$$

Analogously to the situation of compactly supported automorphisms of $(\mathbb{R}^{2n}, \omega_o)$ as studied in [**8**] we have:

Proposition 3.7. *The action spectrum $\mathcal{A}_{\bar\phi} \subset \mathbb{R}$ is compact and nowhere dense.*

Proof:

Lemma 3.8. *The regular values of the Hamiltonian action functional $\mathcal{A}_H$ form a residual subset of $\mathbb{R}$.*

Proof. We will construct a smooth function on a finite-dimensional manifold whose set of critical values contains all critical values of $\mathcal{A}_H$. Thus by Sard's theorem the claim follows.

Let $S_2 = \mathbb{R}/2\mathbb{Z}$ and define the function $F\colon H^{1,2}(S_2, M) \to \mathbb{R}$,

$$F(x) = \int_{D^2} \tilde{x}^*\omega - \int_0^1 H(t, x(t)),$$

where $D^2 = \{z \,|\, |z| \leq 1\}$, $\tilde{x}\colon D^2 \to M$, $\tilde{x}(e^{\pi i t}) = x(t)$, $t \in [0,2]$. Its differential is given by

$$dF(x)(\xi) = \int_0^2 \big\langle -J\big(\dot{x} - \chi_{[0,1]} X_H(t, x(t))\big), \xi(t)\big\rangle dt$$

for $\xi \in H^{1,2}(x^*TM)$. We obtain the embedding of the critical points of $\mathcal{A}_H$ into the critical set of F,

$$i\colon \operatorname{Crit} \mathcal{A}_H \hookrightarrow \operatorname{Crit} F, \quad i(x)(t) = \begin{cases} x(t), & 0 \leq t \leq 1, \\ x(1), & 1 \leq t \leq 2\,. \end{cases}$$

with $\mathcal{A}_H(x) = F(i(x))$ for all $x \in \operatorname{Crit} \mathcal{A}_H$. Clearly, $i(\operatorname{Crit} \mathcal{A}_H) \subset H^{1,2}(S_2, M)$ and F is a smooth function.

Denote now by $U \subset TM$ the injectivity neighbourhood of the exponential map such that

$$\exp\colon U \xrightarrow{\approx} V(\triangle) \subset M \times M, \quad \exp(p, v) = (p, \exp_p(v)).$$

We define the mapping $c\colon X \to H^{1,2}(S_2, M)$ on the open subset $X = \{\, x \in M \,|\, (x, \phi_H^1(x)) \in U \,\}$ by

$$c(p)(t) = \begin{cases} \phi_H^t(p), & 0 \leq t \leq 1, \\ \exp_p\big((2-t)\exp^{-1}(p, \phi_H^1(p))\big), & 1 \leq t \leq 2. \end{cases}$$

The mapping c is smooth and we have $c(x(0)) = i(x)$ for all $x \in \operatorname{Crit} \mathcal{A}_H$. Defining

$$f\colon X \to \mathbb{R}, \quad f(p) = F(c(p))$$

we conclude that the critical values of $\mathcal{A}_H$ form a subset of the critical values of the smooth function f. $\square$

The proof of Proposition 3.7 follows immediately from this Lemma and the compactness of the set of critical points of $\mathcal{A}_H$ since it can be identified with the fixed point set $\mathrm{Fix}^o(\phi) \subset M$. □

We now conclude that $c(\alpha; H)$ only depends on the equivalence class $[H] \in \widetilde{\mathrm{Ham}}(M, \omega)$. Consider H and K such that there exists a homotopy $H \overset{G_s}{\simeq} K$, that is $(G_s)_{s\in[0,1]}$, with $\phi_H^1 = \phi_{G_s}^1 = \phi_K^1$ for all $s \in [0,1]$. After subdividing G_s in homotopies (G_s), $s \in [s_i, s_{i+1}]$, $0 = s_0 < s_1 < \ldots < s_N = 1$, we have

$$\left|c(\alpha; G_{s_{i+1}}) - c(\alpha; G_{s_i})\right| \leq \|G_{s_{i+1}} - G_{s_i}\|.$$

Choosing the subdivision small enough we obtain in view of (12) and Proposition 3.7 that $c(\alpha; H) = c(\alpha; K)$. This idea of an *adiabatic* homotopy implies that $c(\alpha, \cdot)$ can be viewed as sections of the action spectrum bundle over $\widetilde{\mathrm{Ham}}$.

Proposition 3.9. *For every nonzero cohomology class $\alpha \in H^*(M)$ we obtain a section $c(\alpha)$ of the action spectrum bundle $\Sigma \to \widetilde{\mathrm{Ham}}(M, \omega)$ which is continuous with respect to the Hofer-metric $d_{\tilde{H}}(\bar{\phi}, \mathrm{id}) = \inf \left\{ \|H\| \,|\, \bar{\phi} = [H] \right\}$ on the covering $\widetilde{\mathrm{Ham}}(M, \omega)$.*

Note that the continuity does not follow directly from the C^0-continuity $H \mapsto c(\alpha; H)$ as shown in Lemma 2.13. We will prove continuity in Corollary 4.10 below using an additional structure with respect to the pair-of-pants product.

Since the sections $c(\alpha)$ are only defined over the universal covering $\widetilde{\mathrm{Ham}}$ it is important to determine the action of $\pi_1(\mathrm{Ham})$ on them. Let us introduce a more suitable description of Floer homology which has also been used by P. Seidel in [**22**] in order to study $\pi_1(\mathrm{Ham})$.

3.2. The intrinsic Floer homology. Let $\phi \in \mathrm{Ham}(M, \omega)$ be a given automorphism and consider the associated mapping torus as a symplectic fibre bundle over S^1,

$$M_\phi = \{(t, x) \in \mathbb{R} \times M\}/((t, \phi(x)) \sim (t+1, x)) \to S^1. \tag{17}$$

Obviously $\omega \in \Omega^2(M_\phi)$ is a closed 2-form and (M, ω) is the typical fibre. Consider now the following path space

$$\mathcal{L}_\phi = \left\{ x \colon \mathbb{R} \to M \,|\, \phi(x(t+1)) = x(t) \right\},$$

i.e., the space of sections in $M_\phi \to S^1$. We can view the nonempty point set of fixed points $\mathrm{Fix}^o(\phi)$ defined in (14), corresponding to contractible periodic solutions, as a subset of $\mathcal{L}_\phi$. In fact, it is contained entirely in one component of $\mathcal{L}_\phi$, and we define

$$\mathcal{L}_\phi^o = \left\{ x \in \mathcal{L} \,|\, x \simeq x_o \right\}, \quad \text{for any } x_o \in \mathrm{Fix}^o(\phi)\,. \tag{18}$$

Recall that the symplectic action was uniquely associated to an $\bar{\phi} \in \widetilde{\mathrm{Ham}}$ with $\pi(\bar{\phi}) = \phi$ and defined on the set $\mathrm{Fix}^o(\phi)$. For each $\bar{\phi}$ we have a well-defined extension to the component $\mathcal{L}^o_\phi$ by:

Proposition 3.10. *For each $\bar{\phi} \in \widetilde{\mathrm{Ham}}$ there exists a continuous extension of the Hamiltonian action $\mathcal{A}_{\bar{\phi}}$ to a continuous function $l_{\bar{\phi}} \colon \mathcal{L}^o_\phi \to \mathbb{R}$ such that for each differentiable*[1] *path $\gamma \colon [0,1] \to \mathcal{L}^o_\phi$ it holds*

$$l_{\bar{\phi}}\big(\gamma(1)\big) - l_{\bar{\phi}}\big(\gamma(0)\big) = \int_0^1 \int_0^1 \gamma^*\omega.$$

Proof. Essentially, it satisfies to realize that given any two paths $\gamma, \bar{\gamma} \colon [0,1] \to \mathcal{L}^o_\phi$ with coinciding ends, $\gamma(0) = \bar{\gamma}(0)$, $\gamma(1) = \bar{\gamma}(1)$, we have

$$\iint \gamma^*\omega = \iint \bar{\gamma}^*\omega.$$

This follows from $d\omega = 0$ on M_ϕ, $\omega_{|\pi_2(M)} = 0$ and the fact that we restricted $\mathcal{L}_\phi$ to the component of sections which contains the fixed points from contractible periodic solutions. It remains to verify

$$\iint \gamma^*\omega = \mathcal{A}_H(\gamma(1)) - \mathcal{A}_H(\gamma(0)) \tag{19}$$

for a differentiable path between $x, y \in \mathrm{Fix}_o(\phi)$. Set $u(s,t) = \phi^t_H(\gamma(s,t))$, then clearly $u(s,t+1) = u(s,t)$ and one computes

$$\begin{aligned} \int_0^1 \int_0^1 u^*\omega &= \iint \omega(D\phi^t_H \gamma_s, D\phi^t_H \gamma_t + X_H(u))\, ds dt \\ &= \iint \gamma^*\omega + \int_0^1 H(t, u(1,t)) dt - \int_0^1 H(t, u(0,t)) dt. \end{aligned}$$

This proves (19). □

Recall the definition of the moduli space of Floer trajectories $\mathcal{M}_{x,y}$ associated to a Hamiltonian $H \colon M \times S^1 \to \mathbb{R}$ and an ω-calibrated almost complex structure J on $TM \to M \times S^1$,

$$\begin{gathered} u \colon \mathbb{R} \times S^1 \to M, \\ u_s + J(t,u)\big[u_t - X_H(t,u)\big] = 0, \\ \lim_{s \to -\infty} u(s,\cdot) = x, \ \lim_{s \to \infty} u(s,\cdot) = y, \end{gathered}$$

where we consider $x, y \in \mathcal{P}_1(H)$.

In fact, Floer homology HF_* is already uniquely associated to the symplectomorphism ϕ, regardless of the generating Hamiltonian (for the grading,

[1] i.e., differentiable as map $[0,1] \times \mathbb{R} \to M$.

see the remark below). Namely, we consider maps $v\colon \mathbb{R}\times\mathbb{R}\to M$, satisfying

$$\begin{gathered} u(s,t)=\phi_H^t(v(s,t)),\ \text{i.e.},\ v(s,t)=\phi_H^1(v(s,t+1)),\\ v_s+\bar J_H(t,v)v_t=0,\\ \text{for } \bar J_H(t,v)=(D\phi_H^t)^{-1}J(t,\phi_H^t(v))D\phi_H^t,\\ \lim_{s\to-\infty} v(s,t)=x(0),\ \lim_{s\to\infty} v(s,t)=y(0),\quad \text{for all } t\in\mathbb{R}. \end{gathered}$$

We identify v with the section $\bar v(s,t)=[s,t,v(s,t)]$ of $E=\mathbb{R}\times M_\phi\to\mathbb{R}\times S^1$. Let us now generalize this Cauchy-Riemann problem for such mappings v by considering merely the time-1-map $\phi=\phi_H^1$ and

$$\begin{aligned} &v\colon \mathbb{R}\times\mathbb{R}\to M,\quad v(s,t)=\phi(v(s,t+1)),\\ &v_s+\bar J(t,v)v_t=0,\quad v(-\infty,\cdot)=x(0),\ v(\infty,\cdot)=y(0),\\ &\text{for } \bar J \text{ satisfying}\quad D\phi^{-1}(v)J(t+1,\phi(v))D\phi(v)=\bar J(t,v). \end{aligned}\tag{20}$$

As in (19), we have for solutions with $\bar J=\bar J_H$

$$\iint_{\mathbb{R}\times S^1}\bar v^*\omega=\int_{-\infty}^{\infty}\int_0^1|v_s|^2_{\bar J_H}dsdt=\mathcal{A}_H(y(0))-\mathcal{A}_H(x(0)).\tag{21}$$

Moreover, as a corollary we obtain the improvement of (10):

Proposition 3.11. *Given any two equivalent Hamiltonians $H\sim K$ generating $\bar\phi\in\widetilde{\mathrm{Ham}}$, the associated canonical homomorphism between the Floer homologies is compatible with the long exact sequence and respects the filtration by the action,*

$$\Phi_{KH}\colon HF_*^a(H)\to HF_*^a(K).$$

Proof. We define now $\Phi_{KH}\colon C_*(H)\to C_*(K)$ by means of (20) where we allow the generic almost complex structure $\bar J$ on the fibre bundle $\mathbb{R}\times M_\phi\to\mathbb{R}\times S^1$ to be explicitly s-dependent such that

$$\bar J(s,\cdot)=\begin{cases}\bar J_H, & s\le -T,\\ \bar J_K, & s\ge T,\end{cases}$$

for some $T>0$. Observe that such a connecting $\bar J$ exists since the space of ω-compatible almost complex structures on the fibre bundle is fibrewise contractible. By virtue of Proposition 3.10 it follows analogously to (7) that connecting trajectories between $y\in\mathcal{P}_1(H)$ and $x\in\mathcal{P}_1(K)$ solving

$$v_s+\bar J(s,t,v)v_t=0\tag{22}$$

satisfy

$$0\le\iint|v_s|^2_{\bar J}dsdt=\iint v^*\omega=\mathcal{A}_K(y(0))-\mathcal{A}_H(x(0)).$$

Note that such trajectories solving (22) cannot be obtained directly from the "adiabatic" homotopy (G_s) described above since the s-derivative would imply an additional 0-order term. The most general setup formulated in terms of connections will be described in the following section. It is straightforward to see that all operators Φ_{KH} defined on the chain level by such sections in $\mathbb{R} \times M_\phi$ induce identical operators between the homology groups $HF_*(H) \xrightarrow{\cong} HF_*(K)$. □

Remark on $HF_*(\phi)$**.** Clearly, for any two $H \sim K$ generating the same $\bar{\phi} \in \widetilde{\mathrm{Ham}}$, the canonical isomorphism Φ_{HK} viewed as an automorphism of the Floer homology $HF_*(\phi)$ associated to the time-1-map $\phi = \pi(\bar{\phi})$ has to be the identity. This is not true in general for $\bar{\phi}, \tilde{\phi} \in \widetilde{\mathrm{Ham}}$ with $\pi(\bar{\phi}) = \pi(\tilde{\phi})$. P. Seidel shows in [**22**] that the group $\pi_1(\mathrm{Ham})$ operates on $HF_*(\phi)$ in terms of the quantum cup-product, i.e., using the canonical isomorphism $HF_*(\phi) \cong QH^{n-*}(M, \omega)$ with the quantum cohomology ring (cf.[**16**]),

$$q\colon \pi_1(\mathrm{Ham}) \to \mathrm{Aut}(HF_*(\phi))$$

is given by a group homomorphism $\pi_1(\mathrm{Ham}) \to QH^\times(M, \omega)$ into the group of invertibles of the quantum cohomology ring of homogeneous degree. Hence, under our assumption (A) of symplectic asphericity, the quantum cohomology and thus q is trivial. It can also directly be seen that the grading on $\mathrm{Fix}^o(\phi)$ by the Conley-Zehnder index is already well-defined if $c_{1|\pi_2} = 0$.

4. The pair-of-pants product and Poincaré duality.

We will now show that also the canonical ring structure on Floer homology, the pair-of-pants product which was constructed in [**20**], is compatible with symplectic homology. This requires a "sharp" energy estimate which will be proven along the same lines as Proposition 3.11. As a consequence we obtain a crucial sub-additivity property for the section $c(\cdot, \cdot)$.

In [**20**] it was shown that every topological surface with oriented cylindrical ends gives rise to a multi-linear operation on Floer homology. Namely, choosing a conformal structure on the surface Σ and associating a Hamiltonian function to each end, one can generalize the Cauchy-Riemann type equation from $\mathcal{M}_{x,y}(J, H)$ to mappings from Σ into M with respective 1-periodic solutions as boundary conditions. The full theory with the verification of the axioms of a topological field theory is carried out in [**20**]. From the gluing axiom describing the concatenation of such multi-linear operators on Floer homology it follows that the entire theory is already uniquely determined by the multiplication associated to the surface with two exits and one entry, by the standard cylinder, by the cylinder with two exits and by the disk with only one end. From [**18**] it follows that this graded algebra is naturally isomorphic to the cohomology ring of M, if $\omega_{|\pi_2(M)} = 0$.

In [**20**], the construction of the pair-of-pants product on Floer homology was as follows. Let Σ be a Riemann surface with three cylindrical ends, one entry and two exits. Associate the Hamiltonians H and K to the exits and choose any Hamiltonian L for the entry. Then for $x \in \mathcal{P}_1(H)$, $y \in \mathcal{P}_1(K)$ and $z \in \mathcal{P}_1(L)$ we can define a moduli space $\mathcal{M}_{z;x,y}(L;H,K)$ of pair-of-pants solutions $u\colon \Sigma \to M$ converging towards x, y and z on the respective end. As before, under generic choices of the almost complex structure J on M, the solution space is a manifold, which is compact in dimension 0 and its dimension formula is

$$\dim \mathcal{M}_{z;x,y} = \mu(x) + \mu(y) - \mu(z) - n.$$

The multiplication on Floer homology is induced by

$$*\colon C_k(H) \times C_l(K) \to C_{k+l-n}(L),$$
$$x * y = \sum_{\mu(z)=k+l-n} \#_{\mathsf{alg}} \mathcal{M}_{z;x,y}(L;H,K)\, z,$$

and it is isomorphic to the cup-product under the isomorphisms Φ_H, Φ_K and Φ_L (cf. [**16**] and [**18**]),

$$\Phi_H(\alpha) * \Phi_K(\beta) = \Phi_L(\alpha \cup \beta) \tag{23}$$

for all $\alpha, \beta \in H^*(M)$.

In order to combine this multiplicative structure with the refinement of Floer homology by the action filtration we need a suitable energy estimate. For this purpose we again generalize the definition of the moduli space $\mathcal{M}_{z;x,y}$ in terms of J-holomorphic sections in a suitably defined symplectic fibre bundle over the pair-of-pants surface Σ. We obtain:

Proposition 4.1. *Assume that* $[L] = [H] \circ [K]$ *in* $\widetilde{\mathrm{Ham}}(M,\omega)$. *Then, the pair-of-pants product* $*$ *is compatible with the filtration by the Hamiltonian action and the following diagram commutes,*

$$\begin{array}{ccc} HF_*^{(-\infty,a]}(H) \otimes HF_*^{(-\infty,b]}(K) & \xrightarrow{i_*^a \otimes i_*^b} & HF_*(H) \otimes HF_*(K) \\ \big\downarrow * & & \big\downarrow * \\ HF_*^{(-\infty,a+b]}(L) & \xrightarrow{i_*^{a+b}} & HF_*(L)\,. \end{array}$$

For example, as a concrete Hamiltonian generating the composition $\phi_K^1 \circ \phi_H^1$ one can choose

$$(H\#K)(t,x) = H(t,x) + K(t,(\phi_H^t)^{-1}(x)) \tag{24}$$

for $\phi_K^t \circ \phi_H^t$. Note that this operation preserves also the normalization.

Suppose $a > c(\alpha;H)$ and $b > c(\beta;K)$ so that $\Phi_H(\alpha) \in \operatorname{im} i_*^a$ and $\Phi_K(\beta) \in \operatorname{im} i_*^b$. Then, Proposition 4.1 implies that $c(\alpha \cup \beta; H\#K) \le a + b$. This proves:

Theorem 4.2. *Given any* $\alpha, \beta \in H^*(M)$ *with* $\alpha \cup \beta \neq 0$ *and* $\bar{\phi}, \bar{\psi} \in \widetilde{\mathrm{Ham}}(M, \omega)$ *it holds*

$$c(\alpha \cup \beta; \bar{\psi} \circ \bar{\phi}) \leq c(\alpha; \bar{\phi}) + c(\beta; \bar{\psi}) \,.$$

Note that by definition it is obvious that

$$c(\lambda\alpha; \bar{\phi}) = c(\alpha; \bar{\phi}) \quad \text{for all } \lambda \text{ with } \lambda\alpha \neq 0.$$

In particular, $c(\alpha \cup \beta; \bar{\phi}) = c(\beta \cup \alpha; \bar{\phi})$.

Remark on regularity. In order to prove that, for a generic choice of J, the solution space associated to the pair-of-pants model surface is a manifold, one has to exclude solutions which stay constantly on a fixed periodic solution. Such solutions trivially exist for $H = K$ and $L(t, x) = 2H(2t, x)$. This singular situation can be excluded by choosing a generic pair of regular Hamiltonians H, K which have no fixed points for their time-1-maps in common. Finally, in view of Proposition 2.14 we can approximate the case $H = K$ by such generic pairs such that all conclusions about critical levels remain valid.

4.1. The energy estimate for the pair-of-pants. In order to prove Proposition 4.1 we use a more general formulation of (20) given in terms of a connection on a symplectic fibre bundle.

Let Σ_o be a compact Riemann surface of genus 0 with three boundary components two of which are oriented as exits and one as entry, denoted by $\partial_1^+\Sigma_o$, $\partial_2^+\Sigma_o$ and $\partial_-\Sigma_o$. Assume that $E \to \Sigma_o$ is a smooth locally trivial fibre bundle with a closed 2-form $\bar{\omega} \in \Omega^2(E)$, $d\bar{\omega} = 0$, such that $(E_z, \bar{\omega}_{|T(E_z)})$ is a symplectic manifold and the typical fibre is (M, ω). Recall the symplectic fibre bundle $M_\phi \to S^1$ from (17) associated to $\phi \in \mathrm{Ham}(M, \omega)$. Given three symplectomorphisms ϕ, ψ, η we assume that there are symplectic diffeomorphisms

$$M_\phi \xrightarrow{\sim} E_{|\partial_1^+\Sigma_o},\ M_\psi \xrightarrow{\sim} E_{|\partial_2^+\Sigma_o},\ M_\eta \xrightarrow{\sim} E_{|\partial^-\Sigma_o}. \tag{25}$$

Note that such diffeomorphisms lead to trivializations over S^1 if the symplectomorphisms are isotopic to the identity. The simple but crucial observation is:

Lemma 4.3. *Given* ϕ *and* ψ*, such a symplectic fibre bundle* $(E, \bar{\omega})$ *satisfying* (25) *and such that the fibrewise symplectic form* $\bar{\omega}$ *is closed on* E *exists if* $\eta = \psi \circ \phi$.

Proof. Consider the domain $D \subset \mathbb{C}$ as sketched in Figure 1 and impose an equivalence relation $\sim$ on $D \times M$ by identifying boundary points as indicated using the symplectomorphisms ϕ, ψ and $\psi\phi$. Details are left to the reader.

Note that the induced M-fibre bundle $E = (D \times M)/\sim$ over the pair-of-pants surface Σ_o carries the closed fibrewise symplectic form $\bar{\omega} \in \Omega^2(E)$ canonically induced from (M, ω). $\square$

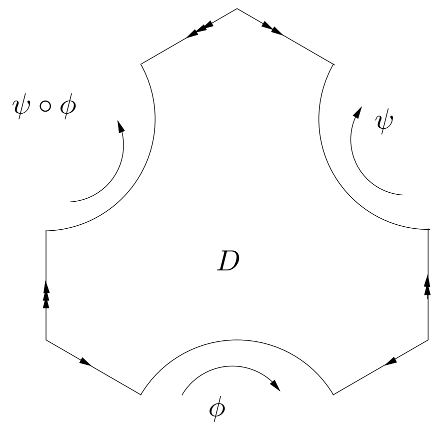


Figure 1. Boundary identification for the pants bundle.

Consider now exactly this fibrewise symplectic form $\bar{\omega}$ and choose an almost complex structure $\bar{J}$ such that the projection map $\pi\colon E \to \Sigma_o$ is $\bar{J}$-holomorphic, and restricts to an ω-compatible almost complex structure on the fibres E_z, $z \in \Sigma_o$. Associated to $\bar{\omega}$ we have the following connection on E,

$$T_\xi E = \mathrm{Ver}_\xi \oplus \mathrm{Hor}_\xi, \quad \xi \in E,$$
$$\mathrm{Ver}_\xi = (\ker D\pi(\xi)\colon T_\xi E \to T_{\pi(\xi)}\Sigma_0),$$
$$\mathrm{Hor}_\xi = \{\, v \in T_\xi E \,|\, \bar{\omega}(v,w) = 0 \text{ for all } w \in \mathrm{Ver}_\xi \,\}.$$

The connection map $K\colon TE \to \mathrm{Ver}$ is the projection w.r.t. this connection. Given a section $v\colon \Sigma_o \to E$ we denote the covariant derivative by

$$\nabla v = K \circ Dv, \quad \nabla v(z) \in \mathrm{Hom}(T_z\Sigma_o, T_{v(z)}(E_z)). \tag{26}$$

Recall that, since we have a fixed conformal structure on Σ_o and a vertically Riemannian structure by $\bar{\omega}(\cdot, \bar{J}\cdot)_{|\,\mathrm{Ver}}$, the L^2-norm $\int_{\Sigma_o} |\nabla v|^2_{\bar{J}}$ of ∇v for any section v is intrinsically defined.

What is crucial for the following is that for the above defined form $\bar{\omega}$ on the bundle $E \to \Sigma_o$ we have

$$\int_{\Sigma_o} |\nabla v|^2_{\bar{J}} = 2\int_{\Sigma_o} v^*\bar{\omega} + \int_{\Sigma_o} |\overline{\partial} v|^2_{\bar{J}} \tag{27}$$

where $\overline{\partial} v = \bar{J}\nabla v\, i + \nabla v$, and consequently

$$\int_{\Sigma_o} v^*\bar{\omega} \geq 0 \quad \text{for } \overline{\partial} v = 0. \tag{28}$$

We call a section v $\bar{J}$-holomorphic if $\overline{\partial}v = 0$. This is not true for an arbitrary closed form $\tilde{\omega} \in \Omega^2(E)$ which restricts fibrewise to ω. The above form $\bar{\omega}$ leads to a so-called *flat* connection.

Let us now combine this energy estimate with the Floer trajectories from (20). Given $\phi, \psi \in \mathrm{Ham}(M,\omega)$ we consider the trivial symplectic fibre bundles

$$M_\phi \times [0,\infty), \quad M_\psi \times [0,\infty), \quad \text{and } M_{\psi\phi} \times (-\infty, 0],$$

and glue them to $E \to \Sigma_o$ along the boundary circles by means of the gluing maps from (25). This gives a symplectic fibre bundle over a Riemann surface Σ, a pair-of-pants, which we denote again by E. We have extensions of the structures on Σ_o, $\bar{\omega}$, $\bar{J}$ and K. Given a continuous section $v\colon \Sigma \to E$ we denote by $v(\partial_i^\pm \Sigma)$ the uniform limits of v as paths in $\mathcal{L}_\phi^o$, $\mathcal{L}_\psi^0$ or $\mathcal{L}_\eta^0$ as we approach the respective ends of Σ, i.e., in cylindrical coordinates $s \to \pm\infty$. Observe that in terms of the cylindrical coordinates we have for a $\bar{J}$-holomorphic section $\overline{\partial}v = 0$

$$\tfrac{1}{2}|\nabla v|^2 = |v_s|^2 ds \wedge dt.$$

Note that the structure $\bar{\omega}$ on E still satisfies the flatness condition (27). The crucial estimate analogous to Proposition 3.10 for this fibre bundle over the surface Σ is:

Proposition 4.4. *Let v be a section of the pair-of-pants bundle $(E \to \Sigma_o, \bar{\omega})$ with boundary values*

$$v_{|\partial\Sigma_o} = (v_\phi, v_\psi, v_{\psi\phi}) \in \mathcal{L}_\phi^o \times \mathcal{L}_\psi^o \times \mathcal{L}_{\psi\phi}^o.$$

Then, we have

$$\int_{\Sigma_o} v^*\bar{\omega} = l_{\bar{\phi}}(v_\phi) + l_{\bar{\psi}}(v_\psi) - l_{\bar{\psi}\bar{\phi}}(v_{\psi\phi}), \quad \textit{for all } \bar{\phi}, \bar{\psi} \in \widetilde{\mathrm{Ham}}(M,\omega).$$

Proof of Proposition 4.4. Observe first that, due to $d\bar{\omega} = 0$ on E and $\omega_{|\pi_2(M)} = 0$, the value of $\int_{\Sigma_o} v^*\bar{\omega}$ only depends on the boundary values $v_{|\partial\Sigma_o}$. The proof is now exactly analogous to the computation of the Fredholm index in [**20**]. Namely, we construct symplectic fibre bundles $E^\pm \to D^\pm$ over the disks with both orientations which restrict respectively to M_ϕ, M_ψ and $M_{\psi\phi}$ over the boundary. Moreover, we have to construct fibrewise symplectic forms on these bundles which are closed. Using the obvious additivity for the integration it thus remains to compute the relevant formula for the symplectic fibre bundle over the disk, if the integration over the closed sphere S^2 as a base gives $\int_{S^2} v^*\bar{\omega} = 0$.

Let us construct the symplectic fibre bundle $E^+ \to D^+$ over the unit disk $D^+ = \{|z| \le 1\}$ in terms of the coordinates $z = \mathrm{e}^{2\pi(s+it)}$, $(s,t) \in$

$(-\infty,0]\times S^1$. Let $\beta\colon \mathbb{R}\to[0,1]$ be a smooth cut-off function

$$\beta(s)=\begin{cases}0, & s\le -2,\\ 1, & s\ge -1,\end{cases}$$

and choose a generating Hamiltonian $H(t,x)$ for $\phi=\phi^1_H$. Consider the smooth 2-parameter family of symplectomorphisms

$$\phi^t_s=\phi^t_{\beta(s)H},\quad \text{i.e., } \partial_t\phi^t_s=\beta(s)X_H(t,\phi^t_s).$$

Define the symplectic fibre bundle $E^+\to D^+$ with fibre (M,ω) by

$$E^+=(-\infty,0]\times\mathbb{R}\times M/\{\,(s,t,x)\sim(s,t-1,\phi^1_s(x))\,\}\to(-\infty,0]\times S^1$$

which is trivialized by

$$\begin{aligned}\Phi\colon E^+ &\xrightarrow{\sim}(-\infty,0]\times S^1\times M\\ [s,t,x]&\mapsto(s,t\pmod 1,\phi^t_s(x)).\end{aligned}$$

Via Φ, sections $\bar v(s,t)=[s,t,v(s,t)]$ of E^+ are identified with maps $u\colon (-\infty,0]\times S^1\to M$, $u(s,t)=\phi^t_s(v(s,t))$. Consider the following 2-form $\omega_E\in\Omega^2(E^+)$, given by $\omega_E=\Phi^*\alpha$ with $\alpha\in\Omega^2(\mathbb{R}\times S^1\times M)$,

$$\alpha_{(s,t,x)}=-\beta'(s)H(t,x)ds\wedge dt+\omega_x+i_{\beta(s)X_H(t,x)}\omega_x\wedge dt.$$

One computes

$$\omega_E=-\beta' H ds\wedge dt+\omega-i_{(D\phi)^{-1}\partial_s\phi}\omega\wedge ds. \tag{29}$$

Straightforward computation shows that

$$\begin{aligned}&\omega_{E|\text{fibre}}=\omega,\\ &d\omega_E=0,\\ &\pi_*\omega_E^{n+1}=0,\end{aligned}\tag{30}$$

where $\pi_*\colon \Omega^{2n+2}(E)\to\Omega^2(\mathbb{R}\times S^1)$ denotes the fibre integration map. The last property holds if H is normalized.

