Pacific Journal of Mathematics

Volume 192 No. 1 January 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. 192, No. 1, 2000

GENERALIZED ELLIPTIC INTEGRALS AND MODULAR EQUATIONS

G.D. Anderson, S.-L. Qiu, M.K. Vamanamurthy, and M. Vuorinen

In geometric function theory, generalized elliptic integrals and functions arise from the Schwarz-Christoffel transformation of the upper half-plane onto a parallelogram and are naturally related to Gaussian hypergeometric functions. Certain combinations of these integrals also occur in analytic number theory in the study of Ramanujan's modular equations and approximations to π. The authors study the monotoneity and convexity properties of these quantities and obtain sharp inequalities for them.

1. Introduction.

In 1995 B. Berndt, S. Bhargava, and F. Garvan published an important paper [**BBG**] in which they studied generalized modular equations and gave proofs for numerous statements concerning these equations made by Ramanujan in his unpublished notebooks. No record of Ramanujan's original proofs has remained. A *generalized modular equation with signature* $1/a$ *and order* (*or degree*) p *is*

$$\frac{F(a,1-a;1;1-s^2)}{F(a,1-a;1;s^2)} = p\frac{F(a,1-a;1;1-r^2)}{F(a,1-a;1;r^2)},\ 0<r<1. \tag{1.1}$$

Here F is the Gaussian hypergeometric function defined in (1.2). The word *generalized* alludes to the fact that the parameter $a \in (0,1)$ is arbitrary. In the classical case, $a = \frac{1}{2}$ and p is a positive integer. Modular equations were studied extensively by Ramanujan, see [**BBG**], who also gave numerous algebraic identities for the solutions s of (1.1) for some rational values of a such as $\frac{1}{6}, \frac{1}{4}, \frac{1}{3}$.

Before proceeding, we introduce some necessary notation. Given complex numbers a, b, and c with $c \neq 0, -1, -2, \ldots$, the *Gaussian hypergeometric function* is the analytic continuation to the slit plane $\mathbf{C} \setminus [1,\infty)$ of

$$F(a,b;c;z) = {}_2F_1(a,b;c;z) \equiv \sum_{n=0}^{\infty} \frac{(a,n)(b,n)}{(c,n)} \frac{z^n}{n!}, \quad |z|<1. \tag{1.2}$$

Here $(a,0)=1$ for $a\neq 0$, and (a,n) is the *shifted factorial function*

$$(a,n)\equiv a(a+1)(a+2)\cdots(a+n-1)$$

for $n\in\mathbf{N}\equiv\{k : k \text{ is a positive integer}\}$. It is well known that $F(a,b;c;z)$ has many important applications, and many classes of special functions in mathematical physics are particular or limiting cases of this function. For these, and for properties of $F(a,b;c;z)$ see, for example, [**AS**, **Ask**, **Be1**, **Be2**, **Be3**, **R**, **Var**, **WW**].

To rewrite (1.1) in a slightly shorter form, we use the decreasing homeomorphism $\mu_a:(0,1)\to(0,\infty)$ defined by

$$\mu_a(r)\equiv\frac{\pi}{2\sin(\pi a)}\frac{F(a,1-a;1;1-r^2)}{F(a,1-a;1;r^2)}, \tag{1.3}$$

for $a\in(0,1)$. We can now write (1.1) as

$$\mu_a(s)=p\,\mu_a(r),\quad 0<r<1. \tag{1.4}$$

The solution of (1.4) is then given by

$$s=\varphi_K^a(r)\equiv\mu_a^{-1}(\mu_a(r)/K),\quad p=1/K. \tag{1.5}$$

We call $\varphi_K^a(r)$ *the modular function with signature* $1/a$ *and degree* $p=1/K$.

For the parameter $K=1/p$ with p a small positive integer, the function (1.5) satisfies several algebraic identities. The main cases studied in [**BBG**] are:

$$a=\frac{1}{6},\frac{1}{4},\frac{1}{3},\quad p=2,3,5,7,11,\dots .$$

For generalized modular equations we use the Ramanujan notation:

$$\alpha\equiv r^2,\quad \beta\equiv\varphi_{1/p}^a(r)^2.$$

We next state a few of the numerous identities [**BBG**] satisfied by $\varphi_{1/p}^a(r)$ for various values of the parameters a and p.

Theorem 1.6 ([**BBG**, Theorem 7.1])**.** *If β has degree* 2 *in the theory of signature* 3, *then, with $a=\frac{1}{3},\alpha=r^2,\beta=\varphi_{1/2}^a(r)^2$,*

$$(\alpha\beta)^{\frac{1}{3}}+\{(1-\alpha)(1-\beta)\}^{\frac{1}{3}}=1.$$

Theorem 1.7 ([**BBG**, Theorem 7.6])**.** *If β has degree* 5 *then, with $a=\frac{1}{3},\alpha=r^2,\beta=\varphi_{1/5}^a(r)^2$,*

$$(\alpha\beta)^{\frac{1}{3}}+\{(1-\alpha)(1-\beta)\}^{\frac{1}{3}}+3\{\alpha\beta(1-\alpha)(1-\beta)\}^{\frac{1}{6}}=1.$$

Theorem 1.8 ([**BBG**, Theorem 7.8])**.** *If β has degree* 11 *then, with $a=\frac{1}{3},\alpha=r^2,\beta=\varphi_{1/11}^a(r)^2$,*

$$\begin{aligned}(\alpha\beta)^{\frac{1}{3}}&+\{(1-\alpha)(1-\beta)\}^{\frac{1}{3}}+6\{\alpha\beta(1-\alpha)(1-\beta)\}^{\frac{1}{6}}\\&+3\sqrt{3}\{\alpha\beta(1-\alpha)(1-\beta)\}^{\frac{1}{12}}\{(\alpha\beta)^{\frac{1}{6}}+\{(1-\alpha)(1-\beta)\}^{\frac{1}{6}}\}=1.\end{aligned}$$

Theorems 1.6-1.8 are surprising, because they provide algebraic identities for the modular function, which itself is defined in terms of the transcendental function $\mu_a(r)$. It is an interesting open problem to determine which of the modular equations in [**BBG**] can be solved algebraically, explicitly in terms of the modular function.

For $r \in (0,1), a \in (0,1)$, and $r' = \sqrt{1-r^2}$, the *generalized elliptic integrals* (cf. [**BB1**, Section 5.5]) are defined by

$$\begin{cases} \mathcal{K}_a = \mathcal{K}_a(r) = \dfrac{\pi}{2} F(a, 1-a; 1; r^2), \\ \mathcal{K}'_a = \mathcal{K}'_a(r) \equiv \mathcal{K}_a(r'), \\ \mathcal{K}_a(0) = \dfrac{\pi}{2},\ \mathcal{K}_a(1) = \infty, \end{cases} \tag{1.9}$$

and

$$\begin{cases} \mathcal{E}_a = \mathcal{E}_a(r) \equiv \dfrac{\pi}{2} F(a-1, 1-a; 1; r^2), \\ \mathcal{E}'_a = \mathcal{E}'_a(r) \equiv \mathcal{E}_a(r'), \\ \mathcal{E}_a(0) = \dfrac{\pi}{2},\ \mathcal{E}_a(1) = \dfrac{\sin(\pi a)}{2(1-a)}. \end{cases} \tag{1.10}$$

Clearly, $\mathcal{K}_a$ is increasing and $\mathcal{E}_a$ is decreasing on $(0,1)$. These functions satisfy the remarkable identity

$$\mathcal{K}'_a \mathcal{E}_a + \mathcal{K}_a \mathcal{E}'_a - \mathcal{K}_a \mathcal{K}'_a = \frac{\pi \sin(\pi a)}{4(1-a)}, \tag{1.11}$$

which is a special case of Elliott's formula (see (3.15)). This identity, along with some other properties of the functions $\mathcal{E}_a$ and $\mathcal{K}_a$, is very useful for the study of the function μ_a. In the particular case $a = \frac{1}{2}$, the functions $\mathcal{K}_a(r)$ and $\mathcal{E}_a(r)$ reduce to $\mathcal{K}(r)$ and $\mathcal{E}(r)$, respectively, which are the well-known *complete elliptic integrals of the first and second kind*, respectively [**Bo**, **BF**]. By symmetry of (1.9), unless stated otherwise, we assume that $a \in (0, \frac{1}{2}]$.

The purpose of this paper is to study the modular function $\varphi^a_K(r)$ for general $a \in (0, \frac{1}{2}]$, as well as the related functions μ_a, $\mathcal{K}_a$, $\mathcal{E}_a$, m_a, and η^a_K, investigating their dependence on r, K, and a, where

$$m_a(r) \equiv \frac{2}{\pi \sin(\pi a)} r'^2 \mathcal{K}_a(r) \mathcal{K}'_a(r), \quad r \in (0,1), \tag{1.12}$$

$$\eta^a_K(x) \equiv \left(\frac{s}{s'}\right)^2, \quad s = \varphi^a_K(r), r = \sqrt{\frac{x}{x+1}}, x, K \in (0, \infty). \tag{1.13}$$

For $a = \frac{1}{2}$ these functions reduce to well-known special cases denoted by $\varphi_K, \mu, \mathcal{K}, \mathcal{E}, m, \eta_K$, which often occur in geometric function theory [**AVV**,

LV], number theory [**BB1**], and analytic function theory. For example, $\mu(r)$ appears in the *classical modular equation of degree* p, $p > 0$ (see [**Be3**] and [**BB1**]), that is, in the particular case $a = \frac{1}{2}$ of formula (1.1), while the upper bound in Schottky's Theorem is given in terms of $\mu(r)$ [**Mart**, Theorem 1.1]. Numerous properties of $\mu(r)$ have been studied (see, for instance, [**AVV**] and [**LV**]). For $a = \frac{1}{3}$, J.M. Borwein and P.B. Borwein [**BB2**] recently proved that there exists $\delta \in (0,1)$ such that the beautiful identity of Ramanujan,

$$\mathcal{K}'_{1/3}\left(\left(\frac{1-r}{1+2r}\right)^{3/2}\right) = (1+2r)\mathcal{K}_{1/3}(r^{3/2}),$$

is valid for all $r \in (0,\delta)$. They used it to derive a cubically convergent algorithm for the computation of π.