Lemma 4.5. *Let $E^+\to(-\infty,0]\times S^1$ and ω_E be given as above. Then*

$$\int_{D^+}\bar v^*\omega_E=l_{[H]}\big(\bar v(0)\big)$$

holds for any section $\bar v$ in E^+.

Proof. We can assume E^+ as canonically extended over $\mathbb{R}\times S^1$. Since $\omega_{|\pi_2}=0$ the left hand side does not depend on the section if $\bar v(\infty)$ is fixed. In view of Proposition 3.10 we can also assume that $\bar{v}(\infty,\cdot) = x \in \mathrm{Fix}^o\,\phi$. Denoting $(s,t,u(s,t)) = \Phi(\bar{v}(s,t))$ and $x(t) = u(\infty,t)$ we compute

$$\begin{aligned}
\iint \bar{v}^*\omega_E &= \iint w^*\alpha \\
&= \iint \alpha(w_s, w_t)\,ds\,dt \\
&= \iint \alpha(\partial_s + u_s, \partial_t + u_t)\,ds\,dt \\
&= -\iint \beta' H(t,u)dsdt + \iint u^*\omega + \iint i_{\beta X_H}\omega(u_s)dsdt \\
&= \iint u^*\omega - \iint [\beta'(s)H(t,u(s,t)) + \beta(s)dH(u(s,t))u_s]dsdt \\
&= \iint u^*\omega - \iint \frac{d}{ds}\big(\beta(s)H(t,u)\big)dsdt \\
&= \mathcal{A}_H(u(+\infty)).
\end{aligned}$$

□

Lemma 4.6. *Let $E \to S^2$ be a symplectic fibre bundle with a form ω_E satisfying* (30)*. Then for any section s of E we have*

$$\int_{S^2} s^*\omega_E = I(\gamma)$$

if E is obtained from gluing two trivial bundles $E^\pm = D^\pm \times M$ along their boundary via a loop representing $\gamma \in \pi_1(\mathrm{Ham})$. In particular, if E is trivial then $\int_{S^2} s^*\omega_E = 0$.

Proof. Since we know that $\mathcal{A}_{\gamma[H]}(x) = \mathcal{A}_{[H]}(x) + I(\gamma)$ it suffices to show that

$$\int_{S^2} s^*\omega_E = 0$$

in case E is trivial. Let $p\colon E \to M$ be the projection map obtained from a trivialization $E \cong S^2 \times M$ and $\pi\colon E \to S^2$ the canonical projection. Denoting by $\sigma \in H^2(S^2)$ the generator normalized by $\sigma(S^2) = 1$ we have

$$\{\omega_E\} = p^*\{\omega\} + a\pi^*\sigma, \quad a = \int_{S^2} s^*\omega_E\,.$$

Then, the last condition of (30) implies that $a = \int_{S^2} \pi_*\omega_E^{n+1} = 0$. □

The proof of Proposition 4.4 now follows from Lemmata 4.5 and 4.6 if we glue the bundles $E^-_{[H]}$, $E^-_{[K]}$ over D^- and $E^+_{[H\#K]}$ over D^+ to $E \to \Sigma_o$ over the pair-of-pants. We obtain a trivial (M,ω)-fibre bundle $\pi\colon \bar{E} \to S^2$ together with a coupling form $\bar{\omega}$ satisfying (30). □

Recalling the energy identity (28) for $\bar{J}$-holomorphic sections in the pair-of-pants bundle $E \to \Sigma_o$ we obtain the energy estimate:

Corollary 4.7. *Every $\bar{J}$-holomorphic section $v\colon \Sigma \to E$ with the boundary condition $v(\partial_1^+\Sigma) = x(0) \in \mathrm{Fix}_o(\phi)$, $v(\partial_2^+\Sigma) = y(0) \in \mathrm{Fix}_o(\psi)$ and $v(\partial^-\Sigma) = z(0) \in \mathrm{Fix}_o(\psi\phi)$ satisfies the energy estimate for automorphisms $\phi = \phi_H^1$, $\psi = \phi_K^1$,*

$$0 \le \int_\Sigma |\nabla v|^2_{\bar{J}} = \int_\Sigma v^*\bar{\omega} = \mathcal{A}_{[H]}(x) + \mathcal{A}_{[K]}(y) - \mathcal{A}_{[K][H]}(z).$$

Observe that such a positivity estimate for $\bar{J}$-holomorphic sections in the bundle $E^\pm \to D^\pm$ does not hold because the closed form $\bar{\omega}$ has non-vanishing $ds \wedge dt$-components, see (29). There is no equivalent of (27).

Let the moduli space of such $\bar{J}$-holomorphic sections of $E \to \Sigma$ with boundary condition as in Corollary 4.7 replace the originally considered space $\mathcal{M}_{z;x,y}(L; H, K)$. Then, analogously to Proposition 3.11 we obtain a multiplication

$$*\colon C_k(\phi) \times C_l(\psi) \to C_{k+l-n}(\psi\phi)$$

which coincides with the original pair-of-pants product on the level of Floer homology. The sharp energy estimate from Corollary 4.7 concludes the proof of Proposition 4.1.

4.2. Poincaré duality. Let us consider the dual Floer complex associated to symplectic fixed points by applying the Hom-functor. For sake of simplicity we restrict ourselves to coefficients in a field $\mathbb{F}$, e.g., $\mathbb{Z}_2$, $\mathbb{Q}$ or $\mathbb{R}$. Given $a \in \mathbb{R}$ we have the cochain complex

$$C^k_{(a,\infty)}(H) = \{\, x \in \mathcal{P}_1(H) \mid \mu(x) = k,\ \mathcal{A}_H(x) > a \,\} \otimes \mathbb{F},$$
$$\delta\colon C^k_{(a,\infty)} \to C^{k+1}_{(a,\infty)}, \quad \delta x = \sum_{\mu(y)=\mu(x)+1} \#_{\mathsf{alg}} \widehat{\mathcal{M}}_{x,y}(J,H)\, y,$$

and we identify

$$C^*(H) = \mathrm{Hom}(C_*(H)), \quad C^*_{(a,\infty)} = \mathrm{Hom}(C_*^{(a,\infty)}),$$
$$\text{and} \quad C^*_{(-\infty,a]} = C^*/C^*_{(a,\infty)} = \mathrm{Hom}\left(C_*^{(-\infty,a]}\right).$$

We observe that the long exact cohomology sequence induced from the short exact sequence of cochain complexes

$$0 \to C^*_{(a,\infty)} \xrightarrow{j_a^\bullet} C^* \xrightarrow{i_a^\bullet} C^*_{(-\infty,a]} \to 0 \tag{31}$$

equals the dual sequence obtained by the Hom-functor from the long exact homology sequence (2). In view of the universal coefficient theorem we have

$$C^*(H) = \mathrm{Hom}(C_*(H), \mathbb{F}), \quad HF^*(H) \cong \mathrm{Hom}(HF_*(H); \mathbb{F}),$$

respectively for $C^*_{(a,\infty)}$, etc.

Let us consider now the dual isomorphism $\Psi^*\colon H_*(M) \to HF^{n-*}(H)$ and the dual generators $[\mathrm{pt}] \in H_0(M)$ and $1 \in H^0(M)$. In view of the long exact cohomology sequence obtained from (31) we have:

Lemma 4.8. *The critical value $c(1;H)$ can be equivalently represented in Floer cohomology by*

$$\inf\left\{\, a \in \mathbb{R} \,|\, j^a_*\big(\Phi_H(1)\big) = 0 \,\right\} \;=\; \sup\left\{\, a \in \mathbb{R} \,|\, i^*_a\big(\Psi^*_H([\mathrm{pt}])\big) = 0 \,\right\}.$$

Proof. We have $\Phi_H(1) \in \operatorname{im} i^a_*$ if and only if there exists a $u_a \in HF_n^{(-\infty,a]}$ such that $\big\langle u_a, i^*_a(\Psi^*_H([\mathrm{pt}]))\big\rangle \neq 0$. That is,

$$j^a_*\big(\Phi_H(1)\big) = 0 \quad \text{if and only if} \quad i^*_a\big(\Psi^*_H([\mathrm{pt}])\big) \neq 0\,.$$

□

We now use the fact that Poincaré duality is represented in terms of Floer homology by the canonical isomorphism between the homology $HF_*(H)$ and the cohomology $HF^*(H^{(-1)})$ of the Hamiltonian generating the inverse symplectomorphism,

$$H^{(-1)}(t,x) = -H(-t,x), \quad \phi^t_{H^{(-1)}} = \phi^{-t}_H.$$

It is straightforward to verify that the identification

$$\mathcal{P}_1(H) \cong \mathcal{P}_1(H^{(-1)}), \quad x^{-1}(t) = x(-t),$$
$$\mathcal{M}_{x,y}(J,H) \cong \mathcal{M}_{y^{-1},x^{-1}}(J,H^{(-1)}), \quad u^{-1}(s,t) = u(-s,-t),$$

provides an identification of the chain complex of H with the cochain complex of $H^{(-1)}$,

$$\big(C_*(H),\partial\big) \cong \big(C^{-*}(H^{(-1)}),\delta\big).$$

Note that it holds

$$\mu(x^{-1}) = -\mu(x) \quad \text{and} \quad \mathcal{A}_{H^{(-1)}}(x^{-1}) = -\mathcal{A}_H(x).$$

Hence we have the identification of $C_*^{(-\infty,a]}(H)$ with $C^{-*}_{[-a,\infty)}(H^{(-1)})$, etc., and the short exact sequence of chain complexes

$$0 \to C_*^{(-\infty,a]}(H) \xrightarrow{i^a_*} C_*(H) \xrightarrow{j^a_*} C_*^{(a,\infty)}(H) \to 0$$

becomes isomorphic to the short exact sequence of cochain complexes

$$0 \to C^*_{[-a,\infty)}(H^{(-1)}) \xrightarrow{j^*_{-a}} C^*(H^{(-1)}) \xrightarrow{i^*_{-a}} C^*_{(-\infty,-a)}(H^{(-1)}) \to 0.$$

In [**20**], Poincaré duality was analyzed in terms of a canonical non-degenerate bilinear form β on Floer homology associated to a Riemann surface with two cylindrical ends both oriented as entries. This is equivalent to the identification of the homology for H with the cohomology for $H^{(-1)}$ because the change $H \to H^{(-1)}$ corresponds to the change of orientation of a cylindrical end. Hence, we have the commutative diagram

$$\begin{array}{ccc} H^k(M) & \xrightarrow{\Phi_H} & HF_{n-k}(H) \\ {\scriptstyle \mathrm{PD}}\big\downarrow & & {\scriptstyle =}\big\downarrow \\ H_{2n-k}(M) & \xrightarrow{\Psi^*_{H^{(-1)}}} & HF^{k-n}(H^{(-1)})\,. \end{array} \tag{32}$$

Proposition 4.9. *The representation of Poincaré duality in Floer homology yields the identity*

$$c([M];\phi) = -c(1;\phi^{-1}), \quad \textit{for all } \phi \in \widetilde{\mathrm{Ham}}(M,\omega).$$

Proof. From (32) we obtain that

$$c([M];H) = \inf\{\, a \mid j^a_*(\Phi_H([M])) = 0\,\} = \inf\{\, a \mid i^*_{-a}(\Psi^*_{H^{(-1)}}([\mathrm{pt}])) = 0\,\}.$$

Hence, the assertion follows by Lemma 4.8. □

An immediate consequence is:

Corollary 4.10. *Given any nonzero cohomology class $\alpha \in H^*(M)$ and $\phi, \psi \in \widetilde{\mathrm{Ham}}(M,\omega)$ we have the estimate*

$$c([M],\psi) \le c(\alpha;\psi\phi) - c(\alpha;\phi) \le c(1,\psi)$$

and $\phi \mapsto c(\alpha,\phi)$ is continuous with respect to $d_{\tilde{H}}(\phi,\mathrm{id}) = \inf\{\|H\| | \phi = [H]\}$.

Proof. First we combine

$$c(\alpha;\phi) \le c(\alpha;\psi\circ\phi) + c(1;\psi^{-1})$$

from Theorem 4.2 with Corollary 4.10. The continuity follows from

$$c(1,[H]) - c([M],[H]) \le \|H\|.$$

Recall that $c(1,[H]) \le E_+(H)$ and $0 \le E_+(H)$ because H is normalized. □

We can now compute the action of $\pi_1(\mathrm{Ham})$ on these continuous sections $c(\alpha)$ in the action spectrum bundle.

Proposition 4.11. *For any $\alpha \in H^*(M) \setminus \{0\}$, $\gamma \in \pi_1(\mathrm{Ham})$, $\phi \in \widetilde{\mathrm{Ham}}(M,\omega)$ we have*

$$c(\alpha;\gamma\phi) = c(\alpha;\phi) + I(\gamma).$$

Proof. Considering the covering $\pi\colon \widetilde{\mathrm{Ham}} \to \mathrm{Ham}$ we identify $\pi_1(\mathrm{Ham}) = \pi^{-1}(\mathrm{id})$. Given $\gamma \in \pi^{-1}(\mathrm{id})$ the action spectrum of γ consist only of one value, $\Sigma_\gamma = \{I(\gamma)\}$, hence

$$c(\alpha,\gamma) = I(\gamma) \quad \text{for all } \alpha \in H^*(M) \setminus \{0\}.$$

Then assertion then follows from Corollary 4.10. □

This result allows us to define the following function on $\mathrm{Ham}(M,\omega)$.

Definition 4.12. Given $\phi \in \widetilde{\mathrm{Ham}}(M,\omega)$ we define

$$c_-(\phi) = c([M];\phi), \quad c_+(\phi) = c(1;\phi) \quad \text{and} \quad \gamma(\phi) = c_+(\phi) - c_-(\phi).$$

We obtain $\gamma\colon \mathrm{Ham}(M,\omega) \to \mathbb{R}$ as a continuous function.

Summing up the above results we have

$$0 \le \gamma(\phi) = \gamma(\phi^{-1}) \quad \text{and} \quad \gamma(\phi \circ \psi) \le \gamma(\phi) + \gamma(\psi) \tag{33}$$

for all $\phi, \psi \in \mathrm{Ham}_{p_o}(M,\omega)$. This function γ plays the role of the nontrivial *selector* of Hofer and Zehnder in the case of compactly supported Hamiltonian automorphisms of $(\mathbb{R}^{2n},\omega_o)$.

4.3. Vanishing of the monodromy I. We now show that the homomorphism $I\colon \pi_1(\mathrm{Ham}) \to \mathbb{R}$ is in fact trivial in the symplectically aspherical case (A).

Let $M^{2n} \overset{i}{\hookrightarrow} E \overset{\pi}{\longrightarrow} S^2$ be a Hamiltonian fibre bundle and $\omega_E \in \Omega^2(E)$ be a coupling form, i.e., satisfying (30). Let $\omega_{|\pi_2(M)} = c_{1|\pi_2(M)} = 0$ and choose a generic almost complex structure J on E such that J is fibrewise ω-compatible, and the projection π is J-i-holomorphic. Then, the space $\mathcal{M}(J)$ of holomorphic sections $s\colon S^2 \to E$ is a closed manifold of dimension $2n$.

Theorem 4.13 ([**22**]). *The evaluation map ev_{z_o} at any point $z_o \in S^2$ induces a homomorphism $ev_*\colon H_{2n}(\mathcal{M}(J),\mathbb{Z}) \to H_{2n}(M,\mathbb{Z})$ of degree ± 1.*

Proof. Given ω_E and a generic J on E, Seidel defines in [**22**] the invariant

$$Q(E,\omega_E,S) = \sum_{\gamma\in\Gamma} \left[\, \mathrm{ev}_{z_o}\left(S(j,J,\gamma+S)\right)\right] \otimes \langle\gamma\rangle \in QH_d(M,\mathbb{Z}_2),$$

where $z_o \in S^2$, $\Gamma = \pi_2(M)/\ker\omega_{|\pi_2} \cap \ker c_{1|\pi_2}$. This is an invariant of a given equivalence class S of a section of E. The degree of the quantum homology class Q is given by $d = 2n + 2c_1(TE^v,\omega_E)(S)$ where TE^v is the vertical subbundle. Since by assumption $\omega_{|\pi_2} = c_{1|\pi_2} = 0$ we can assume coefficients in $\mathbb{Z}$, $\Gamma = 0$, and the quantum homology equals ordinary homology, the invariant

$$Q(E,\omega_E,S) \in H_{2n}(M,\mathbb{Z}) \cong \mathbb{Z}$$

is independent of S. Seidel's main result is that $Q(E,\omega_E)$ is an invertible element in QH_* of homogeneous degree. Hence $Q(E,\omega_E) = \pm 1 \in \mathbb{Z}$. □

Corollary 4.14. *For any section $s\colon S^2 \to E$ it holds $s^*[\omega_E] = 0$ in $H^2(M,\mathbb{R})$.*

Proof. Clearly, the result does not depend on the choice of the section s. Let $u \in \mathcal{M}(J)$ be a holomorphic section for a generic J on E and consider the commutative diagram

$$\begin{array}{ccc} \mathcal{M}(J) \times S^2 & \xrightarrow{\text{ev}} & E \\ \uparrow{\scriptstyle \tilde{u}} & & \downarrow{\scriptstyle \pi} \\ S^2 & \xrightarrow{=} & S^2 \end{array}$$

where $\text{ev}(w, z) = w(z)$ is a bundle map and $\tilde{u}(z) = (u, z)$ is a constant section in the trivial bundle. We have $u^*[\omega_E] = \tilde{u}^* \text{ev}^*[\omega_E]$ and $\text{ev}^*[\omega_E] = \alpha \times 1 + a1 \times \sigma$, $a \in \mathbb{R}$, for $\alpha \in H^2(\mathcal{M}(J), \mathbb{R})$ and $\sigma \in H^2(S^2, \mathbb{Z})$ a generator. Hence,

$$u^*[\omega_E] = a\sigma \in H^2(S^2, \mathbb{R})$$

and we have to prove that $a = 0$.

Since the fibre integration homomorphism $\pi_* \colon H^{2n+2}(E) \to H^2(S^2)$ is an isomorphism, the assumption (30) on the form ω_E implies that the class $[\omega_E]^{n+1}$ vanishes, hence $a\alpha^n \times \sigma = 0$. Since $\alpha = \text{ev}_{z_o}^*[\omega]$ for $\text{ev}_{z_o} \colon \mathcal{M}(J) \to \pi^{-1}(z_o) \approx M$ for any fixed $z_o \in S^2$, we have

$$\big\langle \alpha^n, [\mathcal{M}(J)] \big\rangle = \deg(\text{ev}_*) = \pm 1$$

by Seidel's theorem. Thus, it follows that $a = 0$. □

This proves Theorem 1.1 on the vanishing of the obstruction homomorphism I in the symplectically aspherical case.

Corollary 4.15. *The sections $c(\alpha)$ in the action spectrum bundle are well-defined over* $\text{Ham}(M, \omega)$ *and continuous in the Hofer-metric* $d_H(\text{id}, \phi) = \inf\{\, \|H\| \,|\phi = \phi_H^1 \,\}$.

Proof. Combine Corollary 4.10, Proposition 4.11 and Corollary 4.14. □

This concludes the proof of Theorem 1.2.

5. The bi-invariant metric γ and a relative capacity.

Let us first analyze the relation between the function γ and the well-known displacement energy introduced by Hofer. Let $H \colon [0, 1] \times M \to \mathbb{R}$ be a normalized Hamiltonian and $\psi \in \text{Ham}(M, \omega)$ an automorphism separating the support set of ϕ_H^1

$$\psi(\mathcal{S}(H)) \cap \mathcal{S}(H) = \emptyset \quad \mathcal{S}(H) = \bigcup_{t \in [0,1]} \operatorname{supp} X_H(t, \cdot). \tag{34}$$

As in [**8**] and [**23**] we observe that $\text{Fix}(\phi_H^t \circ \psi) = \text{Fix}(\psi \circ \phi_H^t) = \text{Fix}(\psi)$ for all $t \in [0, 1]$ and for $x \in \text{Fix}^o \psi$, $\psi = \phi_K^1$ we have

$$\mathcal{A}_{[K][H]}(x) = \mathcal{A}_{[K]}(x) + \int_0^1 H(t, x) dt.$$

Setting $a(H) = \int_0^1 H(t,p)dt$ for any $p \notin \mathcal{S}(H)$ we obtain for the action spectra

$$\Sigma_{[K][H]} = \Sigma_{[K]} + a(H).$$

Considering the continuous path $\epsilon \mapsto \epsilon H$, $\epsilon \in [0,1]$ we obtain for every $\alpha \neq 0$ the continuous function

$$\epsilon \mapsto c(\alpha;[\epsilon H][K]) - a(\epsilon H) = c(\alpha;[K][\epsilon H]) - a(\epsilon H)$$

into the nowhere dense action spectrum $\Sigma_{[K]} \subset \mathbb{R}$ which therefore has to be constant. It follows that

$$\gamma(\phi_H^1 \psi) = \gamma(\psi \phi_H^1) = \gamma(\psi). \tag{35}$$

In particular, this implies:

Proposition 5.1. *Given H and ψ as in* (34) *we have*

$$\gamma(\phi_H^k) \leq 2\gamma(\psi)$$

for all $k \in \mathbb{Z}$, where $\phi_H^k = (\phi_H^1)^k$.

Proof. The triangle inequality for γ (33) yields

$$\gamma(\phi_H^1) \leq \gamma(\psi \circ \phi_H^1) + \gamma(\psi).$$

For $k \in \mathbb{Z}$ use $\phi_H^k = \phi_{H^k}^1$ with $H^k(t,x) = kH(kt,x)$ and that condition (34) holds for all $k \in \mathbb{Z}$ since we can assume $H(t+1,x) = H(t,x)$. □

Moreover, we have:

Proposition 5.2. *Given any open subset $U \subset M$ there exists $\phi_H^1 \in$ $\mathrm{Ham}(M,\omega)$ such that $\mathcal{S}(H) \subset U$ and $\gamma(\phi_H^1) > 0$.*

Proof. We pick a smooth positive function $H\colon M \to \mathbb{R}$ independent of t with $\operatorname{supp} H \subset U$ such that the only critical point $p \in \operatorname{supp} H$ is a maximum and the C^2-norm of H is small enough so that the only 1-periodic solution are the constant solutions $p \in \mathcal{P}_1(H)$ and $q \in M \setminus \operatorname{supp} H$. Approximating H suitably by C^2-small Morse functions we obtain $c_-(H) = -H(p)$ and $c_+(H) = 0$, i.e., $\gamma(H) = \|H\| > 0$. Recall that here we do not need H to be normalized. □

Combining this observation with Proposition 5.1 we deduce the metric property of γ.

Theorem 5.3. *For every $\phi \in \mathrm{Ham}(M,\omega)$ we have*

$$\gamma(\phi) > 0 \quad \textit{if} \quad \phi \neq \mathrm{id}\,.$$

That is, $\gamma\colon \mathrm{Ham}(M,\omega) \to \mathbb{R}_+$ defines a metric d_γ by

$$d_\gamma(\phi,\psi) = \gamma(\phi\psi^{-1}).$$

Moreover, this metric is bi-invariant.

Proof. It remains to show that d_γ is bi-invariant. This follows analogously to d_H in [**8**] from the fact that for any $\theta \in \mathrm{Aut}(M,\omega)$ we have

$$\theta \circ \phi_H^t \circ \theta^{-1} = \phi_{H_\theta}^t \quad \text{for all } t,$$

where $H_\theta(t,x) = H(t,\theta^{-1}(x))$. □

Remark 5.4. This result implies another proof for the nontrivial fact that Hofer's metric d_H is in fact a metric. This has been shown for all closed symplectic manifolds in [**10**]. Here we obtain a different Floer-theoretical proof for at least symplectically aspherical manifolds.

An interesting application of this bi-invariant metric is obtained analogously to [**8**].

Theorem 5.5. *Let $\phi \in \mathrm{Ham}(M,\omega)$ such that there exists a uniform bound*

$$\gamma(\phi^n) \leq C \quad \text{for all } n \in \mathbb{N},$$

then ϕ has infinitely many nontrivial geometrically distinct periodic points corresponding to contractible periodic solutions.

Proof. Let H be a normalized Hamiltonian generating $\phi = \phi_H^1$.Then $\phi^n = \phi_{H^{(n)}}^1$. Assume that ϕ has only finitely many nontrivial periodic points, then without loss of generality the spectra are related by the scaling with n, $\Sigma_{\phi^n} = n \cdot \Sigma_\phi$. But if $\phi \neq \mathrm{id}$ then $c_+(\phi^n) - c_-(\phi^n) > 0$ for all $n \in \mathbb{N}$ and necessarily $\gamma(\phi^n) \to \infty$ contradicting the assumption. □

Clearly examples of such automorphisms with uniformly bounded γ-distance exist. For example, if the support $\bigcup_{t\in[0,1]} \mathrm{supp}\,\phi^t$ can be disjoined from itself by a Hamiltonian isotopy.

Let us study the following examples which show that, in general however, there is no upper bound on γ.

5.1. Examples for infinite diameter of $\mathrm{Ham}(M,\omega)$.

Example 5.6. Consider the autonomous Hamiltonian function

$$H\colon S^1 \times S^1 \to \mathbb{R}, \quad H(x,y) = \sin 2\pi x.$$

Since we consider only contractible 1-periodic solutions, $\mathcal{P}_1(H)$ equals the set of critical points of H,

$$\mathcal{P}_1(H) = \{\pi\} \times S^1 \cup \{-\pi\} \times S^1.$$

Moreover, we see that the action spectrum contains only two values, $\sigma(\phi_H^1) = \{-1,1\}$ and since $\gamma(\phi_H^1) > 0$ we have $\gamma(\phi_H^1) = 2$. We conclude that

$$\gamma(\phi_H^k) = 2k \to \infty \quad \text{for } k \to \infty\,.$$

Clearly, it is straightforward to generalize this observation for $\mathbf{T}^2$ to any symplectically aspherical manifold, i.e., $\omega_{|\pi_2} = 0$, admitting an incompressible Lagrangian torus.

A subset $A \subset M$ is called **incompressible** if the inclusion map gives an injection for the fundamental group.

Example 5.7. Assume that $T^n \hookrightarrow M$ is an incompressible Lagrangian torus. By Weinstein's theorem a tubular neighbourhood is symplectomorphic to a neighbourhood of the zero section in T^*T^n which is a product of T^*S^1. We can now find an autonomous Hamiltonian with compact support in this neighbourhood which factorizes according to the product structure and is independent of the base variables. Thus the only nonconstant periodic solutions are non-contractible and we find a Hamiltonian automorphism ϕ_H such that $\gamma(\phi_H^k) \to \infty$ as $k \to \infty$.

Example 5.8. Clearly, this argument also works for a general incompressible Lagrangian submanifold which admits a Riemannian metric all of whose contractible closed geodesics are constant. For example, if it admits a metric of nonpositive sectional curvature. This example recovers the result by Lalonde and Polterovich in [**12**] which is proven without restriction to $\omega_{|\pi_2(M)}$.

Example 5.9. It is also easy to recover here, in the more restrictive case of a symplectically aspherical manifold, the example of products with surfaces of genus ≥ 1 as given in [**11**].

More generally one can consider symplectic fibrations $(M,\omega) \hookrightarrow (E,\omega_E) \xrightarrow{\pi} (B,\sigma)$ where ω_E is a symplectic form such that its restriction to the induced horizontal subbundle of TE with respect to the associated connection equals the pull-back $\pi^*\sigma$. Assume that (B,σ) is a symplectic manifold of the previously mentioned kind with a Hamiltonian of arbitrarily high oscillation norm $\|H\|$ but without nontrivial contractible 1-periodic solutions. Then the Hamiltonian vector field on E of the pull-back π^*H lies in the horizontal sub-bundle of TE and $D\pi$ identifies it with X_H on B. Hence its non-constant periodic solutions have to be non-contractible too.

5.2. Comparison with Hofer-Zehnder capacity. Let us now compare the metric γ with the Hofer-Zehnder capacity for symplectic manifolds which is defined via so-called admissible Hamiltonians.

Definition 5.10. We call a Hamiltonian function $H\colon S^1 \times M \to \mathbb{R}$ **admissible** if its set of 1-periodic contractible solutions $x \in \mathcal{P}_1(H)$ contains only constant solutions $\dot{x} = 0$, i.e., $x(0) = x(t) \in \operatorname{Crit} H(t,\cdot)$ for all $t \in S^1$.

A smooth Hamiltonian $H\colon [0,1] \times M \to \mathbb{R}$ is called **quasi-autonomous** if there exist two points $x_+, x_- \in M$ such that maximum and minimum of H are attained at x_+ and x_- uniformly for all t. Here we will consider a slightly stronger condition:

$$\begin{aligned} &\max_x H(t,x) = H(t,p) \quad \text{for all } t \in [0,1],\ p \in U_+, \\ &\min_x H(t,x) = H(t,p) \quad \text{for all } t \in [0,1],\ p \in U_- \end{aligned} \tag{36}$$

for some disjoint open neighbourhoods $U_-, U_+ \subset M$ of x_-, x_+.