In Section 2 we construct a conformal map of a parallelogram with angles $\pi a, \pi(1-a)$, $0 < a < 1$, onto a half-plane. This map is denoted by sn_a. For $a = \frac{1}{2}$ this map reduces to the well-known Jacobian elliptic function sn [**Bo**].

In Sections 3 and 4 we summarize some of the basic properties of the hypergeometric functions, obtaining a new proof for a formula due to Ramanujan. We also derive differentiation formulas for $\mathcal{K}_a$, $\mathcal{E}_a$, for applications in Section 5, in which we generalize several inequalities for $\mathcal{K}_a$, $\mathcal{E}_a$, μ_a, proved in [**AVV**] for $a = \frac{1}{2}$. Our main results, which are based on the work in Section 5, are proved in Section 6.

We now state some of our main results for the generalized modular function $\varphi_K^a(r)$. The first two results show that this function satisfies simple multiplicative functional inequalities. For the special case $a = \frac{1}{2}$ see [**AVV**, Lemma 10.7 and Theorem 10.28].

Theorem 1.14. *For each $a \in (0, \frac{1}{2}]$ and $K \in (1,\infty)$, the function $f(x) \equiv \log(1/\varphi_K^a(e^{-x}))$ is increasing and convex on $(0,\infty)$, while $g(x) \equiv \log(1/\varphi_{1/K}^a(e^{-x}))$ is increasing and concave on $(0,\infty)$. In particular,*

$$\varphi_K^a(r)\varphi_K^a(t) \le (\varphi_K^a(\sqrt{rt}))^2,$$
$$\varphi_{1/K}^a(r)\varphi_{1/K}^a(t) \ge (\varphi_{1/K}^a(\sqrt{rt}))^2.$$

Theorem 1.15. *For each $a \in (0, \frac{1}{2}], K \in (1,\infty)$, and $r \in (0,1)$, the function $f(x) \equiv \varphi_K^a(rx)/\varphi_K^a(x)$ is increasing from $(0,1)$ onto $(r^{1/K}, \varphi_K^a(r))$, while $g(x) \equiv \varphi_{1/K}^a(rx)/\varphi_{1/K}^a(x)$ is decreasing from $(0,1)$ onto $(\varphi_{1/K}^a(r), r^K)$. In particular,*

$$\varphi_K^a(rt) \le \varphi_K^a(r)\varphi_K^a(t),$$
$$\varphi_{1/K}^a(rt) \ge \varphi_{1/K}^a(r)\varphi_{1/K}^a(t)$$

for each $r, t \in (0,1)$, with equality if and only if $K = 1$.

Because the derivative $\partial\varphi_K^a(r)/\partial r$ is unbounded on $(0,1)$, we now introduce a simple transformation which, when applied to $\varphi_K^a(r)$, yields a function whose derivative has range $(1/K, K)$. Because of its moderating influence on $\varphi_K^a(r)$, we refer to this transformation as a *linearization.*

Theorem 1.16. *Let $p : (0,1) \to (-\infty,\infty)$ and $q : (-\infty,\infty) \to (0,1)$, be given by $p(x) = 2\log(x/x')$ and $q(x) = p^{-1}(x) = \sqrt{e^x/(e^x+1)}$, and for $a \in (0, \frac{1}{2}]$, $K \in (1,\infty)$ let $g, h : (-\infty,\infty) \to (-\infty,\infty)$ be defined by $g = p \circ \varphi_K^a \circ q$, $h = p \circ \varphi_{1/K}^a \circ q$. Then g and h are increasing, g is convex, h is concave, and*

$$\frac{1}{K} \le g'(x) \le K,$$
$$\frac{1}{K} \le h'(x) \le K,$$

for all real x.

In Section 7 we study the dependence of the functions $\mathcal{K}_a$, $\mathcal{E}_a$, μ_a, μ_a^{-1}, and φ_K^a on the parameter a.

Throughout this paper the hyperbolic sine, cosine, and tangent functions and their inverses are denoted by $\mathrm{sh}, \mathrm{ch}, \mathrm{th}$ and $\mathrm{arsh}, \mathrm{arch}, \mathrm{arth}$, respectively. Whenever $r \in (0,1)$ we denote $r' = \sqrt{1-r^2}$.

2. Conformal mapping of a half plane onto a parallelogram.

Definition 2.1. For each number $a \in (0, \frac{1}{2}]$, $r \in (0,1)$, $\mathrm{Im}\, t \ge 0$, let $g(t) \equiv t^{-a}(1-t)^{a-1}(1-r^2t)^{-a}$ denote the analytic branch such that the argument of each of the factors t, $1-t$, and $1-r^2t$, is π whenever it is real and negative. We define the generalized elliptic sine function $\mathrm{sn}_a(w) = \mathrm{sn}_a(w,r)$ to be the inverse of the function

$$w = f(z) \equiv \frac{\sin(\pi a)}{2} \int_0^z g(t)dt, \qquad \mathrm{Im}\ z \ge 0.$$

By [**Bo**, p. 61, Example VI (4)], f is a conformal mapping of the upper half z-plane onto the interior of a parallelogram with angles πa and $\pi(1-a)$. The next result makes this notion more precise.

Theorem 2.2. *Let H denote the closed upper half-plane $\mathrm{Im}\, z \ge 0$, and let $r \in (0,1)$. The function f defined above is a homeomorphism of H onto the parallelogram with vertices $f(0) = 0$, $f(1) = \mathcal{K}_a(r)$, $f(1/r^2) = \mathcal{K}_a(r) + e^{i(1-a)\pi}\mathcal{K}_a'(r)$, and $f(\infty) = e^{i(1-a)\pi}\mathcal{K}_a'(r)$, conformal in the interior.*

Proof. The function f is a Schwarz-Christoffel mapping of H onto the polygon with interior angles $(1-a)\pi$, $a\pi$, $(1-a)\pi$, and $a\pi$, hence a parallelogram [**Mark**, §20]. Clearly $f(0) = 0$ and $f(1) = \mathcal{K}_a(r)$. Next, by [**AVV**, Theorem 1.19 (2)],

$$f\left(\frac{1}{r^2}\right) = \frac{\sin(\pi a)}{2} \int_0^{1/r^2} g(t)dt$$

$$\begin{aligned}
&= \frac{\sin(\pi a)}{2}\int_0^1 g(t)dt + \frac{\sin(\pi a)}{2}\int_1^{1/r^2} g(t)dt \\
&= \mathcal{K}_a(r) + \frac{\sin(\pi a)}{2}\int_1^{1/r^2} g(t)dt \\
&= \mathcal{K}_a(r) + \frac{\sin(\pi a)}{2}\int_0^1 e^{i\pi(1-a)}u^{a-1}(1-u)^{-a}(1-r'^2u)^{a-1}du \\
&= \mathcal{K}_a(r) + e^{i\pi(1-a)}\mathcal{K}'_a(r),
\end{aligned}$$

where in the last step we have made the change of variable $t = 1/(1-r'^2u)$. Finally, by symmetry, $f(\infty) = e^{i\pi(1-a)}\mathcal{K}'_a(r)$. □

Corollary 2.3. *Let P be a parallelogram with sides of length L, L' and angles $\pi(1-a)$, πa, $0 < a < \frac{1}{2}$. Then the conformal modulus of P* [**LV**] *is $\mathcal{K}'(r)/\mathcal{K}(r)$, where*

$$r = r_a \equiv \mu_a^{-1}\left(\frac{\pi}{2\sin(\pi a)}\frac{L'}{L}\right). \tag{2.4}$$

Proof. Choose r so $\mathcal{K}'_a(r)/\mathcal{K}_a(r) = L'/L$. Then P is similar to the parallelogram P' with vertices 0, $\mathcal{K}_a(r)$, $\mathcal{K}_a(r) + e^{i\pi(1-a)}\mathcal{K}'_a(r)$, and $e^{i\pi(1-a)}\mathcal{K}'_a(r)$. Then by (1.3) we have

$$\mu_a(r) = \frac{\pi}{2\sin(\pi a)}\frac{L'}{L},$$

implying that r is as in (2.4). By Theorem 2.2, sn_a maps P' conformally onto the upper half-plane, taking its vertices onto 0, 1, $1/r^2$, and ∞, respectively. Since the Jacobian elliptic sine function $\mathrm{sn} = \mathrm{sn}_{1/2}$ also maps the rectangle with vertices 0, $\mathcal{K}(r)$, $\mathcal{K}(r) + i\mathcal{K}'(r)$, and $\mathcal{K}'(r)$ conformally onto the upper half plane, taking its vertices onto 0, 1, $1/r^2$, and ∞, respectively [**Bo**], it follows that the conformal modulus of P is $\mathcal{K}'(r)/\mathcal{K}(r)$. □

Remark 2.5. Given two parallelograms P and Q with angles $\pi(1-a)$, πa and $\pi(1-b), \pi b$, and sides of lengths A, A' and B, B', respectively, the extremal quasiconformal mapping with least dilatation K from P onto Q, is given by

$$f = \psi \circ \mathrm{sn}_b^{-1}(\cdot, s) \circ \mathrm{sn}(\cdot, s) \circ g \circ \mathrm{sn}^{-1}(\cdot, r) \circ \mathrm{sn}_a(\cdot, r) \circ \varphi,$$

where φ and ψ are similarity mappings and g is an affine mapping. Here

$$r = \mu_a^{-1}\left(\frac{\pi}{2\sin(\pi a)}\frac{A'}{A}\right),\ s = \mu_b^{-1}\left(\frac{\pi}{2\sin(\pi b)}\frac{B'}{B}\right),\ \text{and } K = \max\left\{\frac{\mu(r)}{\mu(s)}, \frac{\mu(s)}{\mu(r)}\right\}.$$

3. Properties of $F(a,b;c;x)$.

In this section, we study some monotoneity properties of the function $F(a,b;c;x)$ and certain of its combinations with other functions. We first recall some well-known properties of this function which will be used in the sequel.