Theorem 5.11. *Every quasi-autonomous Hamiltonian K satisfying* (36) *which is admissible and homotopic to* 0 *through admissible Hamiltonians satisfies*

$$\gamma(\phi_K^1) = \|K\|.$$

Throughout this section we do not assume the Hamiltonian functions to be normalized without further notice. From Proposition 2.14 we know that the functions $c(\alpha, H)$ are well-defined and continuous with respect to $\|H\|$ without any normalization condition. We also still have the symmetry from Poincaré duality, $c_+(H^{-1}) = -c_-(H)$. Moreover, in view of (15), also $\gamma(\phi_H^1) = c_+(H) - c_-(H)$ does not depend on the normalization.

Let K be an admissible Hamiltonian such that there exists an open subset $U \subset M$ with

$$K(t,p) = \max_{x \in M} K(t,x) \quad \text{for all } t \in S^1,\ p \in U. \tag{37}$$

Recall from Definition 2.4, $E_-(K) = -\int_0^1 \max_M K(t,\cdot)\,dt$.

Proposition 5.12. *Assume that an admissible Hamiltonian K satisfying* (37) *is homotopic to 0 through admissible Hamiltonians. Then any regular Hamiltonian $H \in \mathcal{H}_{reg}$ satisfies*

$$c_-(H) \le E_-(K) + \|H - K\|.$$

Clearly, an autonomous Hamiltonian K is homotopic to 0 through admissible Hamiltonians if every T-periodic contractible solution of $\dot{x} = X_K(x)$ for $0 < T \le 1$ is constant, namely take the homotopy $(\tau H)_{\tau \in [0,1]}$.

Proof of Proposition 5.12. The proof is based on a suitable variation of the definition of the map $\Phi_H \colon H^{2n}(M) \to HF_{-n}(H)$ as given in (3). Given any Morse function f, we known that any local maximum $p \in \mathrm{Crit}_{2n} f$ represents the top cohomology class in terms of Morse cohomology, $[M] = \{p\} \in H^{2n}(f)$. Moreover, any negative gradient flow trajectory for f converging towards p has to lie constant in p. Therefore, we can identify the moduli space used for the definition of $\Phi_H([M])$ as

$$\mathcal{M}^-_{y;p}(H,J;f) = \{\, (u,p) \mid u \in \mathcal{M}^-_{y;p}(H,J) \,\},$$

where

$$\mathcal{M}^-_{y;p}(H,J) = \Big\{ u\colon \mathbb{R}\times S^1 \to M \,\Big|\, \partial_s u + J\big(\partial_t u - \beta(s) X_H(u)\big) = 0,$$
$$\int_{-\infty}^{\infty} |\partial_s u|^2 ds dt < \infty,$$
$$u(-\infty) = y, \quad u(+\infty) = p \Big\}$$

and $\beta(s) = 1$ for $s \le -1$, $\beta(s) = 0$ for $s \ge 0$. We therefore have

$$\Phi_H([M]) = \sum_{\mu(y)=-n} \#_{\mathsf{alg}} \mathcal{M}^-_{y;p}(H,J)\, y.$$

Let us now consider the homotopy between H and K

$$G_s(t,x) = \beta(s)H(t,x) + \big(1-\beta(s)\big)K(t,x),$$

so that the associated Hamiltonian vector field satisfies $X_{G_s}(t,x) = 0$ for $s \ge 0$ and $x \in U$. The associated Cauchy-Riemann type flow equation reads

$$u\colon \mathbb{R}\times S^1 \to M, \quad u_s + J\big(u_t - X_{G_s}(u)\big) = 0. \tag{38}$$

Analogously to the solutions in $\mathcal{M}^-_y(H,J)$, every finite energy solution u of (38), i.e., $\iint |u_s|^2 ds dt < \infty$, which satisfies $\lim_{s\to\infty} u(s,t) \in U$ for all $t \in S^1$ has a removable singularity at $+\infty$ and can be smoothly extended over $\mathbb{R}\times S^1 \cup \{+\infty\} \cong \mathbb{C}$. We have thus the well-defined solution space

$$\widetilde{\mathcal{M}}^-_{y;p}(H,K) = \{\, u\colon \mathbb{R}\times S^1 \cup \{\infty\} \to M \,|\, u \text{ solves (38)},$$
$$u(-\infty) = y,\ u(+\infty) = p \,\}.$$

Again, for a generic almost complex structure J, $\widetilde{\mathcal{M}}^-_{y;p}$ is a $(\mu(y)+n)$-dimensional manifold. To be precise, we allow almost complex structures J to be explicitly (s,t)-dependent for $|s| < 2$ and t-dependent for $t \le -2$. (For details cf. [**20**].) Our aim is to define the element $\Phi_{HK} \in HF_{-n}(H)$ by

$$\Phi_{HK} = \sum_{\mu(y)=-n} \#_{\mathsf{alg}} \widetilde{\mathcal{M}}^-_{y;p}(H,K)\, y. \tag{39}$$

The crucial point is to show that the 0-dimensional solution space $\widetilde{\mathcal{M}}^-_{y;p}(H,K)$ is compact. Recall that in view of the asphericity condition $\omega_{|\pi_2} = 0$ the only compactness obstruction can occur by splitting off of cylindrical Floer trajectories at either end. On the negative end, i.e., for $s \to -\infty$ this is prohibited by the regularity assumption on H, the transversality condition and the index restriction that $\dim \widetilde{\mathcal{M}}^-_{y;p} = 0$. It remains to rule out splitting off at the positive end of u. Suppose we have such a weak

C^∞_{loc}-convergence to a split solution for a sequence of suitably reparameterized trajectories in $\widetilde{\mathcal{M}}^-_{y;p}(H,K)$. Then there exists a non-constant solution of

$$v\colon \mathbb{R}\times S^1 \to M, \quad v_s + J(v)\big(v_t - X_K(t,v)\big) = 0$$

with $v(+\infty) = p$ and $\lim_{s\to-\infty} v(s,\cdot) = z \in \mathcal{P}_1(K)$. By assumption of admissibility of K it follows that $z(t) = z(0) = z_o$ for all $t \in S^1$ and we can compute the flow energy of v as

$$\begin{aligned} E(v) = \iint |v_s|^2\, dsdt &= \mathcal{A}_K(p) - \mathcal{A}_K(z) \\ &= -\int_{S^1} K(t,p)dt + \int_{S^1} K(t,z_o)dt. \end{aligned}$$

Since by assumption $K(t,x) \le K(t,p)$ for all $t \in S^1$, $x \in M$, it follows that $E(v) = 0$, that is, v has to be constant. Altogether it follows that the 0-dimensional manifold $\widetilde{\mathcal{M}}^-_{y;p}(H,K)$ is compact and $\Phi_{HK} \in CF_{-n}(H)$ is a well-defined Floer chain.

By the standard arguments in Floer theory we can show that

$$\partial\Phi_{HK} = 0,$$

that is $\{\Phi_{HK}\} \in HF_{-n}(H)$ is well-defined. In order to show that $\{\Phi_{HK}\} = \Phi_H([M])$ we use the typical homotopy-cobordism argument in Floer theory. For this we need the assumption that K is homotopic to 0 through admissible Hamiltonians so that we have compactness up to splitting off of Floer trajectories at the negative end. By a standard argument from Floer theory we can then define a chain $S \in CF_{-n+1}$ such that

$$\Phi_{HK} - \Phi_H([M]) = \partial S.$$

Hence we have

$$\{\Phi_{HK}\} = \Phi_H([M]) \in HF_{-n}(H). \tag{40}$$

An energy estimate analogous to (5) shows that

$$\mathcal{A}_H(y) \le \mathcal{A}_K(p) + \|H-K\| \quad \text{if } \widetilde{\mathcal{M}}^-_{y;p}(H,K) \ne \emptyset.$$

Therefore, it follows that

$$\{\Phi_{HK}\} \in \operatorname{im} i^a_* \quad \text{for } a = E_-(K) + \|H-K\|.$$

□

The assumption that K is homotopic to 0 through admissible Hamiltonians can be replaced by a simpler condition in view of the following. If we assume that the possibly non-autonomous Hamiltonian K has no non-constant contractible T-periodic solution for $0 < T \le 1$, the homotopy $(\tau K)_{\tau\in[0,1]}$ is a homotopy through Hamiltonians which are admissible apart from the fact that they are not anymore 1-periodic in t. But in view of the nonlinear Fredholm analysis for the quasi-linear Cauchy-Riemann type operator of the type $\bar{\partial}\colon W^{1,p} \to L^p$-maps used in the cobordism argument for (40) this non-continuity of H at $t = 1$ is not essential. However, this point is not carried out in further detail since we will apply Proposition 5.12 to autonomous Hamiltonians in view of the Hofer-Zehnder capacity.

From Proposition 5.12 we can now conclude the:

Proof of Theorem 5.11. Note that $K^{-1}(t,x) = -K(-t,x)$ satisfies the same condition as K. For any regular Hamiltonian H we obtain from Proposition 5.12 applied to $H^{\pm 1}$ and $K^{\pm 1}$,

$$\gamma(H) \geq \|K\| - 2\|H - K\|.$$

From the denseness of $\mathcal{H}_{\text{reg}}$ and the continuity of γ we thus obtain

$$\gamma(K) \geq \|K\|$$

which implies the assertion in view of the obvious estimate $\gamma(K) \leq \|K\|$. □

In principle, such admissible Hamiltonian functions are the key ingredient of the definition of the Hofer-Zehnder capacity for symplectic manifolds. However, here we consider an alteration by allowing admissible Hamiltonians to exhibit non-contractible non-constant 1-periodic solutions.

Definition 5.13. Let (U, ω) be a symplectic manifold. Consider the function space

$$\mathcal{H}_c(U) = \{\, H \in C_o^\infty(\operatorname{int} U) \mid H \geq 0,\ H_{|V} = \sup H \text{ for some open subset } V \,\}.$$

We call H **HZ-admissible** if the corresponding Hamiltonian flow has no contractible non-constant T-periodic solution with period $T \leq 1$. Let

$$\mathcal{H}^o_{\mathsf{HZ}}(U, \omega) = \{\, H \in \mathcal{H}_c(U) \mid H \text{ is HZ-admissible} \,\}.$$

Then, the **π_1-sensitive Hofer-Zehnder capacity** of (U, ω) is defined as

$$c^o_{\text{HZ}}(U, \omega) = \sup_{H \in \mathcal{H}^o_{\mathsf{HZ}}(U,\omega)} \|H\|.$$

The definition of the original Hofer-Zehnder capacity $c_{\text{HZ}}(U, \omega)$ also excludes the existence of non-contractible slow non-constant periodic orbits. That is, we have

$$c^o_{\text{HZ}}(U, \omega) \geq c_{\text{HZ}}(U, \omega).$$

We now are able to relate the metric γ with this π_1-sensitive Hofer-Zehnder capacity.

Corollary 5.14. *Let U be an open subset of (M, ω). Then*

$$c^o_{HZ}(U) = \sup\{\gamma(\phi) \mid \phi = \phi^1_H,\ H \text{ HZ-admissible}\}.$$

Proof. This is a direct consequence from the estimate of Theorem 5.11 because every HZ-admissible Hamiltonian is quasi-autonomous, satisfies (36) and is homotopic to 0 through admissible Hamiltonians. □

This comparison result suggests the following definition of a relative capacity based on the Floer-homological approach via γ.

Definition 5.15. Given any subset $A \subset M$ we can define the following relative capacity $c_\gamma(A) \in [0,\infty) \cup \{\infty\}$,

$$c_\gamma(A) = \sup\{\, \gamma(\phi) \mid \phi = \phi_H^1 \in \mathrm{Ham}(M,\omega),\ \mathrm{supp}\, X_H(t,\cdot) \subset A \,\text{for all}\ t\}.$$

According to its definition, c_γ is a priori only a relative capacity, i.e., invariant under global symplectic automorphisms of (M,ω). Following [**8**] and [**11**] we have the displacement energy

$$e(A) = \inf\{d_H(\phi, \mathrm{id}) \mid \phi \in \mathrm{Ham}(M,\omega),\ \phi(A) \cap A = \emptyset\}.$$

From Corollary 5.14 and Proposition 5.1 we have the inequality:

Corollary 5.16. *For any open subset $U \subset M$ we have the inequality of the capacities*

$$c_{\mathrm{HZ}}(U) \le c^o_{\mathrm{HZ}}(U) \le c_\gamma(U) \le 2e(U).$$

It can, in general, not be expected that $c_\gamma(U)$ and $c^o_{\mathrm{HZ}}(U)$ coincide because the Hofer-Zehnder capacity is based on the exclusion of *any* fast non-constant contractible periodic solution, whereas the γ-capacity only considers periodic solutions which represent the top and bottom cohomology class. However, there might be nontrivial fast solutions representing nontrivial intermediate classes, for example associated to the levels $c([\omega]^k)$ with $0 < k < n$.

As example, consider a symplectic embedding of the standard ball $(B^{2n}(r), \omega_o)$ of radius r into M such that the image lies within a set A. This embedding provides a push-forward of HZ-admissible Hamiltonians on $B^{2n}(r)$ to M and we obtain

$$\pi r^2 = c_{\mathrm{HZ}}(B^{2n}(r)) \le c_\gamma(A) \le 2e(A).$$

This reproduces the result from Theorem 1.1 in [**10**] in our Floer homological setup for symplectically aspherical manifolds.

Example 5.17. In the case of the torus, we have seen that

$$c_\gamma(S^1 \times S^1) = \infty.$$

But by passing to a suitably large finite covering one can show that

$$c_\gamma(S^1 \times S^1 \setminus (\{\mathrm{pt}\} \times S^1 \cup S^1 \times \{\mathrm{pt}\})) < \infty.$$

In fact, since the universal covering is $\mathbb{R}^2$, we have

$$c^o_{\mathrm{HZ}}(S^1 \times S^1 \setminus (\{\mathrm{pt}\} \times S^1 \cup S^1 \times \{\mathrm{pt}\})) = 1.$$

To make the covering argument more precise we conclude with the following observations.

Let $\pi\colon \hat{M} \to M$ be a finite, symplectic covering of degree m. Given a Hamiltonian $H\colon S^1 \times M \to \mathbb{R}$ we obtain the pull-back $\pi^* H\colon S^1 \times \hat{M} \to \mathbb{R}$

and define the notation $\pi^*\phi_H^1 = \phi_{\pi^*H}^1$. We observe that for such a finite covering we obtain

$$\gamma(\pi^*\phi) = \gamma(\phi) \quad \text{for all } \phi \in \mathrm{Ham}(M,\omega). \tag{41}$$

Namely, since π is in particular a local symplectomorphism, considering the pull-back operation on Floer homology for this finite covering we deduce that

$$c(\pi^*\alpha, \pi^*H) = c(\alpha, H) \quad \text{for all } \alpha \in H^*(M), \text{ s.t. } \pi^*\alpha \neq 0.$$

The identity (41) then follows from property 1). in Theorem 1.2.

Now suppose we have an open subset $U \subset M$ such that its preimage $\pi^{-1}(U) = U_1 \cup \ldots \cup U_m$ is a disjoint union of m copies of a lift of U, and suppose that H satisfies the condition $\operatorname{supp} X_H(t,\cdot) \subset U$ for all t. We may assume that $\operatorname{supp} H(t,\cdot) \subset U$ for all t. Then define $\phi_i \in \mathrm{Ham}(\hat{M},\omega)$ by $\phi_i = \phi^1_{\chi_i \pi^* H}$ where χ_i is the characteristic function of U_i. It follows for $\phi = \phi_H^1$ that $\pi^*\phi = \phi_1 \circ \ldots \circ \phi_m$, hence by (41)

$$\gamma(\phi) \leq \sum_{i=1}^{m} \gamma(\phi_i) = m\gamma(\phi_i). \tag{42}$$

In the case that for (M, U) there exists a finite symplectic covering $\pi \colon \hat{M} \to M$ and $\psi \in \mathrm{Ham}(\hat{M}, \pi^*\omega)$ such that $\psi(U_i) \cap U_i = \emptyset$ we obtain

$$c_\gamma(U) \leq 2\gamma(\psi) < \infty$$

thus proving the finiteness assertion in Example 5.17.

Remark 5.18. The estimate (42) should not be sharp. One might expect $\gamma(\pi^*\phi) = \gamma(\phi_i)$. This will be studied more closely elsewhere.

References

[1] M. Bialy and L. Polterovich, *Geodesics of Hofer's metric on the group of Hamiltonian diffeomorphisms*, Duke Math. Journal, **76**(**1**) (1994), 273-292.

[2] A. Floer, *Symplectic fixed points and holomorphic spheres*, Comm. Math. Phys., **120** (1989), 575-611.

[3] A. Floer and H. Hofer, *Coherent orientations for periodic orbit problems in symplectic geometry*, Math. Zeit., **212** (1993), 13-38.

[4] ———, *Symplectic homology* I: *Open sets in* $\mathbb{C}^n$, Math. Zeit., **215** (1994), 37-88.

[5] A. Floer, H. Hofer, and K. Wysocki, *Applications of symplectic homology* I, Math. Zeit., **217** (1994), 577-606.

[6] R. Gompf, *On symplectically aspherical manifolds with nontrivial* π_2, Math. Res. Lett., **5** (1998), 599-603.

[7] H. Hofer and E. Zehnder, *A new capacity for symplectic manifolds*, Analysis et cetera (P. Rabinowitz and E. Zehnder, eds.), Academic Press, (1990), 405-428.

[8] ______, *Symplectic invariants and Hamiltonian dynamics*, Birkhäuser advanced texts, Birkhäuser, 1994.

[9] J. Kollar, *private communication*.

[10] F. Lalonde and D. McDuff, *The geometry of symplectic energy*, Ann. of Math., **141** (1995), 349-371.

[11] ______, *Hofer's L^∞-geometry: Energy and stability of Hamiltonian flows*, I, II, Invent. Math., **122**(**1**) (1995), 1-33, 35-69.

[12] F. Lalonde and L. Polterovich, *Symplectic diffeomorphisms as isometries of Hofer's norm*, Topology, **36**(**3**) (1997), 711-727.

[13] H.V. Le and K. Ono, *Cup-length estimate for symplectic fixed points*, Contact and Symplectic Geometry (Cambridge, 1994) (C.B. Thomas, ed.), Publ. Newton Inst., Vol. 8, Cambridge University Press, 1996.

[14] Y.-G. Oh, *Symplectic topology as geometry of action functional*, I, J. Diff. Geom., **46** (1997), 499-577.

[15] ______, *Symplectic topology as geometry of action functional*, II, Comm. Anal. Geom., **7** (1999), 1-54.

[16] S. Piunikhin, D. Salamon, and M. Schwarz, *Symplectic Floer-Donaldson theory and quantum cohomology*, Contact and Symplectic Geometry (Cambridge 1994) (C.B. Thomas, ed.), Publ. Newton Inst., Vol. 8, Cambridge University Press, (1996), 171-200.

[17] D. Salamon and E. Zehnder, *Morse theory for periodic solutions of Hamiltonian systems and the Maslov index*, Comm. Pure Appl. Math., **45** (1992), 1303-1360.

[18] M. Schwarz, *An explicit isomorphism between Floer homology and quantum cohomology*, in preparation.

[19] ______, *Morse homology*, Progress in Mathematics, Vol. 111, Birkhäuser, Basel, 1993.

[20] ______, *Cohomology operations from S^1-cobordisms in Floer homology*, Ph.D. thesis, Swiss Federal Inst. of Techn. Zurich, Diss. ETH No. 11182, 1995.

[21] ______, *A quantum cup-length estimate for symplectic fixed points*, Invent. Math., **133** (1998), 353-397.

[22] P. Seidel, *π_1 of symplectic automorphism groups and invertibles in quantum homology rings*, Geom. Funct. Anal., **7** (1997), 1046-1095.

[23] C. Viterbo, *Symplectic topology as the geometry of generating functions*, Math. Ann., **292**(**4**) (1992), 685–710.

Received September 29, 1998 and revised October 29, 1998. This work was partially supported by NSF grant # DMS 9626430.

Max-Planck-Institute for Mathematics in the Science
D-04102 Leipzig
Germany
E-mail address: mschwarz@mis.mpg.de

PACIFIC JOURNAL OF MATHEMATICS
Vol. 193, No. 2, 2000

HILBERT'S TENTH PROBLEM FOR ALGEBRAIC FUNCTION FIELDS OVER INFINITE FIELDS OF CONSTANTS OF POSITIVE CHARACTERISTIC

Alexandra Shlapentokh

Let K be an algebraic function field of characteristic $p > 2$. Let C be the algebraic closure of a finite field in K. Assume that C has an extension of degree p. Assume also that K contains a subfield K_1, possibly equal to C, and elements u, x such that u is transcendental over K_1, x is algebraic over $C(u)$ and $K = K_1(u, x)$. Then the Diophantine problem of K is undecidable.

Let G be an algebraic function field in one variable whose constant field is algebraic over a finite field and is not algebraically closed. Then for any prime $\mathfrak{p}$ of G, the set of elements of G integral at $\mathfrak{p}$ is Diophantine over G.

1. Introduction.

The interest in the questions of Diophantine definability and decidability goes back to a question which was posed by Hilbert: Given an arbitrary polynomial equation in several variables over $\mathbb{Z}$, is there a uniform algorithm to determine whether such an equation has solutions in $\mathbb{Z}$. This question, otherwise known as Hilbert's 10th problem, has been answered negatively in the work of M. Davis, H. Putnam, J. Robinson and Yu. Matijasevich. (See [**5**] and [**6**].) Since the time when this result was obtained, similar questions have been raised for other fields and rings. Arguably the two most interesting and difficult problems in the area are the questions of Diophantine decidability of $\mathbb{Q}$ and the rings of algebraic integers of arbitrary number fields. One way to resolve the question of Diophantine decidability over a ring of characteristic 0 is to construct a Diophantine definition of $\mathbb{Z}$ over such a ring. This notion is defined below.

Definition 1.1. Let R be a ring and let $A \subset R$. Then we say that A has a Diophantine definition over R if there exists a polynomial $f(t, x_1, \ldots, x_n) \in R[t, x_1, \ldots, x_n]$ such that for any $t \in R$,

$$\exists x_1, \ldots, x_n \in R, f(t, x_1, \ldots, x_n) = 0 \Longleftrightarrow t \in A.$$

If the quotient field of R is not algebraically closed, it can be shown that we can allow Diophantine definitions to consist of several polynomials without changing the nature of the relationship. (For more details see [**6**].) Such Diophantine definitions have been obtained for $\mathbb{Z}$ over rings of algebraic integers of the following fields: Totally real extensions of $\mathbb{Q}$, their extensions of degree 2, fields with exactly one pair of complex conjugate embeddings, some fields of degree 4, and some totally real infinite extensions of $\mathbb{Q}$. For more details concerning these results see [**7**], [**11**], [**12**], [**25**], [**30**], [**29**], [**37**]. However, not much progress has been made towards resolving the Diophantine problem of $\mathbb{Q}$. Further, one of the consequences of a series of conjectures by Barry Mazur and Colliot-Thélène, Swinnerton-Dyer and Skorobogatov is that $\mathbb{Z}$ does not have a Diophantine definition over $\mathbb{Q}$, and thus one would have to look for some other method for resolving the Diophantine problem of $\mathbb{Q}$. (Mazur's conjectures can be found in [**23**] and [**24**]. However, Colliot-Thélène, Swinnerton-Dyer and Skorobogatov have found a counterexample to the strongest of the conjectures in the papers cited above. Their modification of Mazur's conjecture in view of the counterexample can be found in [**4**].) In [**40**], the author of this paper has demonstrated that in certain totally real algebraic number fields there exist recursive integrally closed rings of algebraic numbers where infinite number of primes can appear in denominators and where rational integers have Diophantine definition. (This implies, of course, that Hilbert's Tenth Problem is undecidable over these rings.) The result above was not proved for $\mathbb{Q}$. The general problem of existence of Diophantine definitions of rational and algebraic integers over integrally closed subrings of number fields (including the fields themselves) remains open.

The problem turned out to be much more tractable over function fields. At this point there are several results pertaining to Diophantine undecidability of various function fields and rings. More specifically, we know that the Diophantine problem of the following function fields is undecidable: the rational function fields of characteristic 0 whose constant fields are subfields of some p-adic fields or are formally real (see [**9**] and [**19**]), the rational function fields in two variables over $\mathbb{C}$ (see [**18**]), the rational function fields over the finite fields of constants (see [**25**], [**41**]), rational function fields of positive characteristic whose constant fields do not contain the algebraic closure of a finite field ([**17**]), and algebraic function fields over finite fields of constants ([**38**]). Results concerning various function rings can be found in [**27**], [**31**], [**32**], [**34**], [**35**], [**36**].

In this paper we extend the undecidability results of Pheidas, Kim and Roush, and the author of this paper to a new class of fields of algebraic functions: Algebraic function fields of positive characteristic p such that the algebraic closure of a finite field contained in the fields under consideration has an extension of degree p. More specifically, we will prove the following theorems.

Theorem. *Let K be an algebraic function field of characteristic $p > 2$. Let C be the algebraic closure of a finite field in K. Assume that C has an extension of degree p. Assume also that K contains a subfield K_1, possibly equal to C, and elements u, x such that u is transcendental over K_1, x is algebraic over $C(u)$ and $K = K_1(u, x)$. Then the Diophantine problem of K is undecidable.*

Theorem. *Let G be an algebraic function field whose constant field C is algebraic over a finite field of characteristic $p > 0$. Assume further, that C is not algebraically closed. Then for any prime $\mathfrak{B}$ of G the set of all elements of G integral with respect to $\mathfrak{B}$ is Diophantine over G.*

The proof of the undecidability result is based on the idea first introduced by Denef in [**10**] and further developed by Pheidas in [**26**], Kim and Roush in [**17**], and the author of this paper in [**38**]. This idea can be summarized by the following lemma.

Lemma 1.2. *Let K be an algebraic function field of characteristic $p > 0$. Let $t \in K$ be a nonconstant element of K. Let C_p be the finite field of p elements, and let G be the algebraic closure of $C_p(t)$ in K. Let $\mathfrak{p}$ be a prime of K which lies above a nontrivial prime of G, and assume that the following sets are Diophantine over K.*

$$p(K) = \{(x, w) \in K^2 | \exists s \in \mathbb{N}, \, w = x^{p^s}\}$$

$$INT(\mathfrak{p}),$$

where if $w \in K \cap INT(\mathfrak{p})$ then $\mathrm{ord}_{\mathfrak{p}} w \geq 0$, and if $w \in G$ and $\mathrm{ord}_{\mathfrak{p}} w \geq 0$, then $w \in INT(\mathfrak{p})$. Then the Diophantine problem of K is undecidable.

(The proof of the lemma can be easily derived from the proof of [**38**, Lemma 1.5].)

Section two of the paper is devoted to showing that $p(K)$ is Diophantine over K, while section three contains a proof of the fact that $INT(\mathfrak{p})$ is Diophantine over K. Before we leave this section, we will state one more easy but useful lemma concerning Diophantine definitions.

Lemma 1.3. *Let L be a field, and let*

$$P(w, u_1, \dots, u_m) = 0 \tag{1.1}$$

be a polynomial equation over L. Let

$$\{P_s(w, x_1, \dots, x_n, y_1, \dots, y_r) = 0, s = 1, \dots, v\} \tag{1.2}$$

be a set of equations over L. Then, assuming $k > 0$ is the degree of P in w, there exists a set of equations

$$\{Q_l(u_1, \dots, u_m, t_{1,0}, \dots, t_{n,k-1}, y_1, \dots, y_r) = 0,\ l = 1, \dots, e\} \tag{1.3}$$

over L such that for any $u_1, \dots, u_m, y_1, \dots, y_r \in L$, the system (1.1) *and* (1.2) *has solutions w in the algebraic closure of L and $x_1, \dots, x_n \in L(w)$ if and only if for some $t_{1,0} \dots, t_{n,k-1} \in L$, $t_{1,0}, \dots, t_{n,k-1}$, $u_1, \dots, u_m$, $y_1, \dots, y_r$ are solutions of the system* (1.3).

Proof. Fix $u_1, \dots, u_m, y_1, \dots, y_r \in L$ and assume initially that the values of $u_1, \dots, u_m$ under consideration will not make the leading coefficient of P with respect to w zero. Under this assumption we can use Equation (1.1) to compute $\{A_{i,j}(u_1, \dots, u_m) \in L(u_1, \dots, u_m)\}$ such that for any $i \in \mathbb{N}$,

$$w^i = \sum_{j=0}^{k-1} A_{i,j} w^j. \tag{1.4}$$

Next consider the following system of equations:

$$\left\{P_s\left(w, \sum_{i=0}^{k-1} t_{i,1}w^i, \dots, \sum_{i=0}^{k-1} t_{i,n}w^i, y_1, \dots, y_r\right) = 0, s = 1, \dots, v\right\}. \tag{1.5}$$

If we treat $\{1, \dots, w^{k-1}\}$ as if they were linearly independent over $L(u_1, \dots, u_m, y_1, \dots, y_r)$ and use Equation (1.4), we can replace the system (1.5) by a system of the form (1.3), where every P_i is replaced by k equations corresponding to the coefficients of the first k powers of w. Suppose now (1.2) has solutions as described in the statement of the lemma. Since $0 < [L(w) : L] \leq k$, for $i = 1, \dots, n, x_i = \sum_{i=0}^{k-1} a_{i,j}w^j$, where $a_i \in L$. (If $[L(w) : L] < k$ then for $j = [L(w) : L], \dots, k$, we can set $a_{i,j} = 0$.) Thus, the system (1.3) will clearly be satisfied with $t_{i,j} = a_{i,j}$. Conversely, if for some $a_{i,j}, = Q_l(u_1, \dots, u_m, a_{1,0}, \dots, a_{n,k-1}, y_1, \dots, y_r) = 0, l = 1, \dots, e$, then given the construction of Q_l's and assuming w is a root of P,

$$P_s\left(w, \sum_{i=0}^{k-1} a_{i,1}w^i, \dots, \sum_{i=0}^{k-1} a_{i,n}w^i, y_1, \dots, y_r\right) = 0, s = 1, \dots, v.$$

Finally, we remove the assumption that the leading coefficient of P with respect to w is not zero. To accomplish that we need to consider the following cases: The k-th coefficient is not zero; the k-th coefficient is zero but $k-1$-st coefficient is not zero; ... ; only the free term is nonzero. Conditions for each case can be written down in a Diophantine fashion and all the conditions can be combined together in a Diophantine fashion also.

For the remainder of the paper we will use the following notations.

Notations 1.4.

- K will denote an algebraic function field over a field of constants C_K of characteristic $p > 2$. In other words, K is a finite algebraic extension of $C_K(w)$ for some $w \in K$ transcendental over C_K.
- C will denote the algebraic closure of a finite field in C_K.
- u will denote a nonconstant element of K.
- G will denote the algebraic closure of $C(u)$ in K.
- Given $x_1, \ldots, x_m \in G$, $G_{x_1,\ldots,x_m}$ will denote a subfield of G containing $x_1, \ldots, x_m$ and such that $C_{x_1,\ldots,x_m}$ - the constant field of $G_{x_1,\ldots,x_m}$ is finite.
- t will denote an element of $G \setminus C$ such that the divisor of t is of the form $\mathfrak{p}/\mathfrak{q}$, where $\mathfrak{p}$, $\mathfrak{q}$ are K primes of degree q^h for some rational prime number q and a natural number h. Further, $K/C_K(t)$ is separable and $q^h = n = [K : C_K(t)]$.
- $\tilde{C}_K = C_{\tilde{K}}$, where $\tilde{K} = \tilde{C}_K K$, will denote the algebraic closure of C_K.
- r will denote the number of primes ramifying in the extension $\tilde{C}_K K/\tilde{C}_K(t)$.
- $|C| \geq N(n+2r+5)$, where $N(n+2r+5)$ is a positive constant defined in the proof of Theorem 6.11, or C is infinite.
- C has an extension of degree q, where q is a rational prime possibly equal to p.
- x will denote a generator of G over $C(t)$. (Such a generator exists by Lemma 6.18 and our assumption that $K/C_K(t)$ is separable.)
- $c_0 = 0, c_1 \neq \pm 1, \ldots c_{n+2r+5} \neq \pm 1$ will denote the elements of C such that for $i = 0, 1, \ldots, n + 2r + 5$, the divisor of $t - c_i$ is of the form $\mathfrak{p}_i/\mathfrak{q}$, where $\mathfrak{p}_i$ is a prime divisor. For $i \neq j$, for any natural number $s, c_i^{p^s} \neq c_j$.
- For all i, $\mathfrak{P}_i$ will denote the prime of $C_K(t)$ lying below $\mathfrak{p}_i$, while $\mathfrak{P}$ and $\mathfrak{Q}$ will denote $C_K(t)$-primes below $\mathfrak{p}$ and $\mathfrak{q}$ respectively. For all i, $\mathfrak{P}_i$, $\mathfrak{P}$ and $\mathfrak{Q}$ do not split in the extension $K/C_K(t)$.
- r_i will denote the smallest positive integer such that $c_i^{p^{r_i}} = c_i$. We will let $d_{ij} = c_i^{p^j}, 0 \leq j \leq r_i$.

(The existence of an algebraic function field K over a sufficiently large or infinite field of constants containing t and $c_1, \ldots, c_{2r+n+5}$ satisfying the conditions above follows from Theorem 6.11. In Section 5 of the paper we will give a fuller description of the class of fields satisfying our assumptions.)

2. P-th Power Equations: The case of $p = q$.

In this section we will show that over an algebraic function field K of characteristic $p > 0$, under some assumptions on the constant field, the set $p(K)$ is Diophantine. The method we are going to use has its origins in a paper of Pheidas (see [**26**]). It was extended by the author to prove an analogous result for an algebraic function field over a *finite* field of constants. Unfortunately, in its original form, this extension cannot be used to prove the results pertaining to the algebraic function fields over infinite fields of constants, since it relied on the fact that the class numbers of global fields are finite. To prove the results mentioned above in our case, we will use the fact that under our assumptions on the constant field, the algebraic function field K will have a rational subextension of degree p^h.

Lemma 2.1. *Let F/G be a finite separable extension of fields of positive characteristic p. Let $\alpha \in F$ be such that all the coefficients of its monic irreducible polynomial over G are p-th powers in G. Then α is a p-th power in F.*

Proof. Let $a_0^p + \ldots + a_{m-1}^p T^{m-1} + T^m$ be the monic irreducible polynomial of α over G. Let β be the element of the algebraic closure of F such that $\beta^p = \alpha$. Then β is of degree at most m over G. On the other hand, $G(\alpha) \subseteq G(\beta)$. Therefore, $G(\alpha) = G(\beta)$.

Lemma 2.2. *Let F/G be a finite separable extension of fields of positive characteristic p. Let $[F : G] = n$. Let $x \in F$ be such that $F = G(x)$ and for distinct $a_0^p, \ldots, a_n^p \in G$, $\mathbf{N}_{F/G}(a_i^p - x) = y_i^p$. Then x is a p-th power in F.*

Proof. Let $H(T) = A_0 + A_1 T + A_{n-1}T^{n-1} + T^n$ be the monic irreducible polynomial of x over G. Then for $i = 0, \ldots, n$, $H(a_i^p) = y_i^p$. Further, we have the following linear system of equations:

$$\begin{pmatrix} 1 & a_0^p & \ldots & a_0^{p(n-1)} & a_0^{pn} \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ 1 & a_n^p & \ldots & a_n^{p(n-1)} & a_n^{pn} \end{pmatrix} \begin{pmatrix} A_0 \\ \ldots \\ 1 \end{pmatrix} = \begin{pmatrix} y_0^p \\ \ldots \\ y_n^p \end{pmatrix}.$$

Using Cramer's rule to solve the system, it is not hard to conclude that for $i = 0, \ldots, n$, A_i is a p-th power in G. Then, by Lemma 2.1, x is a p-th power in F.

Lemma 2.3. *Let $w \in K$, let $\mathfrak{a}_1, \ldots, \mathfrak{a}_r$ be primes of K and let $a_1, \ldots, a_{r+1} \in C$ be a set of distinct constants. Then the set $\{w + a_1, \ldots, w + a_{r+1}\}$ contains at least one element of K having no zero at any of the primes $\mathfrak{a}_1, \ldots, \mathfrak{a}_r$.*

Proof. The lemma follows from the fact that each prime $\mathfrak{a}_i$ can be a zero of at most one element of the set $\{w + a_1, \ldots, w + a_{r+1}\}$.

Lemma 2.4. *Let $w \in K$, let $a, b \in C$. Then all the zeros of $\frac{w+a}{w+b}$ are zeros of $w + a$ and all the poles of $\frac{w+a}{w+b}$ are zeros of $w + b$. Further, the height of $\frac{w+a}{w+b}$ is equal to the height of w. (Here by height we mean the degree of zero or pole divisor of an algebraic function.)*

Proof. Let $\mathfrak{p}$ be a prime of K. Then $\mathfrak{p}$ is a pole w if and only if $\mathfrak{p}$ is a pole of $w + a$ and a pole of $w + b$. Moreover, the order of the pole at all the three functions will be the same. On the other hand, any zero of $\frac{w+a}{w+b}$ will come from zeros of $w + a$ or poles of $w + b$. So let $\mathfrak{p}$ be a pole of $w + b$. Then $\text{ord}_{\mathfrak{p}}(w + a) = \text{ord}_{\mathfrak{p}}(w + b)$ and therefore $\text{ord}_{\mathfrak{p}} \frac{w+a}{w+b} = 0$. A similar argument shows that $\frac{w+a}{w+b}$ is a unit at any valuation which is a pole of $w+a$. Consequently, all zeros of $\frac{w+a}{w+b}$ are zeros of $w + a$ and all the poles of $\frac{w+a}{w+b}$ are zeros of $w + b$.

Finally, note that $\frac{w+a}{w+b} = 1 + \frac{a-b}{w+b}$. Let $H_K(\frac{w+a}{w+b})$ denote the K-height of $\frac{w+a}{w+b}$. Then we have the following equalities.

$$H_K\left(\frac{w+a}{w+b}\right) = H_K\left(1 + \frac{a-b}{w+b}\right) = H_K\left(\frac{a-b}{w+b}\right) = H_K(w+b) = H_K(w).$$

The last equality follows from the fact, mentioned above, that the pole divisors of $w + b$ and w are the same.

Lemma 2.5. *Let $u, v, z \in \tilde{C}_K K = \tilde{K}$, let $y \in \tilde{C}_K(z)$, and assume y, z do not have zeros or poles at any valuation of $\tilde{K}$ ramifying in the extension $\tilde{K}/\tilde{C}_K(z)$. Further, assume*

$$y - z = u^p - u \tag{2.1}$$

$$y^{-1} - z^{-1} = v^p - v. \tag{2.2}$$

Then $y = z^{p^s}$, for some natural number $s \geq 0$. (Note that in $\tilde{C}_K(z)$, the zeros and the poles of z are simple. Assuming that z has no zeros or poles at any valuations ramifying in the extension $\tilde{K}/\tilde{C}_K(z)$ amounts, therefore, to assuming that all zeros and poles of z are simple in $\tilde{K}$.)

Proof. The argument below is very similar to the one used in [**26**, Lemma 1, pages 3-4], with the following difference. In this lemma we do not assume that u, v are rational functions in z over $\tilde{C}_K$ and therefore we will have to use the concept of local derivation with respect to a prime in place of the derivative defined in the usual manner on a rational function field. (For a discussion of local and global derivations see [**22**, pages 9-10] and [**13**, pages 144-148].) Let $\mathfrak{A}/\mathfrak{B}$ be the divisor of $z \in \tilde{K}$, where $\mathfrak{A}$ and $\mathfrak{B}$ are relatively prime integral divisors. Further, by assumption all the prime factors of $\mathfrak{A}$ and $\mathfrak{B}$ are distinct. Next note that all the poles of $v^p - v$ and $u^p - u$ in $\tilde{K}$ are of orders divisible by p. Since from the above discussion we know that all the zeros and poles of z are of orders equal to ± 1, we must conclude from (2.1) and (2.2) that the divisor of y is of the form $\mathfrak{U}^p\mathfrak{V}$, where all the prime factors of $\mathfrak{V}$ come from $\mathfrak{A}$ or $\mathfrak{B}$ and are distinct. Further, the factors of $\mathfrak{A}$ will appear to the first power in $\mathfrak{V}$; and the factors of $\mathfrak{B}$ will appear to the power -1 in $\mathfrak{V}$. Indeed, let $\mathfrak{t}$ be a prime which is not a factor of $\mathfrak{A}$ or $\mathfrak{B}$. Without loss of generality assume $\mathfrak{t}$ is a pole of y. Then, since $\mathrm{ord}_{\mathfrak{t}} z = 0$,

$$0 > \mathrm{ord}_{\mathfrak{t}} y = \mathrm{ord}_{\mathfrak{t}}(z - y) = \mathrm{ord}_{\mathfrak{t}}(u^p - u) \cong 0 \text{ modulo } p.$$

Now let $\mathfrak{t}$ be a factor of $\mathfrak{A}$ or $\mathfrak{B}$. Again, without loss of generality, assume that $\mathfrak{t}$ is a pole of y. If $\mathfrak{t}$ is a factor of $\mathfrak{A}$, then $\mathrm{ord}_{\mathfrak{t}}(y - z) = \mathrm{ord}_{\mathfrak{t}} y = \mathrm{ord}_{\mathfrak{t}}(u^p - u)$. Since we assumed $\mathfrak{t}$ to be a pole of y, we must conclude that $\mathfrak{t}$ is a pole of u and thus $\mathrm{ord}_{\mathfrak{t}} y = \mathrm{ord}_{\mathfrak{t}}(u^p - u) \cong 0$ modulo p. If, on the other hand, $\mathfrak{t}$ is a factor of $\mathfrak{B}$. Then we have two possibilities: $\mathrm{ord}_{\mathfrak{t}} y = \mathrm{ord}_{\mathfrak{t}} z = -1$ or again $\mathrm{ord}_{\mathfrak{t}} y = \mathrm{ord}_{\mathfrak{t}}(u^p - u) \cong 0$ modulo p.

On the other hand, since $y \in \tilde{C}_K(z)$, where $\mathfrak{A}$ and $\mathfrak{B}$ are prime divisors, we must conclude that the divisor of y is actually of the form $\mathfrak{U}^p \mathfrak{A}^a \mathfrak{B}^b$, with either $a, b = 0$ or $a = 1, b = -1$. (This follows from the observation that the degree of the zero and the pole divisor of y must be the same. In particular, the degrees must be equal modulo p.) If $a, b = 0$, taking into account the fact that no prime which is a pole or zero of y ramifies in the extension $\tilde{K}/\tilde{C}_K(z)$, we can conclude that the divisor of y in the rational field is also a p-th power of another divisor. Thus, since in the rational field every zero degree divisor is principal, y is a p-th power. Suppose, on the other hand that $a = 1, b = -1$. Then we can conclude using an argument similar to the one above, that yz^{-1} is a p-th power in the rational field. Thus, (2.1) can be rewritten as

$$z(f - 1)^p = u^p - u, \tag{2.3}$$

where $f \in \tilde{C}_K(z)$. Since $f - 1$ is a rational function in z, we can further rewrite (2.3) as

$$z(f_1^p / f_2^p) = u^p - u, \tag{2.4}$$

where f_1, f_2 are relatively prime polynomials in z over $\tilde{C}$ and f_2 is monic. From this equation it is clear that any valuation which is a pole of u, is either a pole of z or a zero of f_2. Further, the absolute value of the order of any pole of u at any valuation which is a zero of f_2, must be the same as the order of f_2 at this valuation. Therefore, $s = f_2 u$ will have poles only at the valuations which are poles of z. Thus we can rewrite (2.4) in the form

$$-zf_1^p + s^p = sf_2^{p-1}.$$

Let $\mathfrak{c}$ be a zero of f_2. Then, since f_2 is a polynomial in z, $\mathfrak{c}$ is not a pole of z. Since, $p - 1 \geq 2$, s is integral over $\tilde{C}_K[z], \mathrm{ord}_{\mathfrak{c}}(s^p - zf_1^p) \geq 2$.

In general, for any $x \in \tilde{K}$ and any $\tilde{K}$-prime $\mathfrak{a}$, let $\partial x / \partial \mathfrak{a}$ denote the local derivative of x with respect to $\mathfrak{a}$. Further, if x has a zero at $\mathfrak{a}$ of order greater than 1, then $\partial x / \partial \mathfrak{a}$ will have a zero at $\mathfrak{a}$. Now observe that

$$\mathrm{ord}_{\mathfrak{c}} \partial(-zf_1^p + s^p)/\partial \mathfrak{c} = \mathrm{ord}_{\mathfrak{c}} \frac{d(-zf_1^p + s^p)}{dz} = \mathrm{ord}_{\mathfrak{c}}(-f_1^p),$$

by Lemma 6.17, since, by assumption f_2 does not have any zeros at valuations ramifying in the extension $\tilde{K}/\tilde{C}_K(t)$. Thus, f_1 has a zero at $\mathfrak{c}$. But f_1 and f_2 are supposed to be relatively prime polynomials. Hence, f_2 does not have any zeros, and thus is equal to 1. Therefore, y is a polynomial in z. Similarly, we can show that $1/y$ is a polynomial in $1/z$. Hence, y is a power of z, and more specifically, unless $y = z$, y must be a power of z divisible by p. If $y = z$ we are done. Otherwise, we have shown that y is a p-th power of another rational function in z over $\tilde{C}_K$. From this point on, the proof of the lemma proceeds in the fashion identical to the proof which can be found in [**26**, Lemma 1, pages 3, 4].

Lemma 2.6. *Let w, u, u_{i,j_i,k,j_k}, v_{i,j_i,k,j_k}, $i,k = 0,\ldots,(r+n+2)$, $j_i = 1,\ldots,r_i$, $j_k = 1,\ldots,r_k$ be elements of K satisfying the following equations for all $i,k = 0,\ldots,(r+n+2)$, and some $1 \leq j_i \leq r_i$, $1 \leq j_k \leq r_k$.*

$$w - t = u^p - u \tag{2.5}$$

$$w^{-1} - t^{-1} = v^p - v \tag{2.6}$$

$$w_{i,j_i,k,j_k} = \frac{w - d_{i,j_i}}{w - d_{k,j_k}}, \tag{2.7}$$

$$t_{i,k} = \frac{t - c_i}{t - c_k} \tag{2.8}$$

$$w_{i,j_i,k,j_k} - t_{i,k} = u^p_{i,j_i,k,j_k} - u_{i,j_i,k,j_k} \tag{2.9}$$

$$\frac{1}{w_{i,j_i,k,j_k}} - \frac{1}{t_{i,k}} = v^p_{i,j_i,k,j_k} - v_{i,j_i,k,j_k}. \tag{2.10}$$

Then $w = t^{p^s}$ for some natural number $s \geq 0$.

Proof. First of all note that $\mathfrak{Q}$, and $\mathfrak{P}_i$ for all i will remain prime in the extension $\tilde{C}_K(t)/C_K(t)$ and their factors will be unramified in the extension $\tilde{K}/\tilde{C}_K$. Indeed, the first assertion is true because all the listed primes are of degree one in $C_K(t)$ and thus will remain prime under any constant field extension. The second assertion is true by Lemma 6.16. Thus, for all $i, k, t_{i,k}$ has neither zeros nor poles at any prime ramifying in the extension $\tilde{K}/\tilde{C}_K(t)$.

Next we note that by Lemma 2.3, for some $i = 0,\ldots,(n+r+2)$ there exist distinct $k_1,\ldots,k_{n+1} \in \{0,\ldots,(r+n+2)\} \setminus \{i\}$ such that for any $1 \leq j_i \leq r_i, 1 \leq j_{k_l} \leq r_{k_l}, w_{i,j_i,k_l,j_{k_l}}, l = 1,\ldots,n+1$ does not have zeros or poles at any prime ramifying in the extension $\tilde{K}/\tilde{C}_K(t)$. Indeed, we can select the required indices in the following manner. First consider the set $\{w - d_{i,j_i}, i = 0,\ldots,n+r+2, 1 \leq j_i \leq r_i\}$. Note, that by assumption, for all $(i,j_i), d_{i,j_i}$ is a constant and all these constants are distinct. Let $\{m_u, u = 1\ldots,s\}$ be the set of all the elements of the set $\{0,\ldots,n+r+2\}$ such that for some j_{m_u} in the set $\{1,\ldots,r_{m_u}\}$, $w - d_{m_u,j_{m_u}}$ has a zero at a valuation of $\tilde{K}$ ramifying in the extension $\tilde{K}/\tilde{C}_K(t)$. Then by Lemma

2.3, $s \leq r$. Therefore, the set $\{0,\dots,n+r+2\}\setminus\{m_1,\dots,m_s\}$ contains at least $n+3$ elements. Choose i in this set. Finally choose $k_1,\dots,k_{n+1}$ in the set $\{0,\dots,n+r+2\}\setminus\{m_1,\dots,m_s,i\}$, containing at least $n+2$ elements. Next consider, $w_{i,j_i,k_l,j_{k_l}} = \frac{w-d_{i,j_i}}{w-d_{k_l,j_{k_l}}}$, where $1 \leq j_i \leq r_i, k_l \in \{0,\dots,n+r+2\}\setminus\{m_1,\dots,m_s,i\}, 1 \leq j_{k_l} \leq r_{k_l}$. Note that neither numerator, nor denominator of this fraction has a zero at a valuation ramifying in the extension $\tilde{K}/\tilde{C}_K(t)$. Thus, by Lemma 2.4, $w_{i,j_i,k_l,j_{k_l}}$ has no zeros or poles at any valuation ramifying in the extension $\tilde{K}/\tilde{C}_K(t)$.

If $w \in C_K(t)$ then we can apply Lemma 2.5 to conclude that our lemma is true. Thus, we may assume $w \notin C_K(t)$. This would imply that $w_{i,j_i,k_l,j_{k_l}} \notin C_K(t)$ for all i, j_i, k_l, j_{k_l} Further, by an argument similar to the one used in the proof of Lemma 2.5, for all $l = 1,\dots,n+1$, Equations (2.9) and (2.10) imply that for some j_i, j_{k_l} the divisor of $w_{i,j_i,k_l,j_{k_l}}$ is of the form $\mathfrak{A}^p\mathfrak{p}_{k_l}^a\mathfrak{p}_i^b$, where a is either -1 or 0 and b is either 1 or 0. Let $K_w = C_K(w,t)$, and note that for all $i, k, j_i, j_k, w_{i,j_i,k,j_k} \in K_w$ and $[K_w : C_K(t)] = p^m$, where $0 < m \leq h$. (The left inequality is strict due to our assumption that $w \notin C_K(t)$.) Further, since for all $l = 1,\dots,n+1, w_{i,j_i,k_l,j_{k_l}}$ does not have any zeros or poles ramifying in the extension $K/C_K(t)$, the divisor of $w_{i,j_i,k_l,j_{k_l}}$ will be of the form $\mathfrak{A}_{K_w}^p\mathfrak{P}_{i,w}^a\mathfrak{P}_{k_l,w}^b$ in K_w, where $\mathfrak{A}_{K_w}$ is the K_w-divisor below the divisor $\mathfrak{A}$, and for all i, $\mathfrak{P}_{i,w}$ denotes the prime below $\mathfrak{p}_i$ in $C_K(t,w)$. Next we note that for all $l = 1,\dots,n+1$ the divisor of $\mathbf{N}_{K_w/C_K(t)}(w_{i,j_i,k_l,j_{k_l}})$ is equal to the corresponding norm of the divisor of $w_{i,j_i,k_l,j_{k_l}}$. On the other hand,

$$\mathbf{N}_{K_w/C_K(t)}\mathfrak{P}_{i,w} = \mathfrak{P}_i^{f(\mathfrak{P}_{i,w}/\mathfrak{P}_i)} = \mathfrak{P}_i^{p^m}.$$

Thus, for all $l = 1,\dots,n+1$, the divisor of the norm of $w_{i,j_i,k_l,j_{k_l}}$ in $C_K(t)$ is a p-th power of some other divisor of $C_K(t)$. Since in $C_K(t)$ every zero degree divisor is principal, we must conclude that for all $l = 1,\dots,n+1$, the $K/C_K(t)$ norm of $w_{i,j_i,k_l,j_{k_l}}$ is a p-th power of some element of $C_K(t)$. On the other hand,

$$\begin{aligned} w_{i,j_i,k_lj_{k_l}}^{-1} &= \frac{w-d_{k_l,j_{k_l}}}{w-d_{i,j_i}} = 1 + \frac{d_{i,j_i}-d_{k_l,j_{k_l}}}{w-d_{i,j_i}} \\ &= (d_{i,j_i}-d_{k_l,j_{k_l}})\left(\frac{1}{d_{i,j_i}-d_{k_l,j_{k_l}}} - \frac{1}{d_{i,j_i}-w}\right). \end{aligned}$$

Thus, we can conclude that for $l = 1,\dots,n+1$,

$$\mathbf{N}_{K_w/C_K(t)}\left(\frac{1}{d_{i,j_i}-d_{k_l,j_{k_l}}} - \frac{1}{d_{i,j_i}-w}\right)$$

is a p-th power. Then, by Lemma 2.2, taking into account our assumption that for all natural numbers s, for $r \neq j, c_r^{p^s} \neq c_j$, we can conclude that $w - d_{i,j_i}$ is a p-th power in K. Consequently, w is a p-th power in K. Thus, $w = \tilde{w}^p$ for some $\tilde{w} \in K$. Next observe the following.

$$w_{i,j_i,k,j_k} = \frac{w - d_{i,j_i}}{w - d_{k,j_k}} = \frac{w - c_i^{p^{j_i}}}{w - c_k^{p^{j_k}}} = \left(\frac{\tilde{w} - c_i^{p^{m_i}}}{\tilde{w} - c_k^{p^{m_k}}}\right)^p = (\tilde{w}_{i,m_i,k,m_k})^p,$$

where $m_i = j_i - 1, m_k = j_k - 1$, if $j_k, j_i > 1$ and $m_i = r_i, m_k = r_k$, if $j_k = 1, j_i = 1$. Note that since for all k, j_k took all values $1, \ldots, r_k$, the same will be true of m_k. Thus Equations (2.9) and (2.10) can be rewritten in the following manner.

$$\tilde{w}_{i,m_i,k,m_k} - t_{i,k} = (u^p_{i,j_i,k,j_k} - \tilde{w}^p_{i,m_i,k,m_k}) - (u_{i,j_i,k,j_k} - \tilde{w}_{i,m_i,k,m_k}), \tag{2.11}$$

$$\frac{1}{\tilde{w}_{i,m_i,k,m_k}} - \frac{1}{t_{i,k}} = \left(v^p_{i,j_i,k,j_k} - \frac{1}{\tilde{w}^p_{i,m_i,k,m_k}}\right) - \left(v_{i,j_i,k,j_k} - \frac{1}{\tilde{w}_{i,m_i,k,m_k}}\right), \tag{2.12}$$

where $1 \leq m_i \leq r_i, 1 \leq m_k \leq r_k$. Equations (2.5) and (2.6) can be rewritten in a similar fashion. Therefore, the previous argument applies to $\tilde{w}$. Note also that the height of $\tilde{w}$ is strictly less than the height of w. Thus after finitely many iterations of this process, we will find ourselves in a situation where (2.5) and (2.6) hold for a $\bar{w} \in K$, whose height is less or equal to the height of t. This would imply that the divisor of $\bar{w}$ and t are the same. In other words, $\bar{w} = at$, where a is a constant. Thus, $(a-1)t = u^p - u$. However, unless $a = 1$, we have a contradiction. Therefore, if we assume that the height of $\bar{w}$ is less or equal to the height of t, we must conclude that $\bar{w} = t$. Consequently, for some natural number s, $w = t^{p^s}$.

Corollary 2.7. *The set* $\{w \in K | \exists s \in \mathbb{N}, w = t^{p^s}\}$ *is Diophantine over* K.

Proof. First we note that for any $x \in K$ and any $s \in \mathbb{N}$

$$x^{p^s} - x = (x^{p^{s-1}} + x^{p^{s-2}} + \cdots + x)^p - (x^{p^{s-1}} + x^{p^{s-2}} + \cdots + x). \tag{2.13}$$

Next we want to show that assuming $w = t^{p^s}$, Equations (2.5)-(2.10) can be satisfied over K. In view of equality (2.13), it is enough to show that for some $1 \leq j_i \leq r_i, 1 \leq j_k \leq r_k$, $w_{i,j_i,k,j_k} = (t_{i,k})^{p^s}$. Choose $j_i \cong s$ modulo r_i. (Such a j_i exists since the set of all possible values of j_i contains a representative of every class modulo r_i.) Then $c_i^{p^s} = (c_i^{p^{j_i}})^{p^{mr_i}} = c_i^{p^{j_i}}$. Similarly, choose $j_k \cong s$ modulo r_k so that $c_k^{p^s} = c_k^{p^{j_k}}$. Now the desired conclusion follows from Equations (2.7) and Equations (2.8).

Lemma 2.8. *Let $\sigma, \mu \in K$. Assume that all the primes that are poles of σ or μ do not ramify in the extension $\tilde{K}/\tilde{C}_K(t)$. Further, assume the following equality is true.*

$$t(\sigma^p - \sigma) = \mu^p - \mu. \tag{2.14}$$

Then $\sigma^p - \sigma = \mu^p - \mu = 0$.

Proof. Let $\mathfrak{A}$, $\mathfrak{B}$ be integral divisors of K, relatively prime to each other and to $\mathfrak{p}$ and $\mathfrak{q}$, such that the divisor of σ is of the form $\frac{\mathfrak{A}}{\mathfrak{B}}\mathfrak{p}^i\mathfrak{q}^k$, where i, k are integers. Then it is not hard to see that for some integral divisor $\mathfrak{C}$, relatively prime to $\mathfrak{B}$, $\mathfrak{p}$ and $\mathfrak{q}$, some integers j, m, the divisor of μ is of the form $\frac{\mathfrak{C}}{\mathfrak{B}}\mathfrak{p}^j\mathfrak{q}^m$. Indeed, let $\mathfrak{t}$ be a pole of μ such that $\mathfrak{t} \neq \mathfrak{p}$ and $\mathfrak{t} \neq \mathfrak{q}$. Then

$$0 > p\mathrm{ord}_{\mathfrak{t}}\mu = \mathrm{ord}_{\mathfrak{t}}(\mu^p - \mu) = \mathrm{ord}_{\mathfrak{t}}(t(\sigma^p - \sigma)) = \mathrm{ord}_{\mathfrak{t}}(\sigma^p - \sigma) = p\mathrm{ord}_{\mathfrak{t}}\sigma.$$

Conversely, let $\mathfrak{t}$ be a pole of σ such that $\mathfrak{t} \neq \mathfrak{p}$ and $\mathfrak{t} \neq \mathfrak{q}$. Then

$$0 > p\mathrm{ord}_{\mathfrak{t}}\sigma = \mathrm{ord}_{\mathfrak{t}}(\sigma^p - \sigma) = \mathrm{ord}_{\mathfrak{t}}(t(\sigma^p - \sigma)) = \mathrm{ord}_{\mathfrak{t}}(\mu^p - \mu) = p\mathrm{ord}_{\mathfrak{t}}\mu.$$

By the Strong Approximation Theorem there exists $b \in K$ such that the divisor of b is of the form $\mathfrak{B}\mathfrak{D}/\mathfrak{q}^l$, where $\mathfrak{D}$ is an integral divisor relatively prime to $\mathfrak{A}, \mathfrak{C}, \mathfrak{p}, \mathfrak{q}$ and l is a natural number. Then $b\sigma = s_1t^i, b\mu = s_2t^j$, where s_1, s_2 are integral over $C_K[t]$ and have zero divisors relatively prime to $\mathfrak{p}$ and $\mathfrak{B}$. Indeed, consider the divisors of $b\sigma$:

$$\frac{\mathfrak{B}\mathfrak{D}}{\mathfrak{q}^l}\frac{\mathfrak{A}}{\mathfrak{B}}\mathfrak{p}^i\mathfrak{q}^k = \mathfrak{D}\mathfrak{A}\mathfrak{p}^i\mathfrak{q}^{k-l} = \mathfrak{D}\mathfrak{A}\mathfrak{q}^{k-l+i}\frac{\mathfrak{p}^i}{\mathfrak{q}^i}.$$

Thus the divisor of s_1 is of the form $\mathfrak{D}\mathfrak{A}\mathfrak{q}^{k-l+i}$ and therefore, $\mathfrak{q}$ is the only pole of s_1, making it integral over $C_k[t]$. Further, by construction $\mathfrak{A}$ and $\mathfrak{D}$ are integral divisors relatively prime to $\mathfrak{p}$ and $\mathfrak{B}$. A similar argument applies to s_2.