It is well known that hypergeometric functions are closely related to the classical *gamma function* $\Gamma(x)$, the *psi function* $\psi(x)$, and the *beta function* $B(x,y)$. For $\operatorname{Re} x > 0$, $\operatorname{Re} y > 0$, these functions are defined by

$$\Gamma(x) \equiv \int_0^\infty e^{-t}t^{x-1}dt,\ \psi(x) \equiv \frac{\Gamma'(x)}{\Gamma(x)},\ B(x,y) \equiv \frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)}, \tag{3.1}$$

respectively (cf. [**WW**]). It is well known that the gamma function satisfies the *difference equation* [**WW**, p. 237]

$$\Gamma(x+1) = x\Gamma(x), \tag{3.2}$$

and the *reflection property* [**WW**, p. 239]

$$\Gamma(x)\Gamma(1-x) = \frac{\pi}{\sin \pi x} = B(x, 1-x). \tag{3.3}$$

We shall also need the function

$$R(a,b) \equiv -2\gamma - \psi(a) - \psi(b),\ R(a) \equiv R(a, 1-a),\ R\left(\frac{1}{2}\right) = \log 16, \tag{3.4}$$

where γ is the *Euler-Mascheroni constant* given by

$$\gamma = \lim_{n\to\infty}\left(\sum_{k=1}^n \frac{1}{k} - \log n\right) = 0.577215\ldots . \tag{3.5}$$

By [**QVu2**, Lemma 2.14 (2)], for $a \in (0, \frac{1}{2}]$,

$$R(a) \equiv R(a, 1-a) \geq A\cdot\left(\frac{1}{2} - a\right)^2 + \log 16, \tag{3.6}$$

with equality if and only if $a = \frac{1}{2}$, where $A = 14\cdot\zeta(3) = 16.82879\ldots$, and where $\zeta(\cdot)$ is the *Riemann zeta function* [**WW**].

The hypergeometric function (1.2) has the following simple differentiation formula

$$\frac{d}{dx}F(a,b;c;x) = \frac{ab}{c}F(a+1, b+1; c+1; x). \tag{3.7}$$

An important tool for our work is the following classification of the behavior of the hypergeometric function near $x = 1$ in the three cases

$a+b<c, a+b=c$, and $a+b>c$:

(3.8)

$$\begin{cases} F(a,b;c;1) = \dfrac{\Gamma(c)\Gamma(c-a-b)}{\Gamma(c-a)\Gamma(c-b)}, \ a+b<c, \\ \\ B(a,b)F(a,b;a+b;x) + \log(1-x) = R(a,b) + \mathrm{O}((1-x)\log(1-x)), \\ \\ F(a,b;c;x) = (1-x)^{c-a-b}F(c-a,c-b;c;x), \ c<a+b. \end{cases}$$

The above asymptotic formula for the *zero-balanced* case $a+b=c$ is due to Ramanujan (see [**Ask**], [**Be2**]). This formula is implied by [**AS**, 15.3.10].

The asymptotic formula (3.8) gives us a precise description of the behavior of the function $F(a,b;a+b;x)$ near the logarithmic singularity $x=1$. In the next theorem we show that by an exponential change of variables we can cancel this singularity and that the transformed function will be nearly linear.

Theorem 3.9. *For $a,b>0$, define $g(x) = F(a,b;a+b;1-e^{-x}), x>0$. Then g is an increasing and convex function with $g'((0,\infty)) = (ab/(a+b), \Gamma(a+b)/(\Gamma(a)\Gamma(b)))$.*

Proof. By [**AS**, 15.3.3] for $|z|<1$,

$$F(a,b;a+b+1;z) = (1-z)F(a+1,b+1;a+b+1;z).$$

From this relation and (3.7) we obtain

$$\begin{aligned} (a+b)g'(x)/(ab) &= F(a+1,b+1;a+b+1;1-e^{-x})e^{-x} \\ &= F(a,b;a+b+1;1-e^{-x}), \end{aligned}$$

so that g' is positive and increasing and has the asserted range, by (3.8). $\square$

Theorem 3.10. *Given $a,b>0$, and $a+b>c$, $d \equiv a+b-c$, the function $f(x) = F(a,b;c;1-(1+x)^{-1/d})$, $x>0$, is increasing and convex, with $f'((0,\infty)) = (ab/(cd), \Gamma(c)\Gamma(d)/(\Gamma(a)\Gamma(b)))$.*

Proof. By [**AS**, 15.2.1 and 15.3.3],

$$\begin{aligned} f'(x) &= \frac{ab}{c}F(a+1,b+1;c+1;1-(1+x)^{-1/d})(1+x)^{-1/d-1}\frac{1}{d} \\ &= \frac{ab}{cd}\frac{((1+x)^{-1/d})^{-d-1}}{(1+x)^{1+1/d}}F(c-a,c-b;c+1;1-(1+x)^{-1/d}) \\ &= \frac{ab}{cd}F(c-a,c-b;c+1;1-(1+x)^{-1/d}), \end{aligned}$$

so that f' is increasing and has the asserted range, by (3.8). $\square$

3.11. Gauss contiguous relations and derivative formula. The six functions $F(a\pm1,b;c;z), F(a,b\pm1;c;z), F(a,b;c\pm1;z)$ are called *contiguous* to $F(a,b;c;z)$. Gauss gave 15 relations between $F(a,b;c;z)$ and pairs of its contiguous functions [**AS**, 15.2.10-15.2.27], [**R**, Section 33]. Using these relations, we shall write the differentiation formula (3.7) as in Theorem 3.12, which will be useful in our study. In particular, the differentiation formulas for the generalized elliptic integrals in Section 4 follow from Theorem 3.12. We prove this result, since this is not included in [**AS**] and since we have not found a proof in the literature. However, parts (1) and (2) of Theorem 3.12 are stated in [**Mi**, p. 267] in a slightly different form. Part (4) of Theorem 3.12 seems to be new.

Theorem 3.12. *For* $a,b,c>0$, $r\in(0,1)$, *let* $u=u(r)=F(a-1,b;c;r)$, $v=v(r)=F(a,b;c;r)$, $u_1=u(1-r)$, $v_1=v(1-r)$. *Then*

$$r\frac{du}{dr}=(a-1)(v-u), \tag{1}$$

$$r(1-r)\frac{dv}{dr}=(c-a)u+(a-c+br)v, \tag{2}$$

or, equivalently,

$$\frac{ab}{c}r(1-r)F(a+1,b+1;c+1;r)=(c-a)u+(a-c+br)v, \tag{3}$$

$$\begin{cases} r(1-r)\dfrac{d}{dr}\big(uv_1+u_1v-vv_1\big)= \\ (1-a-b)\big[(1-r)uv_1-ru_1v-(1-2r)vv_1\big]. \end{cases} \tag{4}$$

Proof. (1) Observing that

$$\begin{aligned}\left(r\frac{d}{dr}+(a-1)\right)u(r)&=\sum_{n=0}^{\infty}(a-1+n)\frac{(a-1,n)(b,n)}{(c,n)n!}r^n\\ &=(a-1)\sum_{n=0}^{\infty}\frac{(a,n)(b,n)}{(c,n)n!}r^n=(a-1)v(r),\end{aligned}$$

we have the desired equality.

(2) If we change $a-1$ to a in (1), then

$$\frac{dv}{dr}=\frac{a}{r}(w-v),$$

where $w=F(a+1,b;c;r)$. Gauss' relation for contiguous functions [**AS**, 15.2.10] yields

$$w=\frac{1}{a(1-r)}((c-a)u+(2a-c-ar+br)v).$$

Thus

$$\begin{aligned}\frac{dv}{dr} &= \frac{a}{r}\left(\frac{1}{a(1-r)}((c-a)u+(2a-c-ar+br)v)-v\right)\\ &= \frac{(c-a)}{r(1-r)}u+\frac{(a-c+br)}{r(1-r)}v,\end{aligned}$$

as desired.

Part (3) follows from (3.7) and part (2). Part (4) follows from (1), (2), and the product and chain rules. □

We next apply Theorem 3.12 to give a new proof in Corollary 3.13 (3) for a special case of Ramanujan's formula for the derivative of the quotient of two hypergeometric functions (cf. [**Be2**, p. 88, Corollary]). In Corollary 3.13 (5) we give a generalization of formula (1.11). Both (1.11) and 3.13 (5) are generalizations of the Legendre relation for elliptic integrals.