Multiplying through by b^p we will obtain the following equation.

$$t(s_1^p t^{ip} - b^{p-1}s_1t^i) = s_2^p t^{jp} - b^{p-1}s_2t^j. \tag{2.15}$$

Suppose $i < 0$. Then the left side of (2.15) has a pole of order $ip + 1$ at $\mathfrak{p}$. This would imply that $j < 0$ and the right side has a pole of order jp at $\mathfrak{p}$. Thus, we can assume that i, j are both nonnegative. We can now rewrite (2.15) in the form

$$(s_1^p t^{ip+1} - s_2^p t^{jp}) = b^{p-1}(s_1t^{i+1} - s_2t^j). \tag{2.16}$$

Let $\mathfrak{t}$ be any prime factor of $\mathfrak{B}$ in $\tilde{K}$. Then $\mathfrak{t}$ does not ramify in the extension $\tilde{K}/C_{\tilde{K}}(t)$ and since $p > 2$, $\mathrm{ord}_{\mathfrak{t}}(s_1^p t^{ip+1} - s_2^p t^{jp}) \geq 2$. Thus, $\mathrm{ord}_{\mathfrak{t}}\partial(s_1^p t^{ip+1} - s_2^p t^{jp})/\partial\mathfrak{t} > 0$. Since $\mathfrak{t}$ is not ramified in the extension $\tilde{K}/C_{\tilde{K}}(t)$, by Lemma 6.17, $\mathrm{ord}_{\mathfrak{t}}\partial(s_1^p t^{ip+1} - s_2^p t^{jp})/\partial\mathfrak{t} = \mathrm{ord}_{\mathfrak{t}}d(s_1^p t^{ip+1} - s_2^p t^{jp})/dt = \mathrm{ord}_{\mathfrak{t}}(s_1^p t^{ip})$. Therefore, since $\mathfrak{t}$, by assumption is not a zero of t, s_1 has a zero at $\mathfrak{t}$. This, however, is impossible. Consequently, $\mathfrak{B}$ is a trivial divisor, and in (2.14) all the functions are integral over $C_K[t]$, i.e., they can have poles at $\mathfrak{q}$ only. Assuming μ is not a constant and thus has a pole at $\mathfrak{q}$, we note that the left side has a pole at $\mathfrak{q}$ of order equivalent to 1 modulo p, while the right side has the pole $\mathfrak{q}$ of order equivalent to 0 modulo p. Thus, μ is a constant. But the only way the product of t and a function integral over $C_K[t]$ can be a constant is for that function to be equal to zero. Consequently, the statement of the lemma is true.

Lemma 2.9. *Let $v \in K$ and assume for some distinct $a_0 = 0, a_1, \dots, a_n \in C_K$, the divisor of $v + a_0, \dots, v + a_n$ is a p-th power of some other divisor of K. Then, assuming for all i, $v + a_i$ does not have any zeros or poles at any prime ramifying in the extension $K/C_K(t)$, v is a p-th power in K.*

Proof. First assume $v \in C_K(t)$. Since $v + a_i$ does not have any zeros or poles at primes ramifying in the extension $K/C_K(t)$, the divisor of $v + a_i$ in $C_K(t)$ is a p-th power of another $C_K(t)$ divisor. Since in $C_K(t)$ every zero degree divisor is principal, v is a p-th power in $C_K(t)$ and therefore in K. Next assume $v \notin C_K(t)$. Note that no zero or pole of $v + a_i$ is at any valuation ramifying in the extension $K/C_K(t, v)$. Hence, in $C_K(t, v)$ the divisor of $v + a_i$ is also a p-th power of another divisor. Finally note that $\mathbf{N}_{C_K(t,v)/C_K(t)}(v + a_i)$ will be a p-th power in $C_K(t)$ and apply Lemma 2.2.

Lemma 2.10. *Let $x, v \in K \setminus \{0\}$, let $u = \frac{x^p+t}{x^p-t}$. Further, assume that the following equations hold for all $i, k = 0, \dots, (2r + n + 5)$, some $1 \le j_i \le r_i, 1 \le j_k \le r_k$, and some $s \ge 0$.*

$$u_{i,k,g} = \frac{u^g + c_i}{u^g + c_k}, g = -1, 1. \tag{2.17}$$

$$v_{i,j_i,k,j_k,g} = \frac{v^g + d_{i,j_i}}{v^g + d_{k,j_k}}, g = -1, 1. \tag{2.18}$$

$$v^{2e}_{i,j_i,k,j_k,g} t^{mp^s} - u^{2e}_{i,k,g} t^m = \mu^p_{i,j_i,k,j_k,e,m,g} - \mu_{i,j_i,k,j_k,e,m,g}, \quad e = -1, 1, m = 0, 1, g = -1, 1. \tag{2.19}$$

$$v^e_{i,j_i,k,j_k,g} - u^e_{i,k,g} = \sigma^p_{i,j_i,k,j_k,e,g} - \sigma_{i,j_i,k,j_k,e,g}, e = -1, 1, g = -1, 1. \tag{2.20}$$

$$(u^g + c_i)^e - (v^g + d_{i,j_i})^e = \mu^p_{i,j_i,e,g} - \mu_{i,j_i,e,g}, e = -1, 1, g = -1, 1. \tag{2.21}$$

Then for some natural number $k, v = u^{p^k}$.

Proof. First of all, we claim that for all i, k, g, $u_{i,k,g}$ has no multiple zeros or poles except possibly at the primes ramifying in $\tilde{K}/\tilde{C}_K(t)$, $\mathfrak{p}$ or $\mathfrak{q}$. Indeed, by Lemma 2.4, all the poles of $u_{i,k,g}$ are zeros of $u^g + c_k$ and all the zeros of $u_{i,k,g}$ are zeros $u^g + c_i$. However, by Lemma 4.5 of [**38**] and by assumption on c_i and c_k, all the zeros of $u^g + c_k$ and $u^g + c_i$ are simple, except possibly for zeros at $\mathfrak{p}$, $\mathfrak{q}$, or primes ramifying in the extension $\tilde{K}/\tilde{C}_K(t)$. For future use, we also note that u is not a p-th power in K, assuming $x \neq 0$. (This can be established by computing the derivative of u, which is not 0, if x is not 0.) We will show that if $s > 0$ then v is a p-th power in K, and if $s = 0$ then $u = v$. Suppose $s > 0$ and let $g = 1$. Next note that by Lemma 2.3, by an argument similar to the one used in Lemma 2.6, there exist $0 \leq i \leq (2r+5+n), 0 \leq k_l \leq (2r+n+5), l = 1, \dots, n+1, k_l \neq i, k_l \neq k_m$ for $m \neq l$, such that for all $1 \leq j_i \leq r_i, 1 \leq j_{k_l} \leq r_{k_l}$, $u_{i,k_l,1}$ and $v_{i,j_i,k_l,j_{k_l},1}$ have no zeros or poles at the primes of $\tilde{K}$ ramifying in the extension $\tilde{K}/\tilde{C}(t)$, or $\mathfrak{p}$ or $\mathfrak{q}$. Note that for thus selected indices, all the poles and zeros of $u_{i,k_l,1}$ are simple for $l = 1, \dots, n+1$.

Pick an $i, k_1, \dots, k_{n+1}, j_i, j_{k_1}, \dots, j_{k_{n+1}}$ such that Equations (2.17)-(2.21) are satisfied for these indices and $u_{i,k_1,1}, v_{i,j_i,k_1,j_{k_1},1}, \dots, u_{i,k_{n+1},1}$, $v_{i,j_i,k_{n+1},j_{k_{n+1}},1}$ have no poles or zeros at primes ramifying in the extension $\tilde{K}/\tilde{C}(t)$, or at $\mathfrak{p}$ or $\mathfrak{q}$. Further, by an argument similar to the one used in the proof of Lemma 2.5, either for $l = 1, \dots, n+1$, the divisor of $v_{i,j_i,k_l,j_{k_l},1}$ in $\tilde{K}$ is a p-th power of another divisor or for some l and some prime $\mathfrak{t}$ not ramifying in $\tilde{K}/\tilde{C}(t)$ and not equal to $\mathfrak{p}$ or to $\mathfrak{q}$, $\text{ord}_{\mathfrak{t}} v_{i,j_i,k_l,j_{k_l},1} = \pm 1$. In the first case, given the assumption that $v_{i,j_i,k_l,j_{k_l},1}$'s do not have poles or zeros at ramifying primes and Lemma 2.9, v is a p-th power in K. So suppose the second alternative holds. In this case, without loss of generality, assume $\mathfrak{t}$ is a pole of $v_{i,j_i,k_l,j_{k_l},1}$. Next consider the following equations

$$v^2_{i,j_i,k_l,j_{k_l},1} t^{p^s} - u^2_{i,k_l,1} t = \mu^p_{i,j_i,k_l,j_{k_l},1,1,1} - \mu_{i,j_i,k_l,j_{k_l},1,1,1}, \tag{2.22}$$

$$v^2_{i,j_i,k_l,j_{k_l},1} - u^2_{i,k_l,1} = \mu^p_{i,j_i,k,j_k,0,1,1} - \mu_{i,j_i,k,j_k,0,1,1}, \tag{2.23}$$

obtained from (2.19) by first making $e = 1, m = 1$ and then $e = 1, m = 0$. (If $\mathfrak{t}$ were a zero of $v_{i,j_i,k_l,j_{k_l},1}, e$ would be equal to -1 in both equations.) Since t does not have a pole or zero at $\mathfrak{t}$ and $p > 2$, we must conclude that

$$\text{ord}_{\mathfrak{t}}(v^2_{i,j_i,k_l,j_{k_l},1} t^{p^s} - u^2_{i,k_l,1} t) = \text{ord}_{\mathfrak{t}}(\mu^p_{i,j_i,k_l,j_{k_l},1,1,1} - \mu_{i,j_i,k_l,j_{k_l},1,1,1}) \geq 0$$

and

$$\text{ord}_{\mathfrak{t}}(v^2_{i,j_i,k_l,j_{k_l},1} - u^2_{i,k_l,1}) = \text{ord}_{\mathfrak{t}}(\mu^p_{i,j_i,k_l,j_{k_l},0,1,1} - \mu_{i,j_i,k_l,j_{k_l},0,1,1}) \geq 0.$$

Thus,

$$\begin{aligned} &\text{ord}_{\mathfrak{t}} v^2_{i,j_i,k_l,j_{k_l},1}(t^{p^s} - t) \\ &= \text{ord}_{\mathfrak{t}}(\mu^p_{i,j_i,k_l,j_{k_l},1,1,1} - \mu_{i,j_i,k_l,j_{k_l},1,1,1} - t\mu^p_{i,j_i,k_l,j_{k_l},0,1,1} + t\mu_{i,j_i,k_l,j_{k_l},0,1,1}) \\ &\geq 0. \end{aligned}$$

Finally, we must deduce that $\text{ord}_{\mathfrak{t}}(t^{p^s} - t) \geq 2|\text{ord}_{\mathfrak{t}} v|$. But in $C_K(t)$ all the zeros of $(t^{p^s} - t)$ are simple. Thus, this function can have multiple zeros only at primes ramifying in the extension $\tilde{K}/\tilde{C}_K(t)$. By assumption $\mathfrak{t}$ is not one of these primes and thus we have a contradiction, unless v is a p-th power.

Suppose now that $s = 0$. Set $g = 1$ again and let $i, k_1, \ldots, k_{n+1}$ be selected as above. Then from (2.22) and (2.23) we obtain

$$\mu^p_{i,j_i,k,j_k,1,1,1} - \mu_{i,j_i,k,j_k,1,1} = t(\mu^p_{i,j_i,k,j_k,0,1,1} - \mu_{i,j_i,k,j_k,0,1,1}).$$

Note here that all the poles of $\mu_{i,j_i,k,j_k,1,1,1}$ and $\mu_{i,j_i,k,j_k,0,1,1}$ are poles of $u_{i,k_l,1}, v_{i,j_i,k_l,j_{k_l},1}$ or t, and thus are not any valuation ramifying in the extension $\tilde{K}/\tilde{C}_K(t)$. By Lemma 2.8 we can then conclude that

$$v^2_{i,j_i,k_l,j_{k_l},1} - u^2_{i,k_l,1} = 0.$$

Thus, $v_{i,j_i,k_l,j_{k_l},1} = \pm u_{i,k_l,1}$. Since all the poles of $u_{i,k_l,1}$ are simple, (2.20) with $k = k_l$ rules out "$-$". Therefore,

$$v_{i,j_i,k_l,j_{k_l},1} = u_{i,k_l,1}. \tag{2.24}$$

Rewriting (2.24) we obtain

$$\frac{d_{i,j} - d_{k_l,j_{k_l}}}{v + d_{k_l,j_{k_l}}} = \frac{c_i - c_{k_l}}{u + c_{k_l}},$$

or

$$v = au + b, \tag{2.25}$$

where a, b are constants.

Now keep $s = 0$, set $g = -1$, pick new distinct $i, k_1, \ldots, k_{n+1}$ such that $u_{i,k_l,-1}, v_{i,j_i,k_l,j_{k_l},-1}, l = 1, \ldots, n+1$ do not have any zeros or poles at valuations ramifying in the extension $\tilde{K}/\tilde{C}_K(t)$, at $\mathfrak{p}$ or $\mathfrak{q}$. Repeat the argument above (with $s = 0$) for $g = -1$ to conclude that

$$v^{-1} = \bar{a}u^{-1} + \bar{b}, \tag{2.26}$$

where $\bar{a}, \bar{b}$ are also constants. Equation (2.25) stipulates that u and v have the same poles. If $\bar{b} \neq 0$, then (2.26) stipulates that u and v have no poles in common. Since u is not constant, and therefore, v is not constant, we must deduce that $\bar{b} = 0$ and $u = av$ for some constant a. If $a \neq 1$, from (2.21), we conclude, using $g = 1$ for all i, that all the zeros of $u + c_i$ are of order divisible by p. Indeed, consider $\frac{1}{u+c_i}$ and $\frac{1}{u+a^{-1}d_{i,j}}$. Either $ac_i = c_i^{p^{j_i}} = d_{i,j_i}$ and

$$\frac{1}{u+c_i} - \frac{1}{v+d_{i,j_i}} = \frac{1}{u+c_i} - a^{-1}\frac{1}{u+c_i} = (1-a^{-1})/(u+c_i),$$

or $u + c_i$ and $v + d_{i,j_i} = au + d_{i,j_i}$ have no common zeros, and $\frac{1}{u+c_i} - \frac{1}{v+d_{i,j_i}}$ has poles at all the valuations at which $\frac{1}{u+c_i}$ has poles, and these poles are of the same order as the poles of $\frac{1}{u+c_i}$. Since this cannot happen, $a = 1$.

If $s > 0$ and v is a p-th power, then Equations (2.18)-(2.21) can be rewritten in the same fashion as equations in Lemma 2.6 with s being replaced by $s-1$ and v replaced by its p-th root. Therefore, after finitely many iterations of this rewriting procedure we will be in the case of $s = 0$. Hence, for some natural number k, $v = u^{p^k}$.

Corollary 2.11. *Let $x \in K$, and let $u = \frac{x^p+t}{x^p-t}$. Then the set $\{w \in K \| \exists s \in \mathbb{N},\ w = u^{p^s}\}$ is Diophantine over K.*

Proof. Given Lemma 2.10, it is enough to show that if $w = u^{p^s}$ for some natural number s, Equations (2.17) - (2.21) can be satisfied in the remaining variables over K. The proof of this assertion is identical to the proof of Corollary 2.7.

Finally we state the main result of this section.

Theorem 2.12. *The set $\{(x, y) \in K^2 \| \exists s \in \mathbb{N},\ y = x^{p^s}\}$ is Diophantine over K.*

Given Corollaries 2.7 and 2.11, the proof of this theorem will be identical to the proof of Theorem 5.12 of [**38**].

3. Integrality at One Prime: The case of $q = p$.

In this section we will show that integrality at one prime is a Diophantine condition over an algebraic function field of characteristic $p > 0$ whose constant field has an extension of degree $p > 0$.

Lemma 3.1. *Let L be a local field or an algebraic function field of positive characteristic p. Let $v \in L$ and let α be a root of the equation*

$$x^p - x - v = 0. \tag{3.1}$$

Then either $\alpha \in L$ or α is of degree p over L. Further, in the second case the extension $L(\alpha)/L$ is cyclic of degree p and the only primes possibly ramified in this extension are the poles of v. On the other hand, if for some L-prime $\mathfrak{a}$, $\text{ord}_{\mathfrak{a}} v \not\cong 0$ modulo p and $\text{ord}_{\mathfrak{a}} v < 0$, then a factor of $\mathfrak{a}$ in $L(\alpha)$ will be ramified completely.

Proof. Let $\alpha = \alpha_1, \ldots, \alpha_p$ be all the roots of (3.1) in the algebraic closure of L. Then we can number the roots so that $\alpha_i = \alpha + i - 1$. Thus, either the left side of (3.1) factors completely or it is irreducible. In the second case α is of degree p over L and $L(\alpha)$ contains all the conjugates of α over L. Thus, the extension $L(\alpha)/L$ is Galois of degree p, and therefore is cyclic. Next consider the different of α. This different is a constant. By [**3**, Lemma 2, page 71], this implies that no prime of L at which α is integral has any ramified factors in the extension $L(\alpha)/L$. Finally, suppose $\mathfrak{a}$ is a prime of L described in the statement of the lemma. Let $\tilde{\mathfrak{a}}$ be an $L(\alpha)$ prime above $\mathfrak{a}$. Then $\text{ord}_{\tilde{\mathfrak{a}}} v \cong 0$ modulo p. Thus, $\tilde{\mathfrak{a}}$ must be totally ramified over $\mathfrak{a}$.

Lemma 3.2. *Let M/K be a Galois extension of algebraic function fields of degree n. Let $\mathfrak{p}$ be a prime of K which does not split in M. Let $h \in K$ be such that* $\text{ord}_{\mathfrak{p}} h \not\cong 0$ *modulo n. Then h is not a norm of an element of M.*

Lemma 3.3. *Let H/F be an unramified extension of local fields of degree n. Let $\mathfrak{t}$ be the prime of F. Let $x \in F$ be such that* $\text{ord}_{\mathfrak{t}} x \cong 0$ *modulo n. Then x is a norm of some element of H.*

Proof. Let π be a local uniformizing parameter for $\mathfrak{t}$. Then $x = \pi^n \varepsilon$, where ε is a unit. Since π^n is an F-norm, x is an F-norm if and only if ε is an F norm. The last statement is true by [**42**, Corollary, page 226].

Lemma 3.4. *Let L be an algebraic function field. Let $\mathfrak{C}$ and $\mathfrak{B}$ be prime divisors of L. Let $v \in L$ be such that the divisor of v is of the form $\mathfrak{C}^{-1}\mathfrak{V}$, where $\mathfrak{V}$ is a divisor of L which has no common factors with $\mathfrak{C}$ or $\mathfrak{B}$. Further, assume v is equivalent to $b^p - b$ modulo $\mathfrak{B}$, where $b \in C_L$, the constant field of L.* (*Such a v exists by the Weak Approximation theorem.*) *Let β be a root of* (3.1). *Let $R_{\mathfrak{B}}$ be the residue field of $\mathfrak{B}$ in L and assume it is separable over C_L. Let δ be an element of the algebraic closure of C_L, such that $C_L(\delta)$ is isomorphic to the Galois closure of $R_{\mathfrak{B}}$ over C_L. Let $\bar{L} = L(\beta, \delta)$. Then in $\bar{L}$,*

$$\mathfrak{C} = \prod \mathfrak{c}_i^p, \tag{3.2}$$

and

$$\mathfrak{B} = \prod \mathfrak{b}_i, \tag{3.3}$$

where for $i \neq j$, $\mathfrak{b}_i \neq \mathfrak{b}_j$, $\mathfrak{c}_i \neq \mathfrak{c}_j$, and for all $i, \mathfrak{b}_i$ is of degree 1.

Proof. By Lemma 3.1, $[L(\beta) : L] = p$, and the prime above $\mathfrak{C}$ in $L(\beta)$ is totally ramified. Thus, in $L(\beta)$, $\mathfrak{C} = \mathfrak{c}^p$. Note that by Lemma 3.1, $\mathfrak{B}$ does not have ramifying factors in the extension $L(\beta)/L$. On the other hand, the left side of (3.1) will factor completely modulo $\mathfrak{B}$. Since all the coefficients of the left side (3.1) are integral at $\mathfrak{B}$ and $\mathfrak{B}$ is not a zero of the discriminant of this polynomial, β generates a local integral basis with respect to $\mathfrak{B}$. Thus the fact that left side of (3.1) will factor completely modulo $\mathfrak{B}$ implies that $\mathfrak{b}$ will split completely in $L(\beta)$. Thus, the residue fields of the factors of $\mathfrak{B}$ in $L(\beta)$ are the same as the residue field of $\mathfrak{B}$ in L. Next note that the constant fields of L and $L(\beta)$ are the same because $\mathfrak{C}$ has a completely ramified factor in this separable extension. Hence the residue fields of the factors of $\mathfrak{b}$ in $L(\beta)$ are separable over $C_{L(\beta)} = C_L$. Consequently, we can apply Lemma 6.14 to assert that in $L(\beta, \delta)$ all the factors of $\mathfrak{b}$ will be of degree 1. Finally we note that no factor of $\mathfrak{b}$ is ramified in the extension $L(\beta, \delta)/L$, so that all the factors in the product (3.3) are distinct. Similarly, no prime ramifies in the extension $L(\beta, \delta)/L(\beta)$, and all the factors in the product (3.2) are distinct.

Lemma 3.5. *Let $\mathfrak{b}$ be a prime of K and let $\mathfrak{B}$ be a prime of G below $\mathfrak{b}$. Assume $\mathfrak{B}$ is not trivial. Let $R_{\mathfrak{B}}$ be the residue field of $\mathfrak{B}$ in G, and let $\delta_{\mathfrak{B}}$ be a generator of the extension of C isomorphic to $R_{\mathfrak{B}}$. Let $\bar{K}$ be a separable extension of K where $\mathfrak{b}$ splits into factors of degree 1. Let $\hat{G}$ be the algebraic closure of G in $\bar{K}$. Then $\hat{G}$ contains $\delta_{\mathfrak{B}}$.*

Proof. Let $\mathfrak{b} = \prod_{i=1}^{m} \mathfrak{b}_i$ be the factorization of $\mathfrak{b}$ in $\bar{K}$. Then each $\mathfrak{b}_i$ lies over a nontrivial prime $\mathfrak{B}_i$ of $\hat{G}$. (This is true because each $\mathfrak{B}_i$ is an extension of $\mathfrak{B}$.) Let $x \in G$ be such that its residue class generates $R_{\mathfrak{B}}$ over C, and let $F(T) \in C[T]$ be the irreducible polynomial of the residue class of x over C. Then $F(\delta_{\mathfrak{B}}) = 0$ and $F(x) \cong 0$ modulo $\mathfrak{B}$. On the other hand, since $\mathfrak{b}_1$ is of degree 1, there exists $a \in C_{\bar{K}}$ such that $x - a \cong 0$ modulo $\mathfrak{b}_1$. Hence, $0 \cong F(x) \cong F(a)$ modulo $\mathfrak{b}_1$. But $F(a)$ is a constant. Therefore, $F(a) = 0$. Since every extension of C is Galois, and a is a conjugate of $\delta_{\mathfrak{B}}$ over C, $\delta_{\mathfrak{B}} \in C(a) \subset \hat{G}$.

Lemma 3.6. *Let $\mathfrak{a}$ be a prime of K with a nontrivial restriction to G. Let $\mathfrak{A}$ be the prime below $\mathfrak{a}$ in G. Let $\bar{K} \supset \bar{G}$ be finite separable extensions of K and G respectively. Let $\bar{\mathfrak{a}}$ be a prime above $\mathfrak{a}$ in $\bar{K}$. Let $\bar{\mathfrak{A}}$ be the prime below $\bar{\mathfrak{a}}$ in $\bar{G}$. Then $\bar{\mathfrak{A}}$ lies above $\mathfrak{A}$ in $\bar{G}$ (and thus is not a trivial prime of $\bar{G}$). Further, if we assume that $e(\bar{\mathfrak{a}}/\mathfrak{a}) = e(\bar{\mathfrak{A}}/\mathfrak{A})$, then $e(\bar{\mathfrak{a}}/\bar{\mathfrak{A}}) = e(\mathfrak{a}/\mathfrak{A})$.*

Proof. Let $\bar{\mathfrak{a}}$ and $\bar{\mathfrak{A}}$ be as in the statement of the lemma. Then, since $\bar{\mathfrak{a}}$ restricts to $\bar{\mathfrak{A}}$ in $\bar{G}$ and to $\mathfrak{A}$ in G, we must conclude that the restriction of $\bar{\mathfrak{A}}$ to G is also $\mathfrak{A}$. Further, we have the following equality. $e(\bar{\mathfrak{a}}/\bar{\mathfrak{A}})e(\bar{\mathfrak{A}}/\mathfrak{A}) = e(\bar{\mathfrak{a}}/\mathfrak{a})e(\mathfrak{a}/\mathfrak{A})$. Thus, $e(\bar{\mathfrak{a}}/\bar{\mathfrak{A}}) = e(\mathfrak{a}/\mathfrak{A})$.

Lemma 3.7. *Let $\mathfrak{a}, \mathfrak{b}$ be two primes of K, restricting to nontrivial primes $\mathfrak{A}$ and $\mathfrak{B}$ of G respectively. Assume the residue field of $\mathfrak{b}$ is separable over C_K. Further, assume that $\mathfrak{b}$ and $\mathfrak{a}$ are unramified over G and are the only factors of $\mathfrak{B}$ and $\mathfrak{A}$ in K. Let $f \in G$ be such that its divisor is of the form*

$$\frac{\mathfrak{B}^s}{\mathfrak{A}\mathfrak{U}}, \tag{3.4}$$

where $\mathfrak{U}$ is an integral divisor of G relatively prime to $\mathfrak{B}$ and $\mathfrak{A}$, and s is a natural number such that $s \not\cong 0$ modulo p. Let $v \in G$ be such that it has a pole of order 1 at all the primes which are factors of the pole divisor of f and is equivalent to $b^p - b$ modulo $\mathfrak{B}$ for some $b \in C$. (Such an $f \in G$ exists by the Strong Approximation Theorem.) Let δ_G, an element of the algebraic closure of C, be a generator of the residue field of $\mathfrak{B}$ over C and let δ_K, an element of the algebraic closure of C_K, be a generator of the Galois closure of the residue field of $\mathfrak{b}$ over C_K. Let β be a root of (3.1) in the algebraic closure of K. Let $w \in G$, and let $\mathfrak{A}_w$, $\mathfrak{B}_w$, $\mathfrak{U}_w$ be the primes and the divisor below $\mathfrak{A}$ and $\mathfrak{B}$ and $\mathfrak{U}$ respectively in $G_{t,x,v,w,f}$. Let $\bar{G}_w = G_{x,t,v,w,f,\beta}$. Let $\bar{K} = K(\beta, \delta_K), \bar{G} = G(\beta, \delta_G)$. Then the following statements are true.

1) $\bar{G}_w \subset \bar{G} \subset \bar{K}$.
2) $\bar{G}/\bar{G}_w$ *is a separable (possibly infinite) constant field extension, and thus no prime ramifies in this extension.*
3) *The primes* $\mathfrak{A}_w$ *and* $\mathfrak{B}_w$ *are distinct in* $\bar{G}_w$.
4) *In* $\bar{K}$, $\mathfrak{b} = \prod \mathfrak{b}_i$, *where* $\mathfrak{b}_i$ *are distinct prime divisors of degree* 1.
5) *In* $\bar{G}$, $\mathfrak{B} = \prod \mathfrak{B}_i$, *where* $\mathfrak{B}_i$ *are distinct prime divisors of degree* 1.
6) *For each* i *there exists* j *such that* $\mathfrak{b}_i$ *lies above* $\mathfrak{B}_j$ *and* $e(\mathfrak{b}_i/\mathfrak{B}_j) = 1$.
7) *In* $\bar{K}$, $\mathfrak{a} = \prod \mathfrak{a}_i^p$, *where* $\mathfrak{a}_i$ *are distinct prime divisors.*
8) *In* $\bar{G}$, $\mathfrak{A} = \prod \mathfrak{A}_i^p$, *where* $\mathfrak{A}_i$ *are distinct prime divisors.*
9) *For each* i *there exists* j *such that* $\mathfrak{a}_i$ *lies above* $\mathfrak{A}_j$ *and* $e(\mathfrak{a}_i/\mathfrak{A}_j) = 1$.
10) *In* $\bar{G}_w$, $\mathfrak{A}_w = \mathfrak{a}_w^p$ *for some prime* $\mathfrak{a}_w$ *of* $\bar{G}_w$.
11) *There exist* $z_1 \in \bar{G}$ *such that its divisor is of the form* $\mathfrak{T}\mathfrak{A}_1^{-1}$, *where* $\mathfrak{T}$ *is a divisor of* $\bar{G}$ *relatively prime to* $\mathfrak{B}$ *and* $\mathfrak{A}$, *such that for some* $b \in C, z_1 \cong b^p - b$ *modulo* $\mathfrak{B}$.
12) *Let* $k > 0, k \not\cong 0$ *modulo* p *be greater than the highest order of any pole of* z_1 *in* $\bar{K}$. *Then there exists* $z_2 \in \bar{G}$ *such that its divisor is of the form* $\mathfrak{W}\mathfrak{A}_1^{-p^l}$, *where* $p^l > p^2k$ *and* $\mathfrak{W}$ *is an integral divisor of* $\bar{G}$ *relatively prime to* $\mathfrak{B}$ *and* $\mathfrak{A}$. *Further,* $z_2 \cong 1$ *modulo* $\mathfrak{B}$.
13) *Let* $z = z_1z_2$. *Then in* $\bar{G}$, z *has a pole of order* $p^l+1 > p^2k$ *at* $\mathfrak{A}_1$ *and is equivalent to* $b^p - b$ *modulo* $\mathfrak{B}$. *All the other poles of* z *are of order less than* k.
14) *In* $\bar{G}_w(z)$, z *has a pole of order* $p^l+1 > p^2k$ *at the prime below* $\mathfrak{A}_1$ *and is equivalent to* $b^p - b$ *modulo the divisor below* $\mathfrak{B}$.
15) *In* $\bar{G}$, $\mathfrak{U}$ *is a* p*-th power of another divisor.*
16) *In* $\bar{G}_w$, $\mathfrak{U}_w$ *is a* p*-th power of another divisor.*

Proof. 1. First of all, $x, t \in G$ by construction, $v, f, w \in G$ by assumption. Therefore, $G_{x,t,v,f.w} \subset G$, and $\bar{G}_w = G_{x,t,v,f,w,\beta} \subset G(\beta, \delta_G) = \bar{G}$. Secondly, by Lemma 6.14, $\mathfrak{b}$ will split into factors of degree 1 in $\bar{K}$. Therefore, $\delta_G \in K(\delta_K)$ by Lemma 3.5. Hence, $\bar{G} \subset \bar{K}$.