Corollary 3.13. *For $a, b, c > 0, r \in (0,1)$, let $u = u(r) = F(a-1, b; c; r)$, $v = v(r) = F(a, b; c; r)$, $u_1 = u(1-r)$, and $v_1 = v(1-r)$. Then*

$$(1) \qquad -r(1-r)v^2\frac{d}{dr}\left(\frac{v_1}{v}\right) = (c-a)\left(uv_1+u_1v\right)+(b+2(a-c))\,vv_1$$

and

$$(2) \qquad r\frac{d}{dr}\left[(1-r)vv_1\right] = (c-a)\left[uv_1-u_1v\right]+\left[(2b-1)r-b\right]vv_1.$$

For $n = 0, 1, 2, \ldots$,

$$(3) \qquad r\frac{d^{n+1}u}{dr^{n+1}} = (a-1)\frac{d^n v}{dr^n}-(a+n-1)\frac{d^n u}{dr^n},$$

(4)

$$r(1-r)\frac{d^{n+1}v}{dr^{n+1}} = (a-c+br+2nr-n)\frac{d^n v}{dr^n}+(n^2+n(b-1))\frac{d^{n-1}v}{dr^{n-1}}+(c-a)\frac{d^n u}{dr^n}.$$

In particular, if $a \in (0,1)$, $b = 1-a < c$, then

$$(5) \qquad uv_1+u_1v-vv_1 = u(1) = \frac{\Gamma(c)^2}{\Gamma(c+a-1)\Gamma(c-a+1)}.$$

In addition, if $c = 1$, then

$$(6) \qquad \frac{d}{dr}\left(\frac{v_1}{v}\right) = -\frac{\sin(\pi a)}{\pi r(1-r)v^2},$$

$$(7) \qquad r\frac{d}{dr}\left((1-r)vv_1\right) = \frac{\sin(\pi a)}{\pi}+v[(1-2a)rv_1-2(1-a)u_1].$$

Proof. Parts (1)-(4) follow from Theorem 3.12 (1) and the Leibnitz Rule for higher derivatives. Next, let $a, b \in (0,1)$ and $b = 1 - a < c$. Then Theorem 3.12 (4) implies that the function in (5) is constant. If we let r tend to 0 and make use of (3.7) and (3.8), it follows that this constant has the asserted value, and (5) follows. If, in addition, $c = 1$, then (6) and (7) follow from (1) and (2), respectively. ☐

3.14. Elliott's formula. E.B. Elliott [**El**] (cf. [**Ba**, p. 85]) proved the identity

(3.15)

$$\begin{aligned}
&F\left(\frac{1}{2}+\lambda, -\frac{1}{2}-\nu; 1+\lambda+\mu; z\right) F\left(\frac{1}{2}-\lambda, \frac{1}{2}+\nu; 1+\nu+\mu; 1-z\right)\\
&+F\left(\frac{1}{2}+\lambda, \frac{1}{2}-\nu; 1+\lambda+\mu; z\right) F\left(-\frac{1}{2}-\lambda, \frac{1}{2}+\nu; 1+\nu+\mu; 1-z\right)\\
&-F\left(\frac{1}{2}+\lambda, \frac{1}{2}-\nu; 1+\lambda+\mu; z\right) F\left(\frac{1}{2}-\lambda, \frac{1}{2}+\nu; 1+\nu+\mu; 1-z\right)\\
&=\frac{\Gamma(1+\lambda+\mu)\Gamma(1+\nu+\mu)}{\Gamma\left(\lambda+\mu+\nu+\frac{3}{2}\right)\Gamma\left(\frac{1}{2}+\mu\right)}.
\end{aligned}$$

If we put $\lambda = \nu = \frac{1}{2} - a$ and $\mu = c + a - \frac{3}{2}$ in Elliott's identity, we obtain the generalized Legendre relation in Corollary 3.13 (5), as was pointed out to us by B.C. Carlson.

Conjecture 3.16. *Let* $u(a,b,c,r) = F(a-1,b;c;r)$, $v(a,b,c,r) = F(a,b;c;r)$, *and let*

$$\begin{aligned}
\mathcal{L}(a,b,c,r) = u(a,b,c,r)v(a,b,c,1-r) + u(a,b,c,1-r)v(a,b,c,r)\\
-v(a,b,c,r)v(a,b,c,1-r)
\end{aligned}$$

for $r \in (0,1)$. *Note that* $\mathcal{L}(a,a,a,r) \equiv 0$ *for* $a > 0$. *Observe also that by the generalized Legendre relation in Corollary* 3.13 (5), $\mathcal{L}(a,1-a,c,r)$ *is a constant for* $a \in (0,1), 1-a < c$. *For* $a, b \in (0,1), a+b \le 1$ (≥ 1), *we conjecture that the function* $\mathcal{L}(a,b,1,r)$ *is concave* (*convex*) *as a function of* r *on* $(0,1)$.

4. Derivative formulas.

The following derivative formulas are analogous to those well-known ones when $a = \frac{1}{2}$. In particular, formulas (1), (2), (3), (4) are analogues of [**BF**, 710.00, 710.02, 710.05, 710.04], respectively, while (4), (5), (6), (7), (8), (9), (10), (11) generalize formulas (9), (19), (22), (23), (24), (18), (25), (26), respectively, in [**AVV**, Appendix E].

Theorem 4.1. *For each $a \in (0, \frac{1}{2}]$ the following derivative formulas hold for $r \in (0,1)$ and $x, y, K \in (0,\infty)$:*

$$\frac{d\mathcal{K}_a}{dr} = \frac{2(1-a)(\mathcal{E}_a - r'^2\mathcal{K}_a)}{rr'^2}. \tag{1}$$

$$\frac{d\mathcal{E}_a}{dr} = \frac{2(a-1)(\mathcal{K}_a - \mathcal{E}_a)}{r}. \tag{2}$$

$$\frac{d}{dr}(\mathcal{K}_a - \mathcal{E}_a) = \frac{2(1-a)r\mathcal{E}_a}{r'^2}. \tag{3}$$

$$\frac{d}{dr}(\mathcal{E}_a - r'^2\mathcal{K}_a) = 2ar\mathcal{K}_a. \tag{4}$$

$$\frac{d\mu_a(r)}{dr} = -\frac{\pi^2}{4rr'^2\mathcal{K}_a(r)^2}. \tag{5}$$

$$\frac{d\mu_a^{-1}(y)}{dy} = -\frac{4rr'^2\mathcal{K}_a(r)^2}{\pi^2}, \quad \textit{where } r = \mu_a^{-1}(y). \tag{6}$$

$$\frac{\partial\varphi_K^a(r)}{\partial r} = \frac{1}{K}\frac{ss'^2\mathcal{K}_a(s)^2}{rr'^2\mathcal{K}_a(r)^2} = \frac{ss'^2\mathcal{K}_a(s)\mathcal{K}_a'(s)}{rr'^2\mathcal{K}_a(r)\mathcal{K}_a'(r)} = K\frac{ss'^2\mathcal{K}_a'(s)^2}{rr'^2\mathcal{K}_a'(r)^2}, \tag{7}$$

$$\frac{\partial\varphi_K^a(r)}{\partial K} = \frac{4ss'^2\mathcal{K}_a(s)^2}{\pi^2}\frac{\mu_a(r)}{K^2}, \tag{8}$$

where $s = \varphi_K^a(r)$.

$$\frac{dm_a(r)}{dr} = \frac{2}{\pi r\sin(\pi a)}\left[\frac{\pi\sin(\pi a)}{2} - 4(1-a)\mathcal{K}_a(r)\mathcal{E}_a'(r) + 2(1-2a)r^2\mathcal{K}_a(r)\mathcal{K}_a'(r)\right]. \tag{9}$$

(10)
$$\frac{\partial\eta_K^a(x)}{\partial x} = \frac{1}{K}\left(\frac{r's\mathcal{K}_a(s)}{rs'\mathcal{K}_a(r)}\right)^2 = K\left(\frac{r's\mathcal{K}_a'(s)}{rs'\mathcal{K}_a'(r)}\right)^2 = \left(\frac{r's}{rs'}\right)^2\frac{\mathcal{K}_a(s)\mathcal{K}_a'(s)}{\mathcal{K}_a(r)\mathcal{K}_a'(r)}.$$

$$\frac{\partial\eta_K^a(x)}{\partial K} = \frac{8\eta_K^a(x)\mu_a(r)\mathcal{K}_a(s)^2}{\pi^2K^2}. \tag{11}$$

In (10) and (11), $r = \sqrt{x/(x+1)}$ and $s = \varphi_K^a(r)$.

Proof. Formulas (1) and (2) follow from Theorem 3.12 and the Chain Rule (see also [**BB1**, (5.5.5)]), while (3) and (4) follow from (1) and (2). Formula (5) follows immediately from Corollary 3.13 (1), and (6) follows immediately from (5).

For (7), we let $s = \varphi_K^a(r)$. Then $\mu_a(s) = (1/K)\mu_a(r)$. Hence, from (5),

$$-\frac{\pi^2}{4ss'^2\mathcal{K}_a(s)^2}\frac{\partial s}{\partial r} = -\frac{1}{K}\frac{\pi^2}{4rr'^2\mathcal{K}_a(r)^2},$$

and (7) follows. Similarly, from (5) we obtain

$$-\frac{\pi^2}{4ss'^2\mathcal{K}_a(s)^2}\frac{\partial s}{\partial K} = -\frac{1}{K^2}\mu_a(r),$$

which yields (8).

Formula (9) follows from Corollary 3.13 (7) and the Chain Rule.

For (10), since $\eta_K^a(x) = (s/s')^2$, where $r = \sqrt{x/(x+1)}$ and $s = \varphi_K^a(r)$, formula (7) gives

$$\begin{aligned}\frac{\partial \eta_K^a(x)}{\partial x} &= 2\left(\frac{s}{s'}\right)^2\left[\frac{1}{s}\frac{\partial s}{\partial x} + \frac{s}{s'^2}\frac{\partial s}{\partial x}\right] = 2\left(\frac{s}{s'}\right)^2\frac{1}{ss'^2}\frac{1}{K}\frac{ss'^2\mathcal{K}_a(s)^2}{rr'^2\mathcal{K}_a(r)^2}\frac{\partial r}{\partial x}\\ &= 2\frac{s^2}{s'^2}\frac{1}{K}\frac{\mathcal{K}_a(s)^2}{rr'^2\mathcal{K}_a(r)^2}\frac{r'^4}{2r} = \frac{1}{K}\left(\frac{r's\mathcal{K}_a(s)}{rs'\mathcal{K}_a(r)}\right)^2\\ &= K\left(\frac{r's\mathcal{K}'_a(s)}{rs'\mathcal{K}'_a(r)}\right)^2 = \left(\frac{r's}{rs'}\right)^2\frac{\mathcal{K}_a(s)\mathcal{K}'_a(s)}{\mathcal{K}_a(r)\mathcal{K}'_a(r)}.\end{aligned}$$

Finally, for (11), with $\eta_K^a(x) = (s/s')^2$ as in (10), formula (8) yields

$$\begin{aligned}\frac{\partial \eta_K^a(x)}{\partial K} &= 2\left(\frac{s}{s'}\right)^2\left[\frac{1}{s} + \frac{s}{s'^2}\right]\frac{\partial s}{\partial K}\\ &= 2\eta_K^a(x)\frac{1}{ss'^2}\frac{4ss'^2\mathcal{K}_a(s)^2}{\pi^2}\frac{\mu_a(r)}{K^2}\\ &= \frac{8}{\pi^2K^2}\eta_K^a(x)\mu_a(r)\mathcal{K}_a(s)^2.\end{aligned}$$

□

4.2. Hypergeometric differential equation. Since the hypergeometric function $y = F(a,b;c;x)$ satisfies the differential equation

$$x(1-x)y'' + [c-(a+b+1)x]y' - aby = 0$$

[**R**, p. 54], it follows from the Chain Rule that $y = F(a,b;c;r^2)$ satisfies

$$rr'^2\frac{d^2y}{dr^2} + [(2c-1)-(2a+2b+1)r^2]\frac{dy}{dr} - 4abry = 0.$$

For a review of the theory of this differential equation, see [**Var**]. In particular, the functions $y = \mathcal{K}_a(r)$ (or $\mathcal{K}'_a(r)$) and $z = \mathcal{E}_a(r)$ (or $\mathcal{E}'_a(r)$) satisfy the differential equations

$$\begin{cases} rr'^2\dfrac{d^2y}{dr^2} + (1-3r^2)\dfrac{dy}{dr} - 4a(1-a)ry = 0, \\ rr'^2\dfrac{d^2z}{dr^2} + r'^2\dfrac{dz}{dr} + 4(1-a)^2rz = 0. \end{cases} \tag{4.3}$$

These reduce to the standard differential equations for $\mathcal{K}$ (or $\mathcal{K}'$) and $\mathcal{E}$ (or $\mathcal{E}'$) [**BF**, 118.02], respectively, when $a = \frac{1}{2}$.

4.4. Particular values. According to [**AS**, 15.1.26], [**BB1**, p. 191, Exercise 22], or [**R**, p. 69, Exercise 3], we have the following formula:

$$\mathcal{K}_a\left(\frac{1}{\sqrt{2}}\right) = \frac{c}{4\sqrt{\pi}}\sin(\pi a), \qquad c = \Gamma\left(\frac{1-a}{2}\right)\Gamma\left(\frac{a}{2}\right), \tag{4.5}$$

while (4.5) and (1.11) give

$$\mathcal{E}_a\left(\frac{1}{\sqrt{2}}\right) = \frac{4\pi^2 + (1-a)c^2\sin(\pi a)}{8\sqrt{\pi}(1-a)c}. \tag{4.6}$$

4.7. Some identities. It is easy to see that, for each $a \in (0, \frac{1}{2}]$, the function μ_a in (1.3) and its inverse are strictly decreasing functions and satisfy the identities

$$\mu_a(r)\mu_a(r') = \frac{\pi^2}{4\sin^2(\pi a)} \tag{4.8}$$

and

$$\mu_a^{-1}(x)^2 + \mu_a^{-1}(y)^2 = 1, \tag{4.9}$$

for $r \in (0,1)$, $r' = \sqrt{1-r^2}$, $x, y \in (0,\infty)$ with $xy = \pi^2/(4\sin^2(\pi a))$ (see the solution of [**AVV**, Exercise 5.45 (1), p. 364]).

For $a = \frac{1}{2}$ the relation (4.8) reduces to a well-known property of the function $\mu(r)$ (cf. [**LV**, p. 61, (2.7)]). The function $m_a(r)$ introduced in (1.12) is easily seen to satisfy the identities

$$m_a(r) + m_a(r') = \frac{2}{\pi\sin(\pi a)}\mathcal{K}_a(r)\mathcal{K}'_a(r) \quad \text{and} \quad r^2 m_a(r) = r'^2 m_a(r'). \tag{4.10}$$

5. Generalized elliptic integrals.

The following *monotone form of l'Hôpital's Rule* [**AVV**, Theorem 1.25] will be extremely useful in our proofs. We have recently learned from R. Kellerhals of a similar result due to M. Gromov which is a handy tool for volume estimation in Riemannian geometry (see [**C**, p. 124, Lemma 3.1]).

Lemma 5.1. *For* $-\infty < a < b < \infty$, *let* $f, g : [a, b] \to \mathbf{R}$ *be continuous on* $[a, b]$, *and be differentiable on* (a, b). *Let* $g'(x) \neq 0$ *on* (a, b). *If* $f'(x)/g'(x)$ *is increasing (decreasing) on* (a, b), *then so are*

$$[f(x) - f(a)]/[g(x) - g(a)] \quad \textit{and} \quad [f(x) - f(b)]/[g(x) - g(b)].$$

If $f'(x)/g'(x)$ *is strictly monotone, then the monotoneity in the conclusion is also strict.*

In the next two lemmas and the following theorem we study how these generalized elliptic integrals $\mathcal{K}_a$ and $\mathcal{E}_a$ depend upon the variable r. When $a = \frac{1}{2}$, Lemma 5.2 (1), (3), (4), (10), (12) reduce to Theorem 3.21 (1), Exercise 3.43 (32), (46), and Theorem 3.31 (6), Exercise 3.43 (15), respectively, in [**AVV**].

Lemma 5.2. *Let* $a \in (0, \frac{1}{2}]$ *be given, and let* $b = 1 - a, c = (\sin(\pi a))/b$. *Then the function*

(1) $f_1(r) \equiv (\mathcal{E}_a - r'^2\mathcal{K}_a)/r^2$ *is increasing and convex from* $(0, 1)$ *onto* $(\pi a/2, c/2)$.
(2) $f_2(r) \equiv r'^2\mathcal{K}_a/\mathcal{E}_a$ *is decreasing from* $(0, 1)$ *onto* $(0, 1)$.
(3) $f_3(r) \equiv (\mathcal{K}_a - \mathcal{E}_a)/(r^2\mathcal{K}_a)$ *is increasing from* $(0, 1)$ *onto* $(b, 1)$.
(4) $f_4(r) \equiv (\mathcal{E}_a - r'^2\mathcal{K}_a)/(r^2\mathcal{K}_a)$ *is decreasing from* $(0, 1)$ *onto* $(0, a)$.
(5) $f_5(r) \equiv (\mathcal{E}_a - r'^2\mathcal{K}_a)/(\mathcal{K}_a - \mathcal{E}_a)$ *is decreasing from* $(0, 1)$ *onto* $(0, a/b)$.
(6) $f_6(r) \equiv r'^2(\mathcal{K}_a - \mathcal{E}_a)/(r^2\mathcal{E}_a)$ *is decreasing from* $(0, 1)$ *onto* $(0, b)$.
(7) $f_7(r) \equiv ((\pi/2)^2 - (r'\mathcal{K}_a)^2)/(\mathcal{E}_a - r'^2\mathcal{K}_a)$ *is increasing from* $(0, 1)$ *onto* $(\pi(a^2 + b^2)/(2a),\ \pi^2/(2c))$.
(8) $f_8(r) \equiv (\mathcal{K}_a - \pi/2)/\log(1/r')$ *is increasing from* $(0, 1)$ *onto* $(\pi ab, bc)$.
(9) $f_9(r) \equiv (\mathcal{E}_a - (1 - r)\mathcal{K}_a)/r$ *is decreasing from* $(0, 1)$ *onto* $(c/2, \pi/2)$.
(10) $f_{10}(r) \equiv r^{-2}[(\mathcal{E}_a - r'^2\mathcal{K}_a)/r^2 - \pi a/2]$ *is increasing and convex from* $(0, 1)$ *onto* $(\pi a^2 b/4, (c - \pi a)/2)$.
(11) $f_{11}(r) \equiv [a(\mathcal{K}_a - \mathcal{E}_a) - (1 - a)(\mathcal{E}_a - r'^2\mathcal{K}_a)]/(2\log(1/r') - r^2)$ *is increasing from* $(0, 1)$ *onto* $(\pi a^2 b/2, abc/2)$.
(12) $f_{12}(r) \equiv (\mathcal{K}_a - \mathcal{E}_a)/\log(1/r')$ *is decreasing from* $(0, 1)$ *onto* $(\sin(\pi a), \pi b)$.
(13) $f_{13}(r) \equiv (\pi/2 - r'^2\mathcal{K}_a)/r^2$ *is increasing and convex from* $(0, 1)$ *onto* $(\pi(a^2 + b)/2, \pi/2)$.
(14) $f_{14}(r) \equiv (\pi^2/4 - r'^2\mathcal{K}_a^2)/r^2$ *is increasing from* $(0, 1)$ *onto* $(\pi^2(a^2 + b^2)/4, \pi^2/4)$.