2. $\bar{G} = C(t, x, \beta, \delta_G) = C\bar{G}_w(\beta)$.

3. Since $v \in \bar{G}_w$ has a pole at $\mathfrak{A}_w$ but not at $\mathfrak{B}_w$, these primes must be distinct in $\bar{G}_w$.

4,5,7,8,10,15,16. These statements follow from Lemma 3.4.

6,9. These statements follow from Lemma 3.6.

11, 12. These statements follow from Lemma 6.15.

13. This statement follows from a direct calculation of the orders of poles of z.

14. This statement follows from the fact that $\bar{G}/\bar{G}_w(z)$ is also a separable constant field extension and thus no prime is ramified.

Lemma 3.8. *Let* $\mathfrak{b}$, $\mathfrak{a}$, $\mathfrak{A}$, $\mathfrak{B}$, $\bar{G}$, $\bar{K}$, $\mathfrak{a}_i$, $\mathfrak{A}_i$, $\mathfrak{b}_i$, $\mathfrak{B}_i, f, z, s, k$ *be as in Lemma* 3.7. *Let* $w \in K$. *If* $w \in G$ *then let* $\bar{G}_w$ *be as in Lemma* 3.7. *Let* $a \in C_{\bar{K}}$ *be* *such that the equation*

$$x^p - x - a = 0 \tag{3.5}$$

has no solution in $C_{\bar{K}}$*, the constant field of* $\bar{K}$*, while* a *is algebraic over a finite field. If* $w \in G$*, without loss of generality, we can assume that* $a \in G_w$*. (*G_w *can be any subfield of* G *containing the elements listed above and such that its constant field is finite.) Let*

$$h = f^{-1}w^{p(s+1)} + f^{-p}. \tag{3.6}$$

Let β_w *be a root of the equation*

$$x^p - x - (h^{-k} + z) = 0. \tag{3.7}$$

Let α *be a root of the Equation* (3.5)*. Then the following statements are true.*

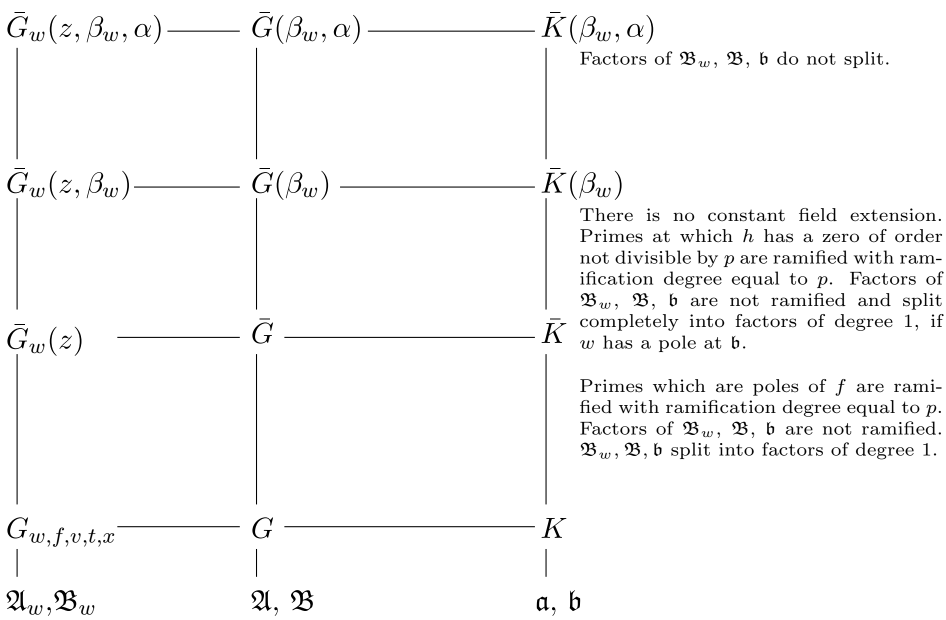
1) *If* $w \in K$ *has a pole at* $\mathfrak{b}$*, then the equation*

$$\prod_{i=0}^{p-1}(a_0 + a_1(\alpha + i) + \cdots + a_{p-1}(\alpha + i)^{p-1}) = h \tag{3.8}$$

has no solution $(a_0, \ldots, a_{p-1}) \in \bar{K}(\beta_w)$*.*

2) *If* $w \in G$ *has no pole at* $\mathfrak{B}$*, then Equation* (3.8) *will have a solution* $(a_0, \ldots, a_{p-1}) \in \bar{G}_w(z, \beta_w) \subset \bar{G}(\beta_w)$*.*

Proof. The following figure describes the extensions involved. The two left columns correspond to the case of $w \in G$.



Before we proceed with the proof we will discuss the following three points. First of all, we will show below that for all $w \in G$,

$$[\bar{K}(\beta_w) : \bar{K}] = [\bar{K}(\beta_w, \alpha) : \bar{K}(\beta_w)] = p, \tag{3.9}$$

$$[\bar{G}(\beta_w) : \bar{G}] = [\bar{G}(\beta_w, \alpha) : \bar{G}(\beta_w)] = p, \tag{3.10}$$

$$[\bar{G}_w(\beta_w, z) : \bar{G}_w(z)] = [\bar{G}_w(z, \beta_w, \alpha) : \bar{G}_w(z, \beta_w)] = p, \tag{3.11}$$

while (3.9) holds for all $w \in K$. Secondly, it is not hard to see that the existence of solutions $a_0, \dots, a_{p-1} \in \bar{K}(\beta_w)$ to (3.8) is equivalent to existence of $u \in \bar{K}(\alpha, \beta_w)$ such that

$$\mathbf{N}_{\bar{K}(\alpha,\beta_w)/\bar{K}(\beta_w)}(u) = h. \tag{3.12}$$

Finally, assume $w \in G$. Then it is also not hard to see that (3.8) has solutions in $\bar{G}_w(z, \beta_w)$ if and only if there exists $u \in \bar{G}_w(z, \alpha, \beta_w)$ such that

$$\mathbf{N}_{\bar{G}_w(z,\alpha,\beta_w)/\bar{G}_w(z,\beta_w)}(u) = h. \tag{3.13}$$

In order to show that (3.9)-(3.11) hold, we will show that in extensions $\bar{K}(\beta_w)/\bar{K}$, $\bar{G}(\beta_w)/\bar{G}$, and $\bar{G}_w(\beta_w, z)/\bar{G}_w(z)$ at least one prime will have ramification degree p while the degree of each extension listed above is at most p. (As above, when we consider the last two extensions, we assume that $w \in G$.) Since all the extensions listed above are separable, the presence of a totally ramified prime will imply that there is no constant field extension in either of the three extensions. Thus, since α was of degree p over $C_{\bar{K}}, C$ and C_w- the constant field of G_w, it will remain of degree p over the constant fields of $\bar{K}(\beta_w), \bar{G}(\beta_w)$ and $\bar{G}_w(\beta_w, z)$. We can assume without loss of generality that a $\bar{K}$-prime $\mathfrak{a}_1$ lies above a $\bar{G}$-prime $\mathfrak{A}_1$. In this case, by Lemma 3.7, in $\bar{K}, f$ has a pole of order p at $\mathfrak{a}_1$, so that f^{-1} and f^{-p} have zeros of order p and p^2 respectively at $\mathfrak{a}_1$. Therefore, if w has a pole at $\mathfrak{a}_1$,

$$\mathrm{ord}_{\mathfrak{a}_1} h = \mathrm{ord}_{\mathfrak{a}_1} f^{-1} w^{(s+1)p} + f^{-p} = ps \, \mathrm{ord}_{\mathfrak{a}_1} w < 0.$$

If w is a unit at $\mathfrak{a}_1$, then

$$\mathrm{ord}_{\mathfrak{a}_1} h = \mathrm{ord}_{\mathfrak{a}_1} f^{-1} w^{p(s+1)} + f^{-p} = -\mathrm{ord}_{\mathfrak{a}_1} f = p.$$

If w has a zero at $\mathfrak{a}_1$, then

$$\mathrm{ord}_{\mathfrak{a}_1} h = \mathrm{ord}_{\mathfrak{a}_1} f^{-1} w^{p(s+1)} + f^{-p} = -p \, \mathrm{ord}_{\mathfrak{a}_1} f = p^2.$$

Thus, at $\mathfrak{a}_1$, h either has a pole or a zero of degree at most p^2. Now consider $h^{-k} + z$. Since at $\mathfrak{a}_1$, z has a pole of order greater than $p^2 k$, $\mathrm{ord}_{\mathfrak{a}_1}(h^{-k} + z) = \mathrm{ord}_{\mathfrak{a}_1} z = -(p^l + 1)$. Therefore, by Lemma 3.1, $\mathfrak{a}_1$ will ramify completely in the extension $\bar{K}(\beta_w)/\bar{K}$. Hence, this extension is of degree p. Since at least one prime is ramified completely and the extension is separable, the constant field of $\bar{K}(\beta_w)$ is the same as the constant field of $\bar{K}$. Thus α is of degree p over $\bar{K}(\beta_w)$. Further we remind the reader that if $w \in G$, $h \in \bar{G}_w(z) \subset \bar{G}$. In these fields, $h^{-k} + z$ will have a pole of order not divisible by p at primes below $\mathfrak{a}_1$. Therefore, by Lemma 3.1 these primes in $\bar{G}_w(z)(\beta_w)$ and in $\bar{G}(\beta_w)$ respectively will have factors with ramification degree p. Consequently, the degrees of the corresponding extensions will be equal to p. Finally, α will remain of degree p over $\bar{G}_w(z)(\beta_w)$ and $\bar{G}(\beta_w)$ for the reasons described above.

For future use, in the case $w \in G$, also note that in all of the three fields, any valuation that is a zero of h is also a pole of $(h^{-k} + z)$. Further, the order of $(h^{-k} + z)$ at any such valuation, except for $\mathfrak{a}_1$ and primes below it, is divisible by p if and only if the order of h at this valuation is divisible by p. Thus, if h has a zero at $\mathfrak{t}$ and $\text{ord}_{\mathfrak{t}} h \not\cong 0$ modulo p in $\bar{G}_w$, then $\mathfrak{t}$ ramifies completely in the extensions $\bar{K}(\beta_w)/\bar{K}$, $\bar{G}(\beta_w)/\bar{G}$, $\bar{G}_w(z, \beta_w)/\bar{G}_w(z)$.

We will now proceed to the proof of the lemma.

1) Suppose $w \in K$ has a pole at $\mathfrak{b}$. Then in K,

$$\text{ord}_{\mathfrak{b}} h = \text{ord}_{\mathfrak{b}}(f^{-1} w^{p(s+1)} + f^{-p}) = p(s+1)\text{ord}_{\mathfrak{b}} w - s \not\cong 0 \text{ modulo } p.$$

Further,

$$\text{ord}_{\mathfrak{b}} h < 0.$$

Further, by construction, no factor of $\mathfrak{b}$ ramifies in the extension $\bar{K}/K$. Thus, in $\bar{K}$, for any factor $\mathfrak{g}$ of $\mathfrak{b}$, $\text{ord}_{\mathfrak{g}} h \not\cong 0$ modulo p and h has a pole at all factors of $\mathfrak{b}$.

Next observe the following. Since h has a pole at $\mathfrak{b}$, and z does not have a pole at any factor of $\mathfrak{b}$, $h^{-k} + z$ does not have a pole at any factor of $\mathfrak{b}$, and so, by Lemma 3.1, no factor of $\mathfrak{b}$ ramifies in the extension $\bar{K}(\beta_w)/\bar{K}$. Thus, the order of h at any factor of $\mathfrak{b}$ is not divisible by p in $\bar{K}(\beta_w)$.

Note also that every factor of $\mathfrak{b}$ is relatively prime to the discriminant of β_w. Further, $h^{-k} + z \cong b^p - b$ modulo every factor of $\mathfrak{b}$ in $\bar{K}$ and thus the left side of (3.7) factors completely modulo every factor of $\mathfrak{b}$. Therefore, by [**21**, Proposition 25, page 27, Proposition 16, page 67], every factor of $\mathfrak{b}$ will split completely in the extension $\bar{K}(\beta_w)/\bar{K}$. Since this extension has no constant field subextension, and every factor of $\mathfrak{b}$ is of degree 1 in $\bar{K}$, we must conclude that in $\bar{K}(\beta_w)$ all factors of $\mathfrak{b}$ are also of degree 1.

Since $\bar{K}$ and $\bar{K}(\beta_w)$ have the same constant field, (3.5) still has no solution in $\bar{K}(\beta_w)$ and consequently, (3.5) has no solution modulo any factor of $\mathfrak{b}$ in $\bar{K}(\beta_w)$. Thus, by [**21**, Proposition 25, page 27, Proposition 16, page 67], every factor of $\mathfrak{b}$ in $\bar{K}(\beta_w)$ remains prime in $\bar{K}(\beta_w, \alpha)$. Hence, by Lemma 3.2, (3.12) will have no solution in $\bar{K}(\alpha, \beta_w)$.

2) Suppose now w does not have a pole at $\mathfrak{b}$ and $w \in G$. We will show that in this case (3.13) will have a solution in $\bar{G}_w(z, \alpha, \beta_w)$.

By the Strong Hasse Norm Principal (see [**2**, Page 185] or [**42**, Propositions 10,11, pages 182-183; Theorem 2, page 206]), it is enough to show that for all primes $\mathfrak{t}$ of $\bar{G}_w(z, \beta_w), h$ is a local norm. Note that no prime ramifies in the extension $\bar{G}_w(z, \alpha, \beta_w)/\bar{G}_w(z, \beta_w)$. Thus if h is a unit at $\mathfrak{t}$, it is automatically a local norm at $\mathfrak{t}$ by [**42**, Corollary, page 226]. Suppose $\mathfrak{t}$ is a pole of h. Then either it is a factor of $\mathfrak{B}_w$ or it is a pole of w. Since w has no pole at $\mathfrak{B}_w$, direct calculation assures us that h will have a pole at every factor of $\mathfrak{B}$ of order divisible by p. On the other hand, if $\mathfrak{t}$ is a pole of w, then again by direct calculation one can see that h will also have a pole at $\mathfrak{t}$ of order divisible by p. Indeed, the only case which has to be considered with some care is the case of $\mathfrak{t}$ being a pole of f or a zero of f^{-1}. In this case,

$$\begin{aligned}\mathrm{ord}_{\mathfrak{t}} h &= \mathrm{ord}_{\mathfrak{t}} \left(f^{-1} w^{p(s+1)} + f^{-p}\right)\\ &= \min\left(\mathrm{ord}_{\mathfrak{t}}\left(f^{-1} w^{p(s+1)}\right),\ \mathrm{ord}_{\mathfrak{t}} f^{-p}\right)\\ &= \mathrm{ord}_{\mathfrak{t}}\left(f^{-1} w^{p(s+1)}\right).\end{aligned}$$

We should note here that by Lemma 3.7, $\mathfrak{t}$ is ramified over $G_{w,f,v,t,x}$ with ramification degree divisible by p. On the other hand, $f \in G_{w,f,v,t,x}$. Thus, $\mathrm{ord}_{\mathfrak{t}}(f^{-1}) \cong 0$ modulo p. Hence, $\mathrm{ord}_{\mathfrak{t}} h \cong 0$ modulo p.

Assume now that $\mathfrak{t}$ is a zero of h. If $\mathfrak{t}$ is a factor of $\mathfrak{A}_w$ then it is ramified with ramification degree divisible by p over $G_{x,t,v,w,f}$ and since $h \in G_{x,t,v,w,f}$, we can conclude that h has a zero of order divisible by p at $\mathfrak{t}$. If $\mathfrak{t}$ is not a factor of $\mathfrak{A}_w$, then it is ramified with ramification degree divisible by p over $\bar{G}_w(z)$ and again we conclude that h has a zero of order divisible by p at $\mathfrak{t}$. Thus, in all the cases cited above, by Lemma 3.3, h is a local norm at $\mathfrak{t}$.

Theorem 3.9. *Let $\mathfrak{a}$ and $\mathfrak{b}$ be primes of K satisfying conditions described in Lemma* 3.7. *Then the set $INT(\mathfrak{b})$ is Diophantine over K.*

4. Integrality at one prime: The case of $q \neq p$.

In this section we will show that in the case C has an extension of degree $q \neq p$, the set of elements of G integral at a prime is Diophantine over G. Most of the work necessary to prove this proposition has been done in [**39**], but we will need to take care of some details. In this section we will assume $q \neq p$.

Lemma 4.1. *Let L be an algebraic function field, let $a \in L$. Let q be a rational prime distinct from the characteristic of the field. Then a prime $\mathfrak{t}$ of K ramifies in the extension $K(a^{1/q})/K$ if and only if $\mathrm{ord}_{\mathfrak{t}} a \not\cong 0$ modulo q.*

Proof. If $\mathrm{ord}_{\mathfrak{t}} a \not\cong 0$ modulo q then $\mathfrak{t}$ will clearly ramify in the extension. Suppose now $\mathrm{ord}_{\mathfrak{t}} a \cong 0$ modulo q. Since we can multiply or divide a by the qth power of some local uniformizing parameter without changing the extension, without loss of generality we can assume that $\mathrm{ord}_{\mathfrak{t}} a = 0$. But in this case the discriminant of the power basis of $a^{1/q}$ will be a unit at $\mathfrak{t}$, and thus $\mathfrak{t}$ will be unramified.

Lemma 4.2. *Let L be an algebraic function field containing primitive q-th roots of unity. Let $\mathfrak{C}$ and $\mathfrak{B}$ be prime divisors of L. Let $v \in K$ be such that the divisor of v is of the form $\mathfrak{C}^{-1}\mathfrak{U}$, where $\mathfrak{U}$ is a divisor of L which has no common factors with $\mathfrak{C}$ or $\mathfrak{B}$. Further, assume v is equivalent to $b^q \neq 0$ modulo $\mathfrak{B}$, where $b \in C_L$. (Such a v exists by the Weak Approximation theorem.) Let β be a root of*

$$T^q - v = 0. \tag{4.1}$$

Let $R_{\mathfrak{B}}$ be the residue field of $\mathfrak{B}$ in L and assume it is separable over C_L, the constant field of L. Let δ be an element of the algebraic closure of C_L such that $C_L(\delta)$ is isomorphic to the Galois closure of $R_{\mathfrak{B}}$ over C_L. Let $\bar{L} = L(\beta, \delta)$. Then in $\bar{L}$

$$\mathfrak{C} = \prod \mathfrak{c}_i^q,$$

and

$$\mathfrak{B} = \prod \mathfrak{b}_i,$$

where for $i \neq j$, $\mathfrak{b}_i \neq \mathfrak{b}_j, \mathfrak{c}_i \neq \mathfrak{c}_j$, and for all $i, \mathfrak{b}_i$ is of degree 1.

(The proof of this lemma is analogous to the one for Lemma 3.4.)

Lemma 4.3. *Let $f \in G$ have the divisor of the form* (3.4), *but with s not congruent to* 0 *modulo q. Let $v \in G$ be such that it has a pole of order* 1 *at all the primes which are factors of the pole divisor of f and is equivalent to b^q modulo $\mathfrak{B}$ for some $b \neq 0$ in the field of p elements. (Such an $f \in G$ exists by the Strong Approximation Theorem.) Let δ_G, an element of the algebraic closure of C, be a generator of the residue field of $\mathfrak{B}$ over C. Let β be a root of* (4.1) *in the algebraic closure of G. Let $w \in G$, and let $\mathfrak{A}_w$, $\mathfrak{B}_w$, $\mathfrak{U}_w$ be the primes and the divisor below $\mathfrak{A}$ and $\mathfrak{B}$ and $\mathfrak{U}$ respectively in $G_{t,x,v,w,f}$. Let $\bar{G}_w = G_{x,t,v,w,f,\beta}$. Let $\bar{G} = G(\beta, \delta_G)$. Then the following statements are true.*

1) *$\bar{G}_w \subset \bar{G}$.*
2) *$\bar{G}/\bar{G}_w$ is separable (possibly infinite) constant field extension, and thus no prime ramifies in this extension.*
3) *The primes $\mathfrak{A}_w$ and $\mathfrak{B}_w$ are distinct in $\bar{G}_w$.*
4) *In $\bar{G}$, $\mathfrak{B} = \prod \mathfrak{B}_i$, where $\mathfrak{B}_i$ are distinct prime divisors of degree* 1.
5) *In $\bar{G}$, $\mathfrak{A} = \prod \mathfrak{A}_i^q$, where $\mathfrak{A}_i$ are distinct prime divisors.*
6) *In $\bar{G}_w$, $\mathfrak{A}_w = \mathfrak{a}_w^q$, where $\mathfrak{a}_w$ is a prime of $\bar{G}_w$.*

7) *There exist* $z_1 \in \bar{G}$ *such that its divisor is of the form* $\mathfrak{T}\mathfrak{A}_1^{-1}$*, where* $\mathfrak{T}$ *is a divisor of* $\bar{G}$ *relatively prime to* $\mathfrak{B}$ *and* $\mathfrak{A}$*, such that for some* $b \in C, b \neq 0, z_1 \cong b^q$ *modulo* $\mathfrak{B}$.
8) *Let* $k > 0$ *be the highest order of any pole of* z_1 *in* $\bar{G}$*. Then there exists* $z_2 \in \bar{G}$ *such that its divisor is of the form* $\mathfrak{W}\mathfrak{A}_1^{-q^l}$*, where* $\mathfrak{W}$ *is an integral divisor of* $\bar{G}$ *relatively prime to* $\mathfrak{B}$ *and* $\mathfrak{A}$*, and* $q^l > kq^2$*. Further,* $z_2 \cong 1$ *modulo* $\mathfrak{B}$.
9) *Let* $z = z_1 z_2$*. Then in* $\bar{G}$*,* z *has a pole of order* $q^l + 1 > k$ *at* $\mathfrak{A}_1$ *and is equivalent to* b^q *modulo* $\mathfrak{B}$.
10) *In* $\bar{G}_w(z)$*,* z *has a pole of order* $q^l + 1 > q^2 k$ *at the prime below* $\mathfrak{A}_1$ *and is equivalent to* b^q *modulo the divisor below* $\mathfrak{B}$.
11) *In* $\bar{G}$*,* $\mathfrak{U}$ *is a* q*-th power of another divisor.*
12) *In* $\bar{G}_w$*,* $\mathfrak{U}_w$ *is a* q*-th power of another divisor.*

(The proof of this lemma is analogous to the proof of Lemma 3.7.)

Lemma 4.4. *Let* $\mathfrak{A}$, $\mathfrak{B}$, $\bar{G}$, $\bar{G}_w$, $\mathfrak{A}_i$, $\mathfrak{B}_i, f, w, z$ *be as in Lemma 4.3. Let* $a \in \bar{G}$ *be such that the equation*

$$x^q - a = 0 \tag{4.2}$$

has no solution in $\bar{G}$*. As above assume without loss of generality that* $a \in \bar{G}_w$*. Let* h *be defined by* (3.6) *but with* q *in place of* p*. Let* β_w *be a root of the equation*

$$T^q - (h^{-k} + z) = 0. \tag{4.3}$$

Let α *be a root of the equation* (4.2)*. Then the following statements are true.*

1) *If* $w \in G$ *has a pole at* $\mathfrak{B}$*, then the equation*

$$\prod_{i=0}^{q-1}(a_0 + a_1\xi_q^i\alpha + \cdots + a_{q-1}\xi_q^{i(q-1)}\alpha^{q-1}) = h \tag{4.4}$$

has no solution $(a_0, \ldots, a_{q-1}) \in \bar{G}(\beta_w)$*, where* ξ_q *is a* q*-th primitive root of unity.*

2) *If* $w \in G$ *has no pole at* $\mathfrak{B}$*, then Equation* (4.4) *will have a solution* $(a_0, \ldots, a_{q-1}) \in \bar{G}_w(z, \beta_w) \subset \bar{G}(\beta_w)$.

(The proof of this lemma is analogous to the proof of Lemma 3.8.)

Lemma 4.4 is the last part required for the proof of the following theorem.

Theorem 4.5. *Let* G *be an algebraic function field whose constant field* C *is algebraic over a finite field of characteristic* $p > 0$*. Assume further, that* C *is not algebraically closed. Then for any prime* $\mathfrak{B}$ *of* G *the set of all elements of* G *integral with respect to* $\mathfrak{B}$ *is Diophantine over* G.

Proof. If C is not algebraically closed, then it has an extension of degree q, where q is a prime. (This can be easily derived from [**20**, Theorem 13, page

185].) Further, it is not hard to show that this extension will be generated either by an equation of the form (3.5) or (4.1). By Lemma 6.6, any finite extension of C will have also have an extension of degree q. Thus, a as described in the proofs of Lemmas 3.8 and 4.4 exist. Finally, we note that by Lemma 1.3, Equations (3.8) and (4.4) can be rewritten as an equivalent system of equations over G. (By the equivalent system, we mean a system of equations over G such that for every $w \in G$, this system will have solutions in G if and only if (3.8), ((4.4) respectively) has solutions in $\bar{G}(\beta_w)$.)

5. Diophantine Undecidability.

In this section we will summarize the discussion above and describe in more detail classes of fields to which our result is applicable.

Theorem 5.1. *Let K be a recursive field satisfying the assumptions of Notations* 1.4 *with $q = p$. Then Diophantine problem of K is undecidable.*

Proof. The proof of this theorem will follow from Lemma 1.2, Theorem 2.12, Theorem 3.9 assuming we demonstrate existence of primes $\mathfrak{a}$ and $\mathfrak{b}$ as described in the statement of Theorem 3.9. We can let $\mathfrak{a} = \mathfrak{p}$ and $\mathfrak{b} = \mathfrak{q}$, where $\mathfrak{p}$ and $\mathfrak{q}$ are described in Notations 1.4.

Theorem 5.2. *Let K be a recursive field of characteristic $p > 2$. Let C be the algebraic closure of a finite field in K. Assume C has an extension of degree p. Assume further that K has a subfield K_1, possibly equal to C, and an element u transcendental over K_1 such that for some x algebraic over $C(u)$, $K = K_1(u, x)$. Then Diophantine problem of K is undecidable.*

Proof. We can consider K as an algebraic function field over a constant field $K_1 = C_K$. By Theorem 6.11, we know that a finite extension of G contains element t and constants $c_1, \ldots$ as described in Notations 1.4. Further, by Lemma 6.13, in the corresponding finite extension of K, t and $c_1, \ldots$ will also posses the required properties. Thus, by Theorem 5.1, the Diophantine problem of K is undecidable.

6. Appendix.

Notations 6.1. In this section the term "algebraic function field K over a constant field C" we will always mean a finite algebraic extension of a rational function field $C(w)$, where w is transcendental over C and C is algebraically closed in K.

Lemma 6.2. *Let H/L be a finite separable extension of algebraic function fields and let C_H be the constant field of H. Let $\mathfrak{u}$ be an integral divisor of L. Then* $\text{degree}_H(\mathfrak{u}) = [H : C_H L]\text{degree}_L(\mathfrak{u})$.

(See [**1**, Theorem 9, page 279 and Theorem 14, page 282].)

Lemma 6.3. *Let M/H be a Galois extension of algebraic function fields over the same field of constants C, algebraic over a finite field. Let F be an algebraic extension of C. Then MF/HF is a Galois extension whose Galois group is isomorphic to the original one.*

Lemma 6.4. *Let C be a field algebraic over a finite field, and let t be transcendental over C. Let H be a finite separable extension of $C(t)$ generated by $\alpha \in H$. Let C_0 be any subfield of C. Then the extension $H/C_0(\alpha, t)$ is a constant field extension.*

Lemma 6.5. *Let H be an algebraic function field over a perfect field of constants C and let t be a nonconstant element of H. Then the following conditions are equivalent.*

1) *t is not a p-th power in H.*
2) *The extension $H/C(t)$ is finite and separable.*

(See [**22**, page 94].)

Lemma 6.6. *Let C be a field algebraic a over finite field of characteristic $p > 0$. Let q be a rational prime possibly equal to p such that C has an extension of degree q. For any natural number n, let F_n be the finite field of p^n elements. Let $F = \bigcup_i^{\infty} F_{q^i}$. Then there exists a natural number r such that $F \cap C = F_{q^r}$.*

Proof. First of all, we note the following well known facts concerning the finite fields:

$$F_m F_n = F_{lcm(m,n)}; \tag{6.1}$$

F_m is of degree m over the field of p elements and it consists of all the solutions to the equation $x^{p^m} - x = 0$. (See [**20**, pages 184 - 185].) Next let α be an element of the algebraic closure of a finite field such that $[C(\alpha) : C] = q$. Let $a_0, \dots, a_{q-1}$ be the coefficients of the irreducible polynomial of α over C. Let m be the smallest positive integer such that $\alpha^{p^m} = \alpha$. Then from (6.1) we conclude that $m = q[F_1(a_0, \dots, a_{q-1}) : F_1]$. Let $m = q^t k$, where $(k, q) = 1$. Then $F_{q^{t-1}} \subset F_{q^{t-1}k} \subset C$ but $F_{q^t} \not\subset C$. Otherwise, C contains $F_{q^t} F_{m/q} = F_m$ and thus α.