Proof. (1) It follows from (1.2), (1.9), and (1.10) that

(5.3)

$$
\begin{aligned}
&\mathcal{E}_a - r'^2\mathcal{K}_a\\
&= \frac{\pi}{2}\Bigg\{\sum_{n=0}^{\infty}\frac{(a-1)(a,n-1)(1-a,n)-(a,n)(1-a,n)}{(n!)^2}r^{2n}\\
&\quad+\sum_{n=0}^{\infty}\frac{(a,n)(1-a,n)}{(n!)^2}r^{2(n+1)}\Bigg\}\\
&= \frac{\pi}{2}\sum_{n=1}^{\infty}\frac{n^2(1-a,n-1)(a,n-1)-(1-a,n)[(1-a)(1,n-1)+(a,n)]}{(n!)^2}r^{2n}\\
&= \frac{\pi a}{2}\sum_{n=1}^{\infty}\frac{n(a,n-1)(1-a,n-1)}{(n!)^2}r^{2n}\\
&= \frac{\pi a}{2}r^2\sum_{n=0}^{\infty}\frac{1}{n+1}a_n r^{2n},
\end{aligned}
$$

where $a_n = (a,n)(1-a,n)(n!)^{-2}$, and hence the result follows immediately.

(2) This follows from $r'^2\mathcal{K}_a/\mathcal{E}_a = 1-((\mathcal{E}_a - r'^2\mathcal{K}_a)/\mathcal{E}_a)$, since $\mathcal{E}_a - r'^2\mathcal{K}_a$ is increasing and $\mathcal{E}_a$ is decreasing by Theorem 4.1 (4), (2), respectively. The limiting values are clear, as in (1).

(3) We write $f_3(r) = g_3(r)/h_3(r)$, where $g_3(r) = \mathcal{K}_a - \mathcal{E}_a$ and $h_3(r) = r^2\mathcal{K}_a$. Then $g_3(0) = h_3(0) = 0$ and

$$
\begin{aligned}
\frac{g_3'(r)}{h_3'(r)} &= \frac{2(1-a)r\mathcal{E}_a}{r'^2\left[2r\mathcal{K}_a + 2(1-a)r(\mathcal{E}_a - r'^2\mathcal{K}_a)/r'^2\right]}\\
&= \frac{(1-a)\mathcal{E}_a}{(1-a)\mathcal{E}_a + ar'^2\mathcal{K}_a} = \frac{1-a}{1-a+ar'^2\mathcal{K}_a/\mathcal{E}_a},
\end{aligned}
$$

which is increasing by (2). Clearly $f_3(1-) = 1$, while $f_3(0+) = 1-a$ by l'Hôpital's Rule. Hence (3) follows from Lemma 5.1.

(4) This follows from (3), since $f_4(r) = 1 - f_3(r)$.

(5) This follows from (3) and (4).

(6) Since $f_6(r) = 1 - f_1(r)/\mathcal{E}_a$, the result follows from (1).

(7) Write $f_7(r) = g_7(r)/h_7(r)$, where $g_7(r) = (\pi^2/4) - (r'\mathcal{K}_a)^2$ and $h_7(r) = \mathcal{E}_a - r'^2\mathcal{K}_a$. Then $g_7(0) = h_7(0) = 0$ and

$$
\begin{aligned}
\frac{g_7'(r)}{h_7'(r)} &= \frac{2r\mathcal{K}_a^2 - 4(1-a)\mathcal{K}_a(\mathcal{E}_a - r'^2\mathcal{K}_a)/r}{2ar\mathcal{K}_a}\\
&= \frac{\mathcal{K}_a}{a}\left[1 - 2(1-a)\frac{\mathcal{E}_a - r'^2\mathcal{K}_a}{r^2\mathcal{K}_a}\right],
\end{aligned}
$$

which is increasing by (4). By l'Hôpital's Rule, $f_7(0+) = \pi(1-2a+2a^2)/(2a)$, while $f_7(1-) = \pi^2/(4\mathcal{E}_a(1)) = \pi^2(1-a)/(2\sin(\pi a))$.

(8) Write $f_8(r) = g_8(r)/h_8(r)$, where $g_8(r) = \mathcal{K}_a - \pi/2$ and $h_8(r) = \log(1/r')$. Then $g_8(0) = h_8(0) = 0$ and $g_8'(r)/h_8'(r) = 2(1-a)(\mathcal{E}_a - r'^2\mathcal{K}_a)/r^2$, which is increasing by (1). Hence $f_8(r)$ is increasing by Lemma 5.1. The limits $f_8(0+) = a(1-a)\pi$ and $f_8(1-) = 2(1-a)\mathcal{E}_a(1) = \sin(\pi a)$ follow by l'Hôpital's Rule.

(9) Write $f_9(r) = g_9(r)/h_9(r)$, where $g_9(r) = \mathcal{E}_a - (1-r)\mathcal{K}_a$ and $h_9(r) = r$. Then $g_9(0) = h_9(0) = 0$ and

$$\begin{aligned}\frac{g_9'(r)}{h_9'(r)} &= -2(1-a)\frac{\mathcal{K}_a - \mathcal{E}_a}{r} + \mathcal{K}_a - \frac{2(1-a)(\mathcal{E}_a - r'^2\mathcal{K}_a)}{r(1+r)}\\ &= -(1-2a)\mathcal{K}_a + 2(1-a)\frac{\mathcal{E}_a}{1+r},\end{aligned}$$

which is decreasing. The limit $f_9(0+)$ follows from l'Hôpital's Rule, while $f_9(1-)$ is clear from (1.9). Hence f_9 is decreasing by Lemma 5.1.

(10) Since

$$f_{10}(r) = \frac{\pi a}{2}\sum_{n=1}^{\infty}\frac{a_n}{n+1}r^{2(n-1)}$$

by (5.3), the result follows immediately.

(11) Write $f_{11}(r) = g_{11}(r)/h_{11}(r)$, where $g_{11}(r)$ and $h_{11}(r)$ are the numerator and denominator, respectively, of the fraction defining $f_{11}(r)$. Then $g_{11}(0) = h_{11}(0) = 0$ and

$$\frac{g_{11}'(r)}{h_{11}'(r)} = a(1-a)\frac{\mathcal{E}_a - r'^2\mathcal{K}_a}{r^2},$$

which is increasing by (1), with limits $\pi a^2(1-a)/2$ and $a(\sin(\pi a))/2$ at 0 and 1, respectively. Hence the result follows from Lemma 5.1.

(12) Write $f_{12}(r) = g_{12}(r)/h_{12}(r)$, where $g_{12}(r) = \mathcal{K}_a - \mathcal{E}_a$ and $h_{12}(r) = \log(1/r')$. Then $g_{12}(0) = h_{12}(0) = 0$ and $g_{12}'(r)/h_{12}'(r) = 2(1-a)\mathcal{E}_a$, and the result follows from Lemma 5.1 since $\mathcal{E}_a$ is decreasing.

(13) It follows from (1.2) and (1.9) that

$$\begin{aligned}f_{13}(r) &= \frac{\pi}{2r^2}\left[\sum_{n=0}^{\infty}\frac{(a,n)(1-a,n)}{(n!)^2}r^{2(n+1)} - \sum_{n=1}^{\infty}\frac{(a,n)(1-a,n)}{(n!)^2}r^{2n}\right]\\ &= \frac{\pi}{2r^2}\sum_{n=1}^{\infty}\frac{(a,n-1)(1-a,n-1)}{(n!)^2}(n+a^2-a)r^{2n}\\ &= \frac{\pi}{2}\sum_{n=0}^{\infty}\frac{(a,n)(1-a,n)}{[(n+1)!]^2}(n+a^2+b)r^{2n},\end{aligned}$$

and hence the result follows.

(14) Write $f_{14}(r) = g_{14}(r)/h_{14}(r)$, where $g_{14}(r) = \pi^2/4 - r'^2\mathcal{K}_a^2$ and $h_{14}(r) = r^2$. Then $g_{14}(0) = h_{14}(0) = 0$, and

$$\frac{g'_{14}(r)}{h'_{14}(r)} = \mathcal{K}_a^2 \cdot \left[1 - \frac{2(1-a)(\mathcal{E}_a - r'^2\mathcal{K}_a)}{r^2\mathcal{K}_a}\right],$$

which is increasing, with $f_{14}(0+) = \pi^2(1-2a+2a^2)/4$, by part (4). The value $f_{14}(1-) = \pi^2/4$ is clear. Hence, the result follows from Lemma 5.1. $\square$

In the next lemma, when $a = \frac{1}{2}$, parts (1), (2), (4), (5) reduce to Theorem 3.21 (7), (8) and Exercise 3.43 (30), (45), respectively, in [**AVV**], while part (3) becomes the monotone property of $\mathcal{E}$.