Lemma 6.7. *Let A_1 be a field algebraic over a finite field. Let A_2 be a finite algebraic extension of A_1. Let α be an element of the algebraic closure of A_1 such that for some rational prime $q, (q, [A_1(\alpha) : A_1]) = 1$. Then $(q, [A_2(\alpha) : A_2]) = 1$.*

Proof. Let $F(T) = a_0 + \dots + T^k$ be the monic irreducible polynomial of α over A_2. Then $a_0, \dots, a_{k-1} \in A_1(\alpha)$ since these are symmetric functions of conjugates of α over A_2 which are also conjugates of α over $A_1 \subset A_2$. Thus, $A_3 = A_1(a_0, \dots, a_{k-1}) \subset A_1(\alpha) \cap A_2$. Since $A_3 \subset A_2, [A_2(\alpha) : A_2] \leq [A_3(\alpha) : A_3]$. On the other hand, since $a_0, \dots, a_{k-1} \subset A_3$, $[A_3(\alpha) : A_3] \leq$

$[A_2(\alpha) : A_2]$. Thus, $[A_2(\alpha) : A_2] = [A_3(\alpha) : A_3]$. On the other hand, $A_3(\alpha) \subset A_1(\alpha)$ and $A_1(\alpha) \subset A_3(\alpha)$ so that $A_1(\alpha) = A_3(\alpha)$. Thus,

$$[A_2(\alpha) : A_2] = [A_3(\alpha) : A_3] = [A_1(\alpha) : A_3] = [A_1(\alpha) : A_1]/[A_3 : A_1],$$

and the lemma is true.

Lemma 6.8. *Let H be an algebraic function field whose constant field C_1 is algebraic over a finite field. Let $\mathfrak{p}$ be a prime of H. Let C be an algebraic extension of C_1 such that for any field $C_2 \subset C$ such that C_2/C_1 is a finite extension, $[C_2 : C_1]$ is prime to the degree$_H$ of $\mathfrak{p}$. Then $\mathfrak{p}$ remains prime in CH.*

Proof. Suppose $\mathfrak{p}$ splits in CH, then for some C_2 as described in the statement of the lemma, $\mathfrak{p}$ splits in C_2H. (This is true because in CH, $\mathfrak{p}$ will have at least two factors, and therefore there will be an element α integral at one but not at the other. Hence, $\mathfrak{p}$ will have to split in $C_1(\alpha)H$.) Let $m = [C_2 : C_1]$ and let $\mathfrak{P}$ be a prime above $\mathfrak{p}$ in C_2H. Since C_2/C_1 is a separable extension, by Theorem 14 on page 282 of [**1**], C_2 is the constant field of C_2H. Next consider the following diagram:

$$\begin{array}{ccc} C_2 & \text{———} & R_{\mathfrak{P}} \\ | & & | \\ C_1 & \text{———} & R_{\mathfrak{p}}. \end{array}$$

Here $R_{\mathfrak{p}}$ and $R_{\mathfrak{P}}$ are residue fields of $\mathfrak{p}$ and $\mathfrak{P}$ respectively. Further, from the diagram we can conclude that

$$[R_{\mathfrak{P}} : C_1] = [R_{\mathfrak{P}} : R_{\mathfrak{p}}][R_{\mathfrak{p}} : C_1] = [R_{\mathfrak{P}} : C_2][C_2 : C_1],$$

or, in other words,

$$f(\mathfrak{P}/\mathfrak{p})\text{degree}_{C_1H}(\mathfrak{p}) = \text{degree}_{C_2H}(\mathfrak{P})m.$$

Thus, since $(m, \text{degree}_{C_1H}(\mathfrak{p})) = 1$, we must conclude that $\text{degree}_{C_1H}(\mathfrak{p})$ divides $\text{degree}_{C_2H}(\mathfrak{P})$. Hence, $\text{degree}_{C_2H}(\mathfrak{P})$ is at least as big as $\text{degree}_{C_1H}(\mathfrak{p}) = \text{degree}_{C_2H}(\mathfrak{p}) \geq \text{degree}_{C_2H}(\mathfrak{P})$. (Here we use the fact that degree of a divisor stays the same under separable constant extensions by Lemma 6.2.) Thus, we must conclude that $\text{degree}_{C_2H}(\mathfrak{p}) = \text{degree}(\mathfrak{P})_{C_2H}$ and $\mathfrak{P}$ is the only prime of C_2H above $\mathfrak{p}$.

Our next task is to prove the main technical theorem of this section. The proof of this theorem will be similar to the proof of Theorem 3.6 of [**38**]. The differences will stem from the fact that we have an infinite constant field here (as opposed to the finite constant field in the theorem cited above), and seek primes which are linear polynomials in a certain element t of K. The proof of the theorem relies on two technical lemmas which we state below.

Lemma 6.9. *Let M be a Galois extension of an algebraic function field L over a finite field of constants, let C_L be the constant field of L, let C_M be the constant field of M, let t be a nonconstant element of L. Let $\sigma \in \mathrm{Gal}(M/L)$, and let $\mathcal{C} = \{\tau\sigma\tau^{-1} | \tau \in \mathrm{Gal}(M/L)\}$. Further, let p^r be the size of C_L, let $\phi = \phi_{C_L}$ be the generator of $\mathrm{Gal}(C_M/C_L)$ sending each element $c \in C_M$ to c^{p^r}, and assume that for every $\psi \in \mathcal{C}, \psi_{|C_M} = \phi^a$ for some natural integer a different from zero. Then if $k \cong a$ modulo $[C_M : C_L]$, $m = [M : C_M L]$, $d = [L : C_L(t)]$, and $C_k(M/L, \mathcal{C}) = \{\mathfrak{p} | \mathfrak{p}$ is a prime of L, $\mathrm{degree}(\mathfrak{p}) = k$, $\mathfrak{p}$ is unramified over $C_L(t)$, and for some β above $\mathfrak{p}$ the Frobenius automorphism of β belongs to $\mathcal{C}\}$,*

$$
\begin{aligned}
(6.2) \qquad & \left| C_k(M/L, \mathcal{C}) - \frac{|\mathcal{C}|}{km} p^{rk} \right| \\
& < \frac{|\mathcal{C}|}{km}((m + 2g_M)p^{rk/2} + m(3g_L + 1)p^{kr/4} + 2(g_M + dm)) \\
& < \frac{|\mathcal{C}|}{k}(7g_M + 4d)p^{rk/2},
\end{aligned}
$$

where g_M, g_L are genus' of M and L respectively.

(For the first inequality see [**15**, Proposition 13.4] and [**14**, Lemma 5.7, p. 59]. The second inequality follows from [**3**, Corollary 2, page 106], [**1**, Theorem 22, page 291], and the fact that the extension M/L is separable.)

Lemma 6.10. *Let M be a Galois extension of an algebraic function field L over a finite field of constants, and assume U is an algebraic function field such that $L \subset U \subset M$, and U is not necessarily Galois over L. Let C_M and C_L denote the constant fields of M and L respectively. Further, let $\mathfrak{p}$ be a prime of L which does not split in U. Let $\mathfrak{p}_U$ be the prime above $\mathfrak{p}$ in U, let β be a prime of M above $\mathfrak{p}$, let $G(\beta)$ be the decomposition group of β, and let $\sigma \in G(\beta)$ be such that its coset modulo the inertia group of β induces the Frobenius automorphism $\phi_{R_\mathfrak{p}}$ on the residue field of $\mathfrak{p}$. Then $\sigma^{f(\mathfrak{p}_U/\mathfrak{p})} \in \mathrm{Gal}(M/U)$, and $f(\mathfrak{p}_U/\mathfrak{p}) = [U : L]$ is the smallest positive exponent such that the corresponding power of σ is in $\mathrm{Gal}(M/U)$. Further, $\sigma_{|C_M} = \phi_{C_L}^{\mathrm{degree}(\mathfrak{p})}$, where ϕ_{C_L} is the Frobenius automorphism of C_L.*

Conversely, suppose β is a prime of M not ramified over L. Let $\mathfrak{p}$ be a prime of L below β and let σ be the Frobenius automorphism of β. Assume further that for some coset $\mathrm{Gal}(M/U)\tau$ of $\mathrm{Gal}(M/U)$ in $\mathrm{Gal}(M/L)$, $\mathrm{Gal}(M/U)\tau\sigma^{[U:L]} = \mathrm{Gal}(M/U)\tau$, while this equality does not hold for any smaller exponent. Then $\mathfrak{p}$ does not split in U.

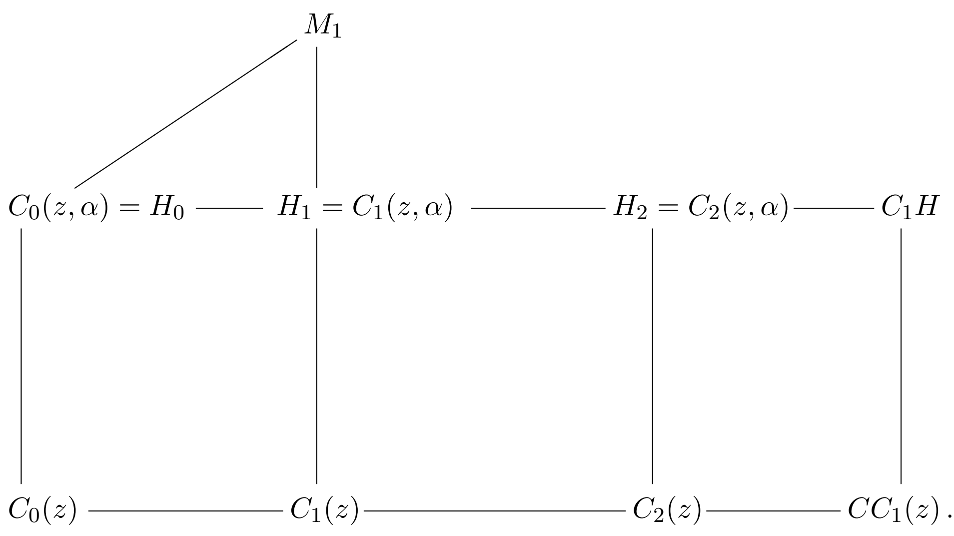
(See [**38**, Lemma 3.3] for part one of the lemma and [**16**, Proposition 2.8, page 101] for part two of the lemma.)

Theorem 6.11. *Let C be an infinite field algebraic over a finite field of characteristic $p > 0$. Assume C has an extension of degree q, where q is* *a rational prime (possibly equal to p). Let H be an algebraic function field whose field of constants is equal to C. Then for any sufficiently large positive integer h, a finite constant extension of H contains a nonconstant element t, infinitely many constants $c_0 = 0, c_1, \dots$, such that for all $i = 0, \dots$, the divisor of $t + c_i$ in H is of the form $\mathfrak{p}_i/\mathfrak{q}$, where $\mathfrak{p}_i$, $\mathfrak{q}$ are primes of H of degree q^h.*

Proof. We will first establish existence of t, and then derive the existence of the required constants. Let z be a nonconstant element of H which is not a p-th power. (Such an element exists by the Weak Approximation Theorem.) Then by Lemma 6.5 the extension $H/C(z)$ is finite and separable and therefore is simple. Thus, for some $\alpha \in H, H = C(z, \alpha)$. Let $C_0 = \bigcup_i^\infty F_{q^i} \cap C$. Let C_1 be the constant field of M_1, the normal closure of $C_0(\alpha, z)$ over $C_0(z)$. Let $H_1 = C_1(\alpha, z)$. Then M_1/H_1 and $M_1/C_1(z)$ are Galois extension and all three fields have the same field of constants.

Let C_2 be a finite extension of C_1 contained in C_1C. Let $H_2 = C_2(z, \alpha)$ and note that H_2/H_1 is a separable constant field extension such that, by Lemma 6.7 and by construction of C_0, its degree is not divisible by q. Indeed, let $\alpha_1, \alpha_2, \dots \in C$ be the generators of C over C_0. Then the degree of α_i over C_0 and consequently, by Lemma 6.7, over $C_1(\alpha_1, \dots, \alpha_{i-1})$ is not divisible by q. Let $\beta \in C_2$, then $\beta \in C_1(\alpha_1, \alpha_2, \dots)$ and consequently the degree of β over C_1 is not divisible by q.

The following diagram describes the extensions involved.



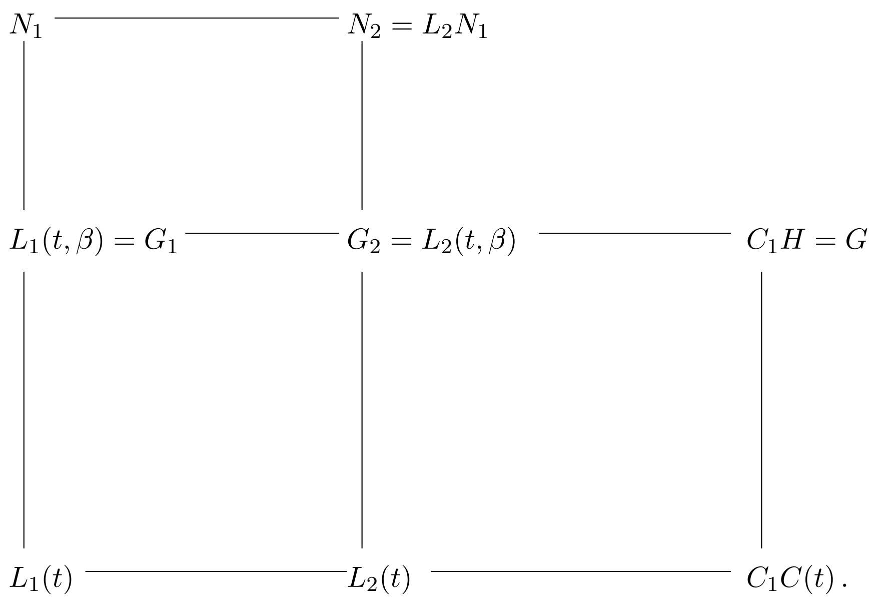
Fix a positive integer h. Let $|C_1| = p^{r_1}$. Then $C_1(z)$ has exactly $\frac{p^{r_1q^h} - p^{r_1q^{h-1}}}{q^h}$ irreducible polynomials of degree q^h. ($p^{r_1q^h} - p^{r_1q^{h-1}}$ is the number of elements of the algebraic closure of C_1 of degree q^h over C_1.

Each of these elements has exactly q^h conjugates.) Let h_{H_1} be the class number of H_1. Then for any sufficiently large h, $C_1(z)$ will contain at least $h_{H_1} + 2$ primes of degree q^h.

Next consider the Galois extension $M_1/C_1(z)$. Let $\mathfrak{t}$ be a prime of $C_1(z)$ of degree q^h. Assume $\mathfrak{t}$ splits completely in M_1. Then, we claim, it splits completely in H_1 and its factors in H_1 are all of degree q^h. Indeed, assume $\mathfrak{t} = \prod_{i=1}^{[H_1:C_1(z)]} \mathfrak{T}_i$ is the factorization of $\mathfrak{t}$ in H_1. For each i, the relative degree of $\mathfrak{T}_i$ over $\mathfrak{t}$ is equal to one. This fact together with the fact that there is no constant field extension from $C_1(z)$ to H_1 implies that $C_1(z)$ degree of $\mathfrak{t}$ must be the same as the H_1 degree of $\mathfrak{T}_i$. Thus, for sufficiently large h, H_1 has at least $h_{H_1} + 2$ primes of degree q^h. Let $\mathfrak{b}_1, \dots, \mathfrak{b}_{h_{H_1}+2}$ be these primes. Next consider the following $h_{H_1} + 1$ H_1-divisors of degree zero: $\mathfrak{b}_2/\mathfrak{b}_1, \dots, \mathfrak{b}_{h_{H_1}}/\mathfrak{b}_1$. At least two of these divisors belong to the same divisor class, and thus for some $1 \le i, j \le h_{H_1} + 2$, $\mathfrak{b}_i/\mathfrak{b}_j$ is a principal divisor. Thus, there exists $t \in H_1$ such that its divisor is of the form $\mathfrak{p}/\mathfrak{q}$, where $\mathfrak{p}, \mathfrak{q}$ are primes of H_1 of degree q^h.

Finally, we note that by Lemma 6.8, divisors $\mathfrak{p}$ and $\mathfrak{q}$ will remain prime in C_1H. Further, since degree of divisors does not change under separable constant field extensions, $\mathfrak{p}$ and $\mathfrak{q}$ will retain their degree. Therefore, C_1H, a finite constant extension of H, will possess the required element t.

We will next address the issue of the existence of the constants $c_1, \dots$ described in the statement of the lemma. To this end let $G = C_1H$ and denote its constant field by $C_G = C_1C$. Note that t is of order 1 at a prime of G and therefore is not a p-th power in G. Thus, the extension $G/C_G(t)$ is separable and finite by Lemma 6.5. Hence, there exists $\beta \in G$ such that $G = C_G(t, \beta)$. Next let L_1 be a finite subfield of C_G such that the following conditions are satisfied: The extension $G/L_1(\beta, t) = G_1$ is an (infinite) constant field extension, the constant field of $L_1(\beta, t)$ is L_1, and $C_1 \subset L_1$. The first condition can be satisfied by any finite field L_1 by Lemma 6.4. Also, by definition of C_1, as in the argument above, the second condition implies that the extension C_G/L_1 contains no finite subextension of degree divisible by q. Note that the prime $\mathfrak{p}_1$ below $\mathfrak{p}$ in G_1 has the same degree in G_1 as $\mathfrak{p}$ in G, by Lemma 6.2. Thus, since there is no constant field extension from $L_1(t)$ to G_1, we can conclude that $[G_1 : L_1(t)]$ is equal to the degree of $\mathfrak{p}_1$: q^h. Let N_1 be the Galois closure of G_1 over $L_1(t)$. Next let $L_2 \subset C_G$ be any finite extension of L_1. Let $G_2 = L_2(t, \beta) = L_2G_1$. Let $N_2 = L_2N_1$. Note that the extensions $N_2/L_2(t, \beta)$ and $N_2/L_2(t)$ are Galois. From the above discussion, it follows that the G_2-divisor of t is of the form $\mathfrak{p}_2/\mathfrak{q}_2$, where $\mathfrak{p}_2$ and $\mathfrak{q}_2$ are G_2-primes of degree q^h. Further, since L_2/L_1 is a separable extension and since G_1 and $L_1(t)$ share the same constant field, G_2 and $L_2(t)$ have the same constant field L_2 and $[G_1 : L_1(t)] = [G_2 : L_2(t)]$ by [**1**, Theorem 11, page 280 and Theorem 14, page 282]. Additionally, $[N_2 : L_2(t)] \le [N_1 : L_1(t)]$, while the genus' of N_2 and G_2 are equal to genus' of N_1 and G_1 respectively by [**1**, Theorem 22, page 291]. The following diagram describes the extensions involved.



Let $\mathfrak{b}_2$ be a factor of $\mathfrak{p}_2$ in N_2. Further, let $\sigma_2 \in \mathrm{Gal}(N_2/L_2(t))$ be an element of the decomposition group $G(\mathfrak{b}_2)$ of $\mathfrak{b}_2$ such that the equivalence class of σ_2 modulo the inertia group of $\mathfrak{b}_2$ is mapped onto the Frobenius automorphism ϕ_{L_2} of L_2 under the canonical homomorphism sending $G(\mathfrak{b}_2)$ to $\mathrm{Gal}(R_{2\mathfrak{b}_2}/L_2)$. Here $R_{2\mathfrak{b}_2}$ is the residue field of $\mathfrak{b}_2$. Then by the first part of Lemma 6.10 we have the following.

1) $\sigma_2^{[G_2:L_2(t)]} \in \mathrm{Gal}(N_2/G_2)$.
2) $[G_2 : L_2(t)]$ is the smallest positive exponent such that the corresponding power of σ_2 is in $\mathrm{Gal}(N_2/G_2)$.
3) σ_2 restricted to the constant field of N_2 is equal to ϕ_{L_2} where ϕ_{L_2} is the Frobenius automorphism of L_2.

Next let $\mathfrak{a}_2$ be a prime of $L_2(t)$ such that σ_2 is the Frobenius automorphism of some N_2-factor $\mathfrak{g}_2$ of $\mathfrak{a}_2$ in N_2. Then by the second part of Lemma 6.10 we can conclude that $\mathfrak{a}_2$ does not split in G_2.

Next we note that in the notations of Lemma 6.9, $a = 1$, and thus, assuming $\mathcal{C}$ is the conjugacy class of σ_2 in $\mathrm{Gal}(N_2/L_2(t))$,

$$|C_1(N_2/L_2(t),\mathcal{C})|$$

$$> \frac{|\mathcal{C}|}{[N_2 : L_2(t)]}|L_2| - |\mathcal{C}|(7g_{N_2} + 4)|L_2|^{1/2} \tag{6.3}$$

$$> |L_2|^{1/2}\left(\frac{1}{[N_1 : L_1(t)]}|L_2|^{1/2} - ([N_1 : L_1(t)](7g_{N_1} + 4))\right). \tag{6.4}$$

Hence, we can conclude that for sufficiently large $|L_2|$, arbitrarily large number of degree one primes of $L_2(t)$ will not split in G_2. For each natural number m, let $N(m)$ be the lower bound on the size of L_2 sufficient for the number of the non-splitting degree one primes to be greater than m. Let $\mathfrak{a}_2$, as above, be such a prime. Then, since there is no constant field extension from $L_2(t)$ to G_2, $[G_2 : L_2(t)] = f(\mathfrak{p}_2/\mathfrak{p}_2 \cap L_2(t)) = q^h = f(\mathfrak{g}_2 \cap G_2/\mathfrak{a}_2)$. Thus, again using the fact that there is no constant field extension from $L_2(t)$ to G_2, we conclude that $\mathfrak{g}_2 \cap G_2$ is of degree q^h in G_2 and will not split in the extension G/G_2.

Finally, we note that, that any degree one prime of $C_2(t)$ which is not a pole of t, is the zero of the element of the form $t + c$, where $c \in L_2$. Thus, keeping in mind that the pole of $t+c$ is the same as pole of t, we can conclude that the divisor of $t + c$ in G will be of the required form.

Lemma 6.12. *Let F/G be a finite separable extension of algebraic function fields. Let $\mathfrak{a}$ be a prime of G which does not split in F, i.e., $\mathfrak{a}$ has only one unramified prime factor $\mathfrak{A}$ in F and $f(\mathfrak{A}/\mathfrak{a}) = [F : G]$. Then there exist $\alpha \in F$ such that $F = G(\alpha)$, α is integral with respect to $\mathfrak{a}$ and such that $\mathfrak{a}$ is not a zero of the discriminant of α.*

Proof. Let $\alpha \in F$ be such that its residue class modulo $\mathfrak{A}$ generates the residue field of $\mathfrak{A}$ over the residue field of $\mathfrak{a}$. (Such an element exists because the residue field of $\mathfrak{A}$ is separable, by assumption, over the residue field of $\mathfrak{a}$.) Then α must be integral with respect to $\mathfrak{A}$ and thus with respect to $\mathfrak{a}$. Further, since the residue class of α is of degree $[F : G]$ over the residue field of $\mathfrak{a}$, $F = G(\alpha)$. Finally, since the residue class of α generate the residue field of $\mathfrak{A}$ over the residue field of $\mathfrak{a}$, $\mathfrak{a}$ cannot be a zero of the discriminant of α. (Otherwise the irreducible polynomial of α modulo $\mathfrak{a}$ will have multiple roots. This is impossible since by assumption the residue field extension is separable.)

Lemma 6.13. *Let H be an algebraic function field over a field of constants C_H. Let K be a constant field extension of H. Let C_K be the constant field of K and assume H is algebraically closed in K. Let $t \in H \setminus C_H$ be such that $H/C_H(t)$ is separable. Let $\mathfrak{a}$ be a prime of $C_H(t)$ remaining prime in*

H and such that its residue field is separable over C_H. Then $\mathfrak{a}$ will have just one prime factor in K.

Proof. Without loss of generality assume $\mathfrak{a}$ is not a pole of t and let $P(t)$ be the polynomial in t over C_H corresponding to $\mathfrak{a}$. By Lemma 6.12, there exists $\alpha \in H$ such that $H = C_H(\alpha, t)$, α is integral with respect to $\mathfrak{a}$, and $\mathfrak{a}$ is not a zero of the discriminant of α. Let $G(T)$ be the monic irreducible polynomial of α over $C_H(t)$. Then, given our assumptions on α and $\mathfrak{a}$, by [**21**, Proposition 25, page 27], $G(T)$ does not split modulo $\mathfrak{a}$. Next consider $P(t)$ over $C_K(t)$. Since H is algebraically closed in K, C_H is algebraically closed in C_K, and thus $P(t)$ will not factor in $C_K(t)$ by [**1**, Theorem 11, page 280]. Hence, $\mathfrak{a}$ will remain prime in $C_K(t)$. Next we want to show that $G(T)$ will not factor modulo $\mathfrak{a}$ over $C_K(t)$. First of all, observe that since $P(t)$ is separable over C_H, the residue field of $\mathfrak{a}$ as a prime of $C_H(t)$ is algebraically closed in the residue field of $\mathfrak{a}$ as a prime $C_K(t)$ by [**1**, Theorem 13, page 281]. Let $\bar{G}(T)$ is the image of $G(T)$ modulo $\mathfrak{a}$. By assumption, $\bar{G}(T)$ is irreducible over the residue field of $\mathfrak{a}$ as a prime of $C_H(t)$. Finally, since the residue field of $\mathfrak{a}$ as a prime of $C_H(t)$ is algebraically closed in the residue field of $\mathfrak{a}$ as a prime $C_K(t)$, again by [**1**, Theorem 11, page 280], $\bar{G}(T)$ will remain prime over the residue field of $\mathfrak{a}$ in $C_K(t)$. Since $K = C_K H = C_K(t, \alpha)$, we can use [**21**, Proposition 25, page 27] to conclude that $\mathfrak{a}$ will remain prime in K.

Lemma 6.14. *Let K be an algebraic function field over a field of constants C_K. Let $\mathfrak{t}$ be a prime of K. Let $R_{\mathfrak{t}}$ be the residue field of $\mathfrak{t}$ isomorphic to a finite extension of $C_{\mathfrak{t}}$ of C_K. Assume that $C_{\mathfrak{t}}$ is separable over C_K. Let C_{Gal} be the Galois closure of $C_{\mathfrak{t}}$ over C_K. Then in the extension $C_{\text{Gal}}K/K$, $\mathfrak{t}$ will split into degree 1 factors. Further, the same statement will apply to any separable constant field extension of $C_{\text{Gal}}K$.*

Proof. Let $\alpha \in K$ be such that the residue class of α modulo $\mathfrak{t}$ generates $C_{\mathfrak{t}}$ over C_K. Let $F(T) \in C_K[T]$ be the monic irreducible polynomial of the residue class of α over C_K. By assumption $F(T)$ is a separable polynomial. Let C_{Gal} be the splitting field of F. Let $a_1, \ldots, a_m$ be all the distinct roots of $F(T)$ in the algebraic closure of C_K. Since $F(T)$ does not factor over K (otherwise some symmetric function of a subset of $a_1, \ldots, a_m$ would be in $K \setminus C_K$), $a_1, \ldots, a_m$ are conjugates over C_K and K. Next note that $\mathbf{N}_{K(a_i)/K}(\alpha - a_i) = F(\alpha) \cong 0$ modulo $\mathfrak{t}$. Thus, for each i, $(\alpha - a_i)$ has a zero at a factor of $\mathfrak{t}$ in $K(a_i)$ and $K(a_1, \ldots, a_m)$. Further, $\alpha - a_i$ and $\alpha - a_j$ have no common zeros for $i \neq j$ because these elements differ by a nonzero constant. Hence, $\mathfrak{t}$ has at least degree($F(T)$) factors in $K(a_1, \ldots, a_m)$. On the other hand, degree of $\mathfrak{t}$ over K is equal to the degree of $F(T)$ and this degree remains the same in $K(a_1, \ldots, a_m)$- a separable constant field extension of K. Thus all the factors of $\mathfrak{t}$ in $C_{\text{Gal}}K$ are of degree 1. Finally, under any separable constant field extension of $C_{\mathrm{Gal}}K$ all the divisors including factors of $\mathfrak{t}$ will retain their degree.

Lemma 6.15. *Let $\{\mathfrak{q}, \mathfrak{p}_1, \ldots, \mathfrak{p}_m\}$ be a set of primes of K. Let $\{b_1, \ldots, b_m\}$ be a set of elements of K such that for each $i = 1, \ldots, m, b_i$ is integral at $\mathfrak{p}_i$. Let $\{n_1, \ldots, n_m\}$ be a set of natural numbers. Then there exists $y \in K$ satisfying the following requirements:*

1) $\mathrm{ord}_{\mathfrak{q}} y = -p^l$, *for some* $l \in \mathbb{N}$;
2) *y is integral at all the other primes of K;*
3) $\mathrm{ord}_{\mathfrak{p}_i}(y - b_i) \geq n_i$.

Proof. By the Strong Approximation Theorem ([**14**, page 21, Proposition 2.11]), there exists $y_1 \in K$ such that $\mathrm{ord}_{\mathfrak{p}_i}(y_1 - b_i) \geq n_i$, y_1 has a pole at $\mathfrak{q}$ and is integral at all the other primes. By a corollary of the Riemann-Roch Theorem, for any sufficiently large l, which we can assume to be greater than $\log_p(\mathrm{ord}_{\mathfrak{q}} y_1)$, there exists $y_2 \in K$ with a sole pole of order p^l at $\mathfrak{q}$, and for each $i = 1, \ldots, m$, with a zero of order greater than $\mathrm{ord}_{\mathfrak{p}_i}(y_1 - b_i)$ at $\mathfrak{p}_i$. Next consider $y = y_1 + y_2$. Note that y will have the pole of the required order and $\mathrm{ord}_{\mathfrak{p}_i}(y - b_i) = \mathrm{ord}_{\mathfrak{p}_i}(y_1 + y_2 - b_i) \geq n_i$.