Lemma 5.4. *Let $a \in (0, \frac{1}{2}]$. Then the function*

(1) *$f_1(r) \equiv r'^c\mathcal{K}_a$ is decreasing if and only if $c \geq 2a(1-a)$, in which case $r'^c\mathcal{K}_a$ is decreasing from $(0,1)$ onto $(0, \pi/2)$. Moreover, $\sqrt{r'}\mathcal{K}_a$ is decreasing for each $a \in (0, \frac{1}{2}]$.*

(2) *$f_2(r) \equiv r'^c\mathcal{E}_a$ is increasing if and only if $c \leq -2(1-a)^2$. In particular, $r'^{-2(1-a)^2}\mathcal{E}_a$ is increasing from $(0,1)$ onto $(\pi/2, \infty)$. Moreover, $\mathcal{E}_a/r'^2$ is increasing for each $a \in (0, \frac{1}{2}]$.*

(3) *$f_3(r) \equiv \mathcal{E}_a + (1-2a)(\mathcal{E}_a - r'^2\mathcal{K}_a)$ is decreasing from $(0,1)$ onto $(\sin(\pi a), \pi/2)$.*

(4) *$f_4(r) \equiv r\mathcal{K}_a/\operatorname{arth} r$ is decreasing from $(0,1)$ onto $(\sin(\pi a), \pi/2)$.*

(5) *For $b = (\sin(\pi a))/(2(1-a))$ and $0 < c \leq 2(1-a+a^2)$, the function $f_{5,c}(r) \equiv (\mathcal{E}_a - r'^2\mathcal{K}_a)/(1-r'^c)$ is decreasing from $(0,1)$ onto $(b, \pi a/c)$. In particular, $f_{5,3/2}$ is decreasing from $(0,1)$ onto $(b, 2\pi a/3)$.*

Proof. (1) By Theorem 4.1 (1),

$$rr'^{2-c}f'_1(r) = -cr^2\mathcal{K}_a + 2(1-a)(\mathcal{E}_a - r'^2\mathcal{K}_a),$$

which is nonpositive if and only if

$$c \geq 2(1-a)\sup_r \frac{\mathcal{E}_a - r'^2\mathcal{K}_a}{r^2\mathcal{K}_a} = 2a(1-a),$$

by Lemma 5.2 (4) . Finally, we note that since $\max\{2a(1-a) : 0 < a \leq \frac{1}{2}\} = \frac{1}{2}$, the function $\sqrt{r'}\mathcal{K}_a$ will be decreasing for each $a \in (0, \frac{1}{2}]$. The limiting value at $r = 0$ is obvious, while that at $r = 1$ follows from l'Hôpital's Rule, Theorem 4.1 (1), and Lemma 5.2 (1).

(2) By Theorem 4.1 (2), we have

$$rf'_2(r) = -cr'^{c-2}r^2\mathcal{E}_a + 2(a-1)r'^c(\mathcal{K}_a - \mathcal{E}_a),$$

which is nonnegative if and only if

$$-c \geq 2(1-a)\sup_r \frac{r'^2(\mathcal{K}_a - \mathcal{E}_a)}{r^2\mathcal{E}_a} = 2(1-a)^2,$$

by Lemma 5.2 (6). Finally, since $\max\{2(1-a)^2 : 0 < a \le \frac{1}{2}\} = 2$, it follows that $\mathcal{E}_a/r'^2$ will be increasing for each $a \in (0, \frac{1}{2}]$. The limiting values are clear from (1.9).

For (3) we may write

$$rf_3'(r) = -2(1-a)(\mathcal{K}_a - \mathcal{E}_a) + 2a(1-2a)r^2\mathcal{K}_a.$$

Hence, by Lemma 5.2 (4),

$$\frac{f_3'(r)}{2r\mathcal{K}_a} = 2a(1-a) - 1 + (1-a)\frac{\mathcal{E}_a - r'^2\mathcal{K}_a}{r^2\mathcal{K}_a} < 3a(1-a) - 1 \le -\frac{1}{4},$$

and so f_3 is decreasing on $(0,1)$. The limiting values are clear.

(4) Write $f_4(r) = g_4(r)/h_4(r)$, where $g_4(r) = r\mathcal{K}_a$ and $h_4(r) = \operatorname{arth} r$. Then $g_4(0) = h_4(0) = 0$ and $g_4'(r)/h_4'(r) = f_3(r)$, hence decreasing by (3). By l'Hôpital's Rule, we have $f_4(0+) = \mathcal{K}_a(0) = \pi/2$ and

$$f_4(1-) = \lim_{r\to 1-}[2(1-a)(\mathcal{E}_a - r'^2\mathcal{K}_a)] = 2(1-a)\mathcal{E}_a(1) = \sin(\pi a).$$

(5) The limit $f_{5,c}(1-)$ is clear by (1) and (1.9). For monotoneity write $f_{5,c}(r) = g_5(r)/h_5(r)$, where $g_5(r) = \mathcal{E}_a - r'^2\mathcal{K}_a$ and $h_5(r) = 1 - r'^c$. Then $g_5(0) = h_5(0) = 0$ and

$$\frac{g_5'(r)}{h_5'(r)} = \frac{2a}{c}r'^{2-c}\mathcal{K}_a,$$

and the monotoneity of $f_{5,c}$ follows from (1) and Lemma 5.1. By l'Hôpital's Rule, $f_{5,c}(0+) = \pi a/c$. Finally, since $2(1-a+a^2) \ge 3/2$, the result for $f_{5,3/2}$ follows. □

When $a = \frac{1}{2}$, parts (1), (2), (3), (4), (6) of the next theorem are contained in [**AVV**, Theorems 3.21 (3), 5.13 (2), 3.30 (1), 5.16 (1), (2)], respectively.

Theorem 5.5. *For each $a \in (0, \frac{1}{2}]$ the function*

(1) $f_1(r) \equiv \mathcal{K}_a(r)/\sin(\pi a) + \log r'$ *is decreasing from* $(0,1)$ *onto* $(R(a)/2, \pi/(2\sin(\pi a)))$.
(2) $f_2(r) \equiv \mu_a(r) + \log r$ *is decreasing and concave from* $(0,1)$ *onto* $(0, R(a)/2)$.
(3) $f_3(r) \equiv m_a(r) + \log r$ *is decreasing and concave from* $(0,1)$ *onto* $(0, R(a)/2)$.
(4) $f_4(r) \equiv \mu_a(r)/\log(1/r)$ *is increasing from* $(0,1)$ *onto* $(1,\infty)$.
(5) $f_5(r) \equiv \mu_a(r)\operatorname{arth} r$ *is increasing from* $(0,1)$ *onto* $(0, \pi^2/(4\sin^2(\pi a)))$.
(6) $f_6(r) \equiv \mu_a(r)/\log(e^{(R(a)/2)}/r)$ *is decreasing from* $(0,1)$ *onto* $(0,1)$.
(7) $f_7(r) = \mathcal{K}_a(r)(\mu_a(r) + \log r)$ *is increasing from* $(0,1)$ *onto* $(\pi R(a)/4, \pi^2/(4\sin(\pi a)))$.
(8) $f_8(r) = ((\operatorname{arth} r)/r)(\mu_a(r) + \log r)$ *is increasing from* $(0,1)$ *onto* $(R(a)/2, \pi^2/(4\sin^2(\pi a)))$.
(9) $f_9(r) \equiv [(R(a)/2) - (\mu_a(r) + \log r)]/r^2$ *is increasing from* $(0,1)$ *onto* $((1-2a+2a^2)/2, R(a)/2)$.

Proof. Part (1) follows from the fact that $B(a,b)F(a,b;a+b;r)+\log(1-r)$ is decreasing from $(0,1)$ onto $(R(a,b),B(a,b))$ [**ABRVV**, Theorem 1.3 (2)]. Parts (2), (3), and (6) were obtained in [**QVu2**, Corollary 3.12], [**QVu3**, Theorem 2.29 (1)], and [**QVu3**, Theorem 1.13 (3)], respectively, while parts (4) and (5) are implied by [**QVu3**, Theorem 1.13 (1)] and [**QVu2**, Theorem 1.28 (2)], respectively.

For (7), we write $f_7(r) = g_7(r)/h_7(r)$, where $g_7(r) = \mu_a(r) + \log r$ and $h_7(r) = 1/\mathcal{K}_a(r)$. Then $g_7(1-) = h_7(1-) = 0$ and, after simplification,

$$\frac{g_7'(r)}{h_7'(r)} = \frac{\left(\frac{\pi}{2}\right)^2 - (r'\mathcal{K}_a)^2}{2(1-a)(\mathcal{E}_a - r'^2\mathcal{K}_a)},$$

which is increasing by Lemma 5.2 (7). Hence, by Lemma 5.1, $f_7(r)$ is also increasing. Finally, $f_7(0+) = \pi R(a)/4$ follows from (2) and (1.9), while $f_7(1-) = \pi^2/(4\sin(\pi a))$, since

$$\mathcal{K}_a(r)\log r = (r'^2\mathcal{K}_a(r))\frac{\log r}{r'^2}$$

tends to 0 as r tends to 1, by Lemma 5.4 (1) and l'Hôpital's Rule.

Part (8) follows from (7) and Lemma 5.4 (4).

For (9), $f_9(r) = g_9(r)/h_9(r)$, where $g_9(r) = (R(a)/2) - (\mu_a(r) + \log r)$ and $h_9(r) = r^2$. Then $g_9(0) = h_9(0) = 0$, and

$$\frac{g_9'(r)}{h_9'(r)} = \frac{1}{2}\frac{1}{(r'\mathcal{K}_a)^2}\frac{(\pi^2/4) - (r'\mathcal{K}_a)^2}{r^2},$$

which is increasing, with $f_9(0+) = \pi^2(1-2a+2a^2)/4$, by Lemma 5.2 (14). The value $f_9(1-) = R(a)/2$ is clear. Hence, the result follows from Lemma 5.1. ☐

Theorem 5.6. *For each $a \in (0, \frac{1}{2})$, the function $f(x) \equiv \mu_a(1/\mathrm{ch}\ x)$ is increasing and concave from $(0,\infty)$ onto $(0,\infty)$. In particular, for $r, s \in (0,1)$,*

$$\mu_a\left(\frac{rs}{1+r's'}\right) \le \mu_a(r) + \mu_a(s) \le 2\mu_a\left(\frac{\sqrt{2rs}}{\sqrt{1+rs+r's'}}\right),$$

with equality in the second inequality if and only if $r = s$.