Lemma 6.16. *Let K be an algebraic function field over a field of constants C_K. Let t be a nonconstant element of K such that the extension $K/C_K(t)$ is finite and separable. Let $\tilde{C}_K$ be the algebraic closure of C_K and let $\tilde{K} = \tilde{C}_K K$. Then the extension $\tilde{K}/\tilde{C}_K(t)$ is separable. Further, let $\tilde{\mathfrak{T}}$ be a prime of $\tilde{C}_K(t)$ with ramified factors in $\tilde{K}$. Let $\mathfrak{T}$ be the prime below $\tilde{\mathfrak{T}}$ in $C_K(t)$. Then $\mathfrak{T}$ has ramified factors in K.*

Proof. Since $K/C_K(t)$ is a finite and separable extension, this extension is simple. Let α be a generator. Then the monic irreducible polynomial of α over $C_k(t)$ has no multiple roots. On the other hand, α will also generate $\tilde{K}$ over $\tilde{C}_K(t)$, and hence $\tilde{K}/\tilde{C}_K(t)$ is separable.

Next let $\hat{C}_K$ be the inseparable closure of C_K and let $\hat{K} = \hat{C}_K K$. Then the extension $\tilde{C}_K/\hat{C}_K$ is separable. Further, since $K/C_K(t)$ is separable, $n = [\hat{K} : \hat{C}_K(t)] = [K : C_K(t)]$. Assume $\mathfrak{T}$ has no ramified factors in K. Let $\{\omega_1, \ldots, \omega_n\}$ be an integral basis with respect to $\mathfrak{T}$. Then by [**3**, Lemma 2, page 71], $\mathfrak{T}$ is neither a zero nor a pole of the discriminant of this basis. But $\{\omega_i\}_{i=1,\ldots,n}$ is also a basis of $\hat{K}/\hat{C}_K(t)$. Thus, by the above cited lemma, no factor of $\mathfrak{T}$ in $\hat{C}_K(t)$ has ramified factors in $\hat{K}$. Finally consider the extension tower $\tilde{K} - \hat{K} - \hat{C}_K(t)$. Since the extension $\tilde{K}/\hat{K}$ is a separable constant field extension, no primes ramify. Thus, $\hat{\mathfrak{T}}$, any factor of $\mathfrak{T}$ in $\hat{C}_K(t)$, has no ramified factors in $\tilde{K}$. Finally, we note that the extension $\tilde{K}/\tilde{C}_K(t)$ is a subextension of $\tilde{K}/\hat{C}_K(t)$, and thus $\tilde{\mathfrak{T}}$ has no ramified factors in $\tilde{K}$.

Lemma 6.17. *Let $\tilde{K}$ be an algebraic function field over an algebraically closed field of constants $\tilde{C}_K$. Let t be transcendental over $\tilde{C}_K$. Let $\tilde{K}$ be a separable extension of $\tilde{C}(t)$. Let $\mathfrak{t}$ be a prime of $\tilde{K}$ not ramifying in the extension $\tilde{K}/\tilde{C}(t)$ and not a pole of t. Let $x \in \tilde{K}$. Then $\text{ord}_{\mathfrak{t}}\partial x/\partial \mathfrak{t} = \text{ord}_{\mathfrak{t}} dx/dt$.*

Proof. By [**22**, page 96], $\text{ord}_{\mathfrak{t}}\partial x/\partial \mathfrak{t} = \text{ord}_{\mathfrak{t}}\partial t/\partial \mathfrak{t} + \text{ord}_{\mathfrak{t}} dx/dt$. However, if $\mathfrak{t}$ is not ramified over $\tilde{C}(t)$, for some $a \in \tilde{C}$, $t + a$ is a local uniformizing parameter for $\mathfrak{t}$. Therefore, $\text{ord}_{\mathfrak{t}}\partial t/\partial \mathfrak{t} = 0$.

Lemma 6.18. *Let K be an algebraic function field over the constant field C_K. Let C be the algebraic closure of a finite field in K. Let G be the algebraic closure of $C(u)$ in K, where $u \in K \setminus C_K$ and K is separable over $C_K(u)$. Then the extension $G/C(u)$ is finite.*

Proof. First of all, observe that $C(u)$ is algebraically closed in $C_K(u)$. Next, let $f \in G$. Then by [**1**, Theorem 11, page 280], $[C(f, u) : C(u)] = [C_K(f, u) : C_K(u)] \leq [K : C_K(u)]$. Thus, since G is separable over $C(u)$, the extension $G/C(u)$ must be finite.

References

[1] E. Artin, *Algebraic Numbers and Algebraic Functions*, Gordon and Breach, New York, 1986.

[2] J.W.S. Cassels and A. Frölich, *Algebraic Number Theory*, Thompson Book Co, Washington, D.C., 1967.

[3] C. Chevalley, *Introduction to the theory of algebraic functions of one variable*, Mathematical Surveys, **6**, AMS, 1951.

[4] J.-L. Colliot-Thélène, A.N. Skorobogatov and P. Swinnerton-Dyer, *Double Fibres and Double Covers: Paucity of Rational Points*, preprint.

[5] M. Davis, *Hilbert's tenth problem is unsolvable*, American Mathematical Monthly, **80** (1973), 233-269.

[6] M. Davis, Yu. Matijasevich and J. Robinson, *Positive Aspects of a Negative Solution*, Proc. Sympos. Pure Math., AMS, Providence, RI, 1976

[7] J. Denef, *Hilbert's tenth problem for quadratic rings*, Proc. Amer. Math. Soc., **48** (1975), 214-220.

[8] ———, *Diophantine sets over* $\mathbb{Z}[T]$, Proc. Amer. Math. Soc., **69** (1978), 148-150.

[9] ———, *The Diophantine problem for polynomial rings and fields of rational functions*, Trans. Amer. Math. Soc., **242** (1978), 391-399.

[10] ———, *The Diophantine Problem for Polynomial Rings of Positive characteristic*, Logic Colloquium 78, M. Boffa, D. van Dalen, K. MacAloon (eds.), North Holland, (1979), 131-145.

[11] ———, *Diophantine sets of algebraic integers*, II, Trans. of Amer. Math. Soc., **257**(**1**) (1980), 227-236.

[12] J. Denef and L. Lipshitz, *Diophantine sets over some rings of algebraic integers*, J. London Math. Soc., **18**(**2**) (1978), 385-391.

[13] M. Eichler, *Introduction to Algebraic Numbers and Algebraic Functions*, Academic Press, New York, 1966.

[14] M. Fried and M. Jarden, *Field Arithmetic*, Springer-Verlag, New York, 1986.

[15] W.-D. Geyer and M. Jarden, *Bounded Realization of l-groups over Global Fields*, to appear in Nagoya Mathematical Journal.

[16] G. Januz, *Algebraic Number Fields*, Academic Press, New York, 1973.

[17] K.H. Kim and F.W. Roush, *Diophantine unsolvability for function fields over certain infinite fields of characteristic p*, Journal of Algebra, **152(1)** (1992), 230-239.

[18] ______, *Diophantine undecidability of $\mathbb{C}(t_1, t_2)$*, Journal of Algebra, **150(1)** (1992), 35-44.

[19] ______, *Diophantine unsolvability over p-adic function fields*, Journal of Algebra, **176** (1995), 83-110.

[20] S. Lang, *Algebra*, Addison Wesley, Reading, MA, 1971.

[21] ______, *Algebraic Number Theory*, Addison-Wesley, Reading, MA, 1970.

[22] R. Mason, *Diophantine Equations over Function Fields*, London Mathematical Society Lecture Notes Series, 96, Cambridge University Press, Cambridge, 1984.

[23] B. Mazur, *The topology of rational points*, Experimental Mathematics, **1(1)** (1992), 35-45.

[24] ______, *Questions of decidability and undecidability in number theory*, Journal of Symbolic Logic, **59(2)** (June 1994), 353-371.

[25] T. Pheidas, *Hilbert's tenth problem for a class of rings of algebraic integers*, Proc. of Amer. Math. Soc., **104(2)** (1988).

[26] ______, *Hilbert's tenth problem for fields of rational functions over finite fields*, Inventiones Mathematicae, **103** (1991), 1-8.

[27] ______, *The diophantine theory of a ring of analytic functions*, Journal für die reine und angewandte Mathematik, **463** (1995), 153-167.

[28] ______, *An undecidability result for power series rings of positive characteristic*, II, Proceedings of American Mathematical Society, **100** (1987), 526-530.

[29] H. N. Shapiro and A. Shlapentokh, *Diophantine relations between algebraic number fields*, Communications on Pure and Applied Math., **XLII** (1989), 1113-1122.

[30] A. Shlapentokh, *Extension of Hilbert's tenth problem to some algebraic number fields*, Communications on Pure and Applied Math., **XLII** (1989), 939-962.

[31] ______, *Hilbert's tenth problem for rings of algebraic functions of characteristic zero*, Journal of Number Theory, **40(2)** (1992), 218-236.

[32] ______, *Hilbert's tenth problem for rings of algebraic functions in one variable over fields of constants of positive characteristic*, Transaction of AMS, **333(1)** (1992), 275-298.

[33] ______, *A Diophantine definition of rational integers in some rings of algebraic numbers*, Notre Dame Journal of Formal Logic, **33(3)** (1992), 299-321.

[34] ______, *Diophantine relations between rings of S-integers of fields of algebraic functions in one variable over constant fields of positive characteristic*, Journal of Symbolic Logic, **58(1)** (March 1993), 158-192.

[35] ______, *Diophantine classes of holomorphy rings of global fields*, Journal of Algebra, **169(1)** (October 1, 1994), 139-175.

[36] ______, *Diophantine undecidability for some holomorphy rings of algebraic function fields of characteristic* 0, Communications in Algebra, **22(11)** (1994), 4379-4404.

[37] ______, *Diophantine undecidability in some rings of algebraic numbers of totally real infinite extensions of* $\mathbb{Q}$, Annals of Pure and Applied Logic, **68** (1994), 299-325.

[38] ______, *Diophantine undecidability over algebraic function fields over finite fields of constants*, Journal of Number Theory, **58(2)** (June 1996), 317-342.

[39] ______, *The logic of pseudo S-integers*, Israel Journal of Mathematics, **101** (1997), 229-254.

[40] ______, *Diophantine definability over some rings of algebraic numbers with infinite number of primes allowed in the denominator*, Inventiones Mathematicae, **129** (1997), 489-507.

[41] C. Videla, *Hilbert's tenth problem for rational function fields in characteristic* 2, Proceedings of American Math. Society, **120(1)**, (January 1994), 249-253.

[42] A. Weil, *Basic Number Theory*, Springer Verlag, New York, 1974.

Received September 23, 1998 and revised January 15, 1999. The research for this paper has been partially supported by NSA grants MDA904-96-1-0019 and MDA904-98-1-0510.

East Carolina University
Greenville, NC 27858
E-mail address: shlapentokh@math.ecu.edu

PACIFIC JOURNAL OF MATHEMATICS
Vol. 193, No. 2, 2000

GENERALIZED WRONSKIANS AND WEIERSTRASS WEIGHTS

Christopher Towse

Given a point P on a smooth projective curve C of genus g, one can determine the Weierstrass weight of that point by looking at a certain Wronskian. In practice, this computation is difficult to do for large genus. We introduce a natural generalization of the Wronskian matrix, which depends on a sequence of integers $s = m_0, \ldots, m_{g-1}$ and show that the determinant of our matrix is nonzero at P if and only if s is the non-gap sequence at P.

As an application, we compute the weights of certain points on the F_9 and F_{10}, the 9th and 10th Fermat curves. These weights correspond to the expected weights predicted in an earlier paper.

Introduction.

One of the fundamental facts about Weierstrass points (and generalizations of Weierstrass points) is that any given curve has only a finite number of them; that the total Weierstrass weight of all the points on a curve is finite. Thus, one of the basic problems in the study of Weierstrass points is, once found, to count the Weierstrass weight of those points. One would at least like to know if one has found all of the Weierstrass points on one's curve of interest.

Fermat curves $F_n : X^n + Y^n + Z^n = 0$ are of particular interest since they have so many automorphisms, and thus (by Lewittes' Theorem [**L**]) so many known Weierstrass points. Hasse [**Ha**] showed that the weight of a point with $XYZ = 0$ is $(n-1)(n-2)(n-3)(n+4)/24$ by demonstrating the existence of certain holomorphic differentials. In [**T**], we exploited the simplicity of order two automorphisms to get a lower bound on the Weierstrass weight of a second class of well-known Weierstrass points on Fermat curves. Further, we showed, with the aid a computer, that the lower bound corresponded to the actual weights for $n \leq 8$. For $n = 8$, this was a new result. It seemed unlikely to be able to repeat this for higher degree (more relevantly, higher genus) Fermat curves. Below, however, we do demonstrate that the lower bound given in [**T**] is exact, for $n = 9$ and 10.

1. Preliminaries.

Throughout, we will consider monotonically increasing sequences of nonnegative integers $s = (m_0, \dots, m_{g-1})$. We define the weight of such a sequence to be $\mathrm{wt}(s) = \mathrm{wt}(m_0, \dots, m_{g-1}) = \sum_{i=0}^{g-1}(m_i - i)$.

The idea of the weight is to measure how much a given sequence differs from a "standard" sequence. In this case, we are merely comparing s to $s_0 = (0, 1, \dots, g-1)$.

Let C be a smooth, projective curve of genus g defined over $K = \mathbb{C}$ or any other algebraically closed, characteristic zero field. Let P be a point on C.

Definition. Given any k-dimensional K-vector space, A, of holomorphic differentials on C, we say a basis for A, $\omega_0, \dots, \omega_{k-1}$, is *adapted to* P if $0 \le \mathrm{ord}_P\,\omega_0 < \mathrm{ord}_P\,\omega_1 < \cdots < \mathrm{ord}_P\,\omega_{k-1}$.

It is well-known (see [**F-K**]) that any such A has a basis adapted to P, for any $P \in C$. It is easy to see that the numbers $\mathrm{ord}_P\,\omega_i$ do not depend on the choice of basis, as long as the basis is adapted for P.

Definition. If we let A be the space of *all* holomorphic differentials on C, and we let $n_i = \mathrm{ord}_P\,\omega_i$, then the sequence of (monotonically increasing, nonnegative) integers $n_0, \dots, n_{g-1}$ is called the *nongap sequence of* P.

Definition. We define the (Weierstrass) weight of P to be the weight of the nongap sequence at P.

Note that this differs from the "classical" nongap sequence defined using orders of poles of functions at P. However, the ideas are related via the Riemann-Roch Theorem, and the Weierstrass weight of a point is the same, using either approach.

2. Main Results.

Let $\{\omega_0, \dots, \omega_{g-1}\}$ be a basis of holomorphic differentials on C. Let x be a local uniformizing parameter on some Zariski open set of C containing P. Then we can write each differential as $\omega_j = f_j\,dx$ for some function f_j. Let $\mathcal{F} = \{f_0, \dots, f_{g-1}\}$.

A fundamental fact (due to Hurwitz [**Hu**]) is that the weight of a point P is equal to the order of vanishing at P of the determinant of the Wronskian matrix whose first row is $\mathcal{F}$. First, we relate the weight of P to the weights of various sequences s.

Definition. Suppose we have chosen any fixed set of functions $\Phi = \{\phi_j(x)\}_{j=0}^{k-1}$. We define $M_\Phi[m_0, \dots, m_{k-1}] = M[m_0, \dots, m_{k-1}] = M[s]$ to be the matrix whose ith row ($i = 0, \dots k-1$) is $(\phi_0^{(m_i)}, \dots, \phi_{k-1}^{(m_i)})$ where $\phi^{(m)}$ denotes the mth derivative with respect to x.

So $M[0,\dots,k-1]$ is a Wronskian matrix with first row $[\phi_0,\dots,\phi_{k-1}]$. We will abbreviate this as W_Φ.

Proposition 1. *The dth derivative of the Wronskian determinant is equal to a sum of determinants of matrices $M[s]$ where s has weight d.*

Proof. We use induction on d. For the case $d = 0$, we note that the only sequence of weight zero is $s_0 = (0, 1, \dots, k-1)$. As noted above, $M[s_0]$ is the Wronskian itself.

For general d, we look at the derivative of the determinant of one of the matrices $M[s]$. Let D denote differentiation with respect to x. Using linearity of the determinant in the rows of M, we see that

$$D(\det M[m_0,\dots,m_{k-1}]) = \det M[m_0+1, m_1,\dots,m_{k-1}] + \cdots + \det M[m_0, m_1,\dots,m_{k-1}+1].$$

If the old sequence $(m_0,\dots,m_{k-1})$ has weight d, then any of the new sequences $(m_0,\dots,m_i+1,\dots,m_{k-1})$ has weight $d+1$. □

It should be noted that not all of such sequences will be strictly monotonic, but for any of those sequences, s, $M[s]$ will have a repeated row, and will therefore have determinant identically zero. Note also that the same $M[s]$ may appear more than once in the sum. For instance, with $g = 4$ and $k = 3$, we get

$$D^{(3)} \det W = \det M[0,2,3,4] + 2\det M[0,1,3,5] + \det M[0,1,2,6].$$

The point of the proposition is the observation that the sequences appearing in our $M[s]$ notation are related to the weight of a point. The following theorem elaborates on this connection.

Theorem 2. *Let $s = (m_0,\dots,m_{g-1})$ be a monotonically increasing sequence of nonnegative integers. Let P be a point of C and let $\mathcal{F}$ be as above. Suppose* $\mathrm{wt}(s) \le \mathrm{wt}(P)$. *Then* $\det M_{\mathcal{F}}[s](P) \neq 0$ *if and only if s is the nongap sequence of P.*

That is, the determinant of the matrix $M = M[m_0,\dots,m_{g-1}]$, evaluated at the point P, is zero for all sequences of weight less than or equal to the weight at P, except for the actual nongap sequence of P.

Proof. It is clear that changing the basis $\mathcal{F}$ will only change the determinant by a nonzero multiple. So we may assume that $\omega_0,\dots,\omega_{g-1}$ is a basis for the space of holomorphic differentials of C adapted to P. Let $(n_0,\dots,n_{g-1})$ be the nongap sequence at P.

Since x is a local uniformizing parameter at P, P is not in the support of the divisor (dx). In other words, $\mathrm{ord}_P\, dx = 0$. Thus, $\mathrm{ord}_P\, f_i = n_i$. So $D^{(k)}(f_i)(P) = 0$ for $k < n_i$ and $D^{(n_i)}(f_i)(P) \neq 0$.

In particular, if we consider the case when $m_i = n_i$ for all i, we see that M is lower triangular, with nonzero entries on the diagonal. Thus, $\det M \neq 0$.

Now, suppose $m_i \neq n_i$ for some i. Since $\mathrm{wt}(m_0, \dots, m_{g-1}) \leq \mathrm{wt}(n_0, \dots, n_{g-1})$, we know there is some $m_i < n_i$. Let I be the smallest index with $m_I < n_I$. Then the first I rows of M can only have nonzero entries in (at most) the first $I-1$ places. They are all contained in the $(I-1)$-dimensional K-vector space $K^{I-1} \times \{0\}^{g-I+1}$. Thus, the first I rows of M are linearly dependent. We conclude that $\det M = 0$. □

Let $\Phi = \{\phi_0, \dots, \phi_{k-1}\}$. As above, W_Φ is the corresponding Wronskian matrix. A basic property of Wronskians is this: If $\mathcal{G} = \{h\phi_0, \dots, h\phi_{k-1}\}$ then

$$\det W_{\mathcal{G}} = h^k \det W_\Phi.$$

This equality is not true if we replace the Wronskians by general matrices of the form $M[s]$. However, we do have the following result.

Proposition 3. *Let $\mathcal{F}$ be a set of functions $\{f_1, \dots, f_g\}$ so that the differentials $\omega_i = f_i\,dx$ form a basis for the space of holomorphic differentials of C. Let $h = h(x)$ be another function and let $\mathcal{G} = \{hf_0, \dots, hf_{g-1}\}$. Let $n_0, \dots, n_{g-1}$ be the nongap sequence at P. Let $s = (m_0, \dots, m_{g-1})$ be a sequence with $m_i \leq n_i$ for all i. Then*

$$\det M_{\mathcal{G}}[s](P) = h^g \det M_{\mathcal{F}}[s](P).$$

Proof. The ith row of $M_{\mathcal{G}}$ is

$$\begin{aligned} &\left[(hf_0)^{(m_i)} \quad \cdots \quad (hf_{g-1})^{(m_i)}\right] \\ &= \left[\sum_{r=0}^{m_i} \binom{r}{m_i} h^{(r)} f_0^{(m_i-r)} \quad \cdots \quad \sum_{r=0}^{m_i} \binom{r}{m_i} h^{(r)} f_{g-1}^{(m_i-r)}\right]. \end{aligned}$$

Expanding the determinant of $M_{\mathcal{G}}[s]$ by linearity in the rows, we get a sum of determinants of matrices with ith row equal to

$$\left[h^{(r)} f_0^{(m_i-r)} \quad \cdots \quad h^{(r)} f_{g-1}^{(m_i-r)}\right]$$

where $0 \leq r \leq m_i$. We can pull out $h^{(r)}$ from these rows.

Let us consider one of the matrices whose determinant is in our sum. Its ith row is of the form

$$\left[f_0^{(t_i)} \quad \cdots \quad f_{g-1}^{(t_i)}\right]$$

where $t_i \leq m_i \leq n_i$. So this matrix is just $M_{\mathcal{F}}[\tilde{s}]$, where $\tilde{s} = (t_0, \dots, t_{g-1})$. Plugging in the point P, we know that $\det M_{\mathcal{F}}[\tilde{s}](P) = 0$ unless $t_i = n_i$ for all i, by the theorem.

We have two cases. First, if $m_i < n_i$ for some i, then we see that all the determinants in our sum vanish at P, by the theorem. And we see that $\det M_{\mathcal{F}}[s](P) = 0$ by the theorem, as well.

If, on the other hand, s was the nongap sequence of P to begin with, then there is exactly one nonvanishing term in our sum. It is the one with $\tilde{s} = s$. That is, $r = 0$ in all cases. So we have factored out $h^{(0)} = h$ from each of the g rows. Also, the factors $\binom{r}{m_i} = \binom{0}{m_i}$ are all 1. We get $\det M_{\mathcal{G}}[s](P) = h^g \det M_{\mathcal{F}}[s](P)$. □

3. Examples - Fermat curves.

Consider the nth Fermat curve $F_n : X^n + Y^n + Z^n = 0$. Dehomogenized at $Z \neq 0$, this is given by $x^n + y^n + 1 = 0$.

Definition. Any point $P \in F_n$ with $XYZ = 0$ will be called a *trivial* Weierstrass point.

Let ζ_n be a primitive nth root of unity. Consider the involutions, T, of F_n given by

$$[X, Y, Z] \mapsto [\zeta_n^j Y, \zeta_n^{-j} X, Z], \quad j = 0, \dots, n-1.$$

We could also have T switch X and Z or Y and Z, of course, for a total of $3n$ involutions.

Definition. Any Weierstrass point $P \in F_n$ fixed by one of the involutions T will be called a *diagonal* Weierstrass point.

We should note that for odd n there are diagonal Weierstrass points which are also trivial Weierstrass points. Since the trivial points are well understood (again, see [**Ha**]), we will ignore them in the following discussion.

We showed in [**T**] that any diagonal point, P, has exactly q odd nongaps and $g - q$ even ones. Here

$$q = \begin{cases} (n-1)(n-3)/4 & n \ \text{odd} \\ (n-2)^2/4 & n \ \text{even} \end{cases}$$

is the genus of the any of the quotient curves $F_n/\langle T \rangle$.

Let

$$s = (0, 1, \dots, 2q-1, 2q, 2q+2, \dots, 2(g-q-1)).$$

The weight of s is the so-called expected weight of P, $\mathrm{wt}_e(P)$. We know from [**T**] that

$$\mathrm{wt}(P) \geq \mathrm{wt}_e(P) = \begin{cases} (n-1)(n-3)/8 & n \ \text{odd} \\ (n-2)(n-4)/8 & n \ \text{even.} \end{cases}$$

Proposition 4. *Let*

$$\mathcal{H} = \{x^i y^j : 0 \leq i \leq n-3-j, 1 \leq j \leq n-3\},$$

$P = ((-1/2)^{(1/n)}, (-1/2)^{(1/n)})$, *and* $s' = (n-2, n-1, \dots, 2q-1, 2q, 2q+2, \dots, 2(g-q-1))$. *Then* $\det M_{\mathcal{H}}[s'](P) \neq 0$ *if and only if the weight of* P *is equal to the expected weight,* $\mathrm{wt}_e(P)$, *given above.*

Proof. It is easy to check that

$$\{x^i/y^j\,dx : \quad 0 \le i \le j-2 \quad \text{and} \quad j = 2, \dots, n-1\}$$

is a basis for the space of holomorphic differentials on F_n. So we may let $\mathcal{F}$ be the corresponding set of functions of the form x^i/y^j. Let $h = y^{n-1}$. Then $\mathcal{G} = \{x^i y^j : 0 \le i \le n-3-j, 0 \le j \le n-3\}$. Proposition 3 says that $\det M_{\mathcal{G}}[s](P) = h^g \det M_{\mathcal{F}}[s](P)$ for any sequence s of weight less than or equal to the weight of P.

Further, since the first $n-2$ functions of $\mathcal{G}$ are $\{1, x, \dots, x^{n-3}\}$, we see that

$$y^{-g(n-1)} \det M_{\mathcal{F}}[s](P) = \det M_{\mathcal{G}}[s](P) = \alpha \det M_{\mathcal{H}}[s'](P)$$

where $\alpha = \prod_{i=0}^{n-2} i!$. Since $\det M_{\mathcal{F}}[s](P)$ is nonzero if and only if $\mathrm{wt}(P) = \mathrm{wt}(s)$, by the theorem and since $\mathrm{ord}_P\, h = 0$, we are done. □

Proposition 5. *The nontrivial diagonal Weierstrass points on* F_9 *have weight* 6.

Proof. Using the set of functions $\mathcal{H}$ from the corollary, we had Mathematica compute the matrix $M = M_{\mathcal{H}}[7, \dots, 24, 26, 28, 30](P)$.

Factoring out as much as we could from the rows and columns of the matrix, we were able to have Mathematica compute the remaining determinant:

$$\det M = (-1)^{2/3}(2)^{2/9} 2^{438} 3^{129} 5^{67} 7^{44} 11^{25} 13^{18} 17^{23} 19^{9} 23^{6} 29^{2} (53)(967)(2141)\kappa$$

where κ is a composite integer on the order of 10^{85} with no small factors:

$$\kappa = 4972125503975388123549399196928230101413\\ 365278909345657357099864972271058186013692161. \tag{1}$$

This is nonzero and hence we conclude that the Weierstrass weight of the 243 conjugates of P (including itself, but excluding the points with $XYZ = 0$) are all 6, as "expected." Furthermore, we have shown that the nongap sequence at all of these points is $(0, 1, \dots, 24, 26, 28, 30)$. □

We can do a similar computation for F_{10}. The genus is 36, and the expected gap sequence is $0, 1, \dots, 31, 32, 34, 36, 38$.

Proposition 6. *The diagonal Weierstrass points on* F_{10} *have weight* 6.

Proof. Using Mathematica, we compute (with the appropriate set of functions $\mathcal{H}$) the matrix $M = M_{\mathcal{H}}[8, 9, \dots, 31, 32, 34, 36, 38](P)$. Rather than attempting to compute the exact determinant, we factor out powers of $2^{1/10}$ and powers of $(-1)^{1/10}$. This leaves a matrix with integer entries. In order to show that the determinant of this matrix is nonzero, we need only show that it is nonzero modulo p for some integer p. First, we reduce the entries of the matrix modulo p and obtain a matrix of residues. Next, we can we can easily find the exact determinant of this matrix. Lastly, we reduce this determinant modulo p again.

In this case, the determinant is zero modulo p for all primes less than 43. Of course this only means that these primes divide the determinant. The prime 43 is the smallest which does not divide the determinant. Modulo 43, we get a nonzero determinant. (We got 17 modulo 43.) This shows that the actual nongap sequence is the one we "expected." And thus, the weight of each of the 300 points on F_{10} conjugate to $((-1/2)^{1/10}, (-1/2)^{1/10})$, including itself, is 6. □

It is interesting to note that the primes which divide the determinant of M, in this case, include (but are not limited to) all primes less than 43, as well as 59, 79, and 997. No other primes less than 3571 (the 500th prime) divide the determinant. (Compare with the determinant of M in Proposition 5.)

In general, one should be able to check whether diagonal points have the minimal possible nongap sequence by calculating $\det M_{\mathcal{H}}[s](P)$.

The difficulty is in computing the higher derivatives in order to create the matrix M. Further, one cannot numerically calculate the determinant of M without running into round-off error problems. For $n = 9$, we computed this determinant exactly. This seems too difficult and slow a computation for larger n, at this time. For $n = 10$, we reduced modulo some small primes. This works as long as the determinant *is*, in fact, nonzero. It is interesting, however, that the evidence does seem to suggest that the expected gap sequences are the actual ones, in general.

Also, it is intriguing that the primes p for which the diagonal points coalesce with other points on the reduction of F_n modulo p (that is, the primes which divide $\det M$) seem to include most small primes as well as just a few scattered large primes.

References

[A] R. Accola, *Topics in the theory of Riemann surfaces*, **1595**, Springer-Verlag, 1994, Lecture Notes in Mathematics.

[F-K] H. Farkus and I. Kra, *Riemann Surfaces*, Springer-Verlag, 1980.

[Ha] H. Hasse, *Über den algebraischen Funktionenkörper der Fermatschen Gleichung*, in 'Mathematische Abhandlungen', Walter de Gruyter, 1975; originally in Acta Univ. Szeged Sect. Math., **13** (1950), 195-207.

[Hu] A. Hurwitz, *Ueber algebraische Gebilde mit eindeutigen Transformationen in sich*, Mathematiche Annalen, **41** (1893), 403-442.

[L] J. Lewittes, *Automorphisms of compact Riemann surfaces*, Amer. J. Math., **85** (1963), 732-752.

[T] C. Towse, *Weierstrass weights of fixed points of an involution*, Math. Proc. Camb. Phil. Soc., **122**(**3**) (1997), 385-392.

Received August 19, 1998 and revised March 23, 1999.

Department of Mathematics and Statistics
Pomona College
Claremont, CA 91711-6348
E-mail address: ctowse@pomona.edu

CONTENTS

Volume 193, no. 1 and no. 2

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 193 No. 2 April 2000

0030-8730(200004)193:2;1-5