Proof. Let $r = 1/\mathrm{ch}\ x$ and $s = 1/\mathrm{ch}\ y$. Then

$$f'(x) = \frac{\pi^2}{4r'\mathcal{K}_a(r)^2},$$

which is positive and increasing in r by Lemma 5.4 (1), hence decreasing in x. Therefore, f is increasing and concave on $(0,\infty)$. In particular, we have

$f((x+y)/2) \geq (f(x)+f(y))/2$, with equality if and only if $x=y$. Now

$$\mathrm{ch}^2\left(\frac{x+y}{2}\right) = \frac{1+rs+r's'}{2rs}.$$

Hence

$$\frac{1}{2}(f(x)+f(y)) \leq f\left(\frac{x+y}{2}\right)$$

gives

$$\mu_a(r)+\mu_a(s) \leq 2\mu_a\left(\sqrt{\frac{2rs}{1+rs+r's'}}\right),$$

with equality if and only if $r=s$. Finally, since $f(0+)=0$ and since $f'(x)$ is decreasing in x, it follows from Lemma 5.1 that $f(x)/x$ is decreasing on $(0,\infty)$. Hence $f(x+y) \leq f(x)+f(y)$ by [**AVV**, Lemma 1.24]. □

Corollary 5.7. *For each* $a \in (0,\frac{1}{2}]$ *and* $r,s \in (0,1)$, $\mu_a(r)+\mu_a(s) \leq 2\mu_a(\sqrt{rs})$, *with equality if and only if* $r=s$.

Proof. Since μ_a is decreasing,

$$\mu_a\left(\sqrt{\frac{2rs}{1+rs+r's'}}\right) \leq \mu_a(\sqrt{rs}) \iff 1+rs+r's' \leq 2 \iff r's' \leq 1-rs$$

$$\iff 1-r^2-s^2+r^2s^2 \leq 1-2rs+r^2s^2 \iff (r-s)^2 \geq 0,$$

with equality at each step if and only if $r=s$. Hence the result follows from Theorem 5.6. □

6. Modular functions.

The next lemma gives some basic properties of the functions $\varphi_K^a(r)$ and $\eta_K^a(r)$, generalizing Theorem 10.5 (1) and Exercise 10.65 (13) in [**AVV**].

Lemma 6.1. *For each* $a \in (0,1/2]$, $K \in (0,\infty)$, $r \in (0,1)$, $x \in (0,\infty)$, *we have*

$$\varphi_K^a(r)^2 + \varphi_{1/K}^a(r')^2 = 1, \tag{1}$$

$$\eta_K^a(x)\eta_{1/K}^a(1/x) = 1. \tag{2}$$

Proof. Let $s=\varphi_K^a(r), u=\varphi_{1/K}^a(r')$. Then $\mu_a(s)=\mu_a(r)/K$ and $\mu_a(u)=K\mu_a(r')$. Hence, by (4.8),

$$\mu_a(s)\mu_a(u) = \mu_a(r)\mu_a(r') = \frac{\pi^2}{4\sin^2(\pi a)} = \mu_a(s)\mu_a(s').$$

Thus $s'=u$, so that (1) follows.

(2) Let $r=\sqrt{x/(x+1)}$. Then, by (1), $\eta_K^a(x)=(s/s')^2$ and $\eta_{1/K}^a(1/x)=(s'/s)^2$, so that (2) follows. □

The next lemma will be needed for the proofs of some of the main results stated in the Introduction. Parts (1), (2), (3) generalize Lemma 10.7 (3), (1), (2), respectively, in [**AVV**].

Lemma 6.2. *For each* $a \in (0, \frac{1}{2}], K \in (1, \infty)$, $r \in (0,1)$, *let* $s = \varphi_K^a(r)$ *and* $t = \varphi_{1/K}^a(r)$. *Then the function*

(1) $f(r) \equiv \mathcal{K}_a(s)/\mathcal{K}_a(r)$ *is increasing from* $(0,1)$ *onto* $(1,K)$,
(2) $g(r) \equiv s'\mathcal{K}_a(s)^2/(r'\mathcal{K}_a(r)^2)$ *is decreasing from* $(0,1)$ *onto* $(0,1)$,
(3) $h(r) \equiv s\mathcal{K}'_a(s)^2/(r\mathcal{K}'_a(r)^2)$ *is decreasing from* $(0,1)$ *onto* $(1,\infty)$,
(4) $F(r) \equiv \mathcal{K}_a(t)/\mathcal{K}_a(r)$ *is decreasing from* $(0,1)$ *onto* $(1/K,1)$,
(5) $G(r) \equiv t'\mathcal{K}_a(t)^2/(r'\mathcal{K}_a(r)^2)$ *is increasing from* $(0,1)$ *onto* $(1,\infty)$,
(6) $H(r) \equiv t\mathcal{K}'_a(t)^2/(r\mathcal{K}'_a(r)^2)$ *is increasing from* $(0,1)$ *onto* $(0,1)$.

Proof. By Theorem 4.1 (1), (7), we have

$$\begin{aligned}
&\mathcal{K}_a(r)^2 f'(r)\\
&= 2(1-a)\left[\mathcal{K}_a(r)\frac{\mathcal{E}_a(s)-s'^2\mathcal{K}_a(s)}{ss'^2}\frac{ds}{dr} - \mathcal{K}_a(s)\frac{\mathcal{E}_a(r)-r'^2\mathcal{K}_a(r)}{rr'^2}\right]\\
&= \frac{2(1-a)\mathcal{K}_a(s)}{rr'^2\mathcal{K}'_a(r)}\big[\mathcal{K}'_a(s)(\mathcal{E}_a(s)-s'^2\mathcal{K}_a(s))\\
&\qquad - \mathcal{K}'_a(r)(\mathcal{E}_a(r)-r'^2\mathcal{K}_a(r))\big].
\end{aligned}$$

Thus $f'(r)$ is positive since $s > r$ and

$$\mathcal{K}'_a(x)(\mathcal{E}_a(x)-x'^2\mathcal{K}_a(x)) = (x^2\mathcal{K}'_a(x))\frac{\mathcal{E}_a(x)-x'^2\mathcal{K}_a(x)}{x^2}$$

is increasing in x by Lemmas 5.4 (1) and 5.2 (1). Hence f is increasing. The limiting values are clear.

Next, $g'(r)$ is negative if and only if

$$\begin{aligned}
&\mathcal{K}'_a(s)[s^2\mathcal{K}_a(s) - 4(1-a)(\mathcal{E}_a(s)-s'^2\mathcal{K}_a(s))]\\
&\qquad - \mathcal{K}'_a(r)[r^2\mathcal{K}_a(r) - 4(1-a)(\mathcal{E}_a(r)-r'^2\mathcal{K}_a(r))]
\end{aligned}$$

is positive. This is true if

$$g_1(r) \equiv \mathcal{K}'_a(r)[r^2\mathcal{K}_a(r) - 4(1-a)(\mathcal{E}_a(r)-r'^2\mathcal{K}_a(r))]$$

is increasing. Now,

$$g_1(r) = r^2\mathcal{K}_a(r)\mathcal{K}'_a(r)\left[1-4(1-a)\frac{\mathcal{E}_a(r)-r'^2\mathcal{K}_a(r)}{r^2\mathcal{K}_a(r)}\right],$$

which is positive and increasing, by Lemmas 5.2 (4) and 5.4 (1).

(3) Since $h(r) = 1/g(s')$, $F(r) = 1/f(t)$, $G(r) = 1/g(t)$, and $H(r) = 1/h(t)$, parts (3)-(6) follow from parts (1) and (2). ☐

6.3. Proof of Theorem 1.14. Let $r = e^{-x}$ and $s = \varphi_K^a(r)$. Then

$$f'(x) = \frac{1}{K}\left(\frac{s'\mathcal{K}_a(s)}{r'\mathcal{K}_a(r)}\right)^2,$$

which is positive and increasing in x, by Lemma 6.2 (1), (2). Hence f is strictly increasing and convex. Thus $f((x+y)/2) \le (f(x)+f(y))/2$, so that

$$1/\varphi_K^a(e^{-(x+y)/2}) \le 1/(\varphi_K^a(e^{-x})\varphi_K^a(e^{-y}))^{1/2}.$$

Now putting $r = e^{-x}$ and $t = e^{-y}$ gives the result. The proof for $g(x)$ is similar. $\square$

6.4. Proof of Theorem 1.15. Let $t = rx$, $u = \varphi_K^a(t)$, and $s = \varphi_K^a(x)$. Then

$$\begin{aligned}\frac{f'(x)}{f(x)} &= \frac{1}{u}\frac{1}{K}\frac{ruu'^2\mathcal{K}_a(u)^2}{tt'^2\mathcal{K}_a(t)^2} - \frac{1}{s}\frac{1}{K}\frac{ss'^2\mathcal{K}_a(s)^2}{xx'^2\mathcal{K}_a(x)^2}\\ &= \frac{1}{Kx}\left[\left(\frac{u'\mathcal{K}_a(u)}{t'\mathcal{K}_a(t)}\right)^2 - \left(\frac{s'\mathcal{K}_a(s)}{x'\mathcal{K}_a(x)}\right)^2\right],\end{aligned}$$

which is positive by Lemma 6.2 (1), (2) since $t < x$. The proof for $g(x)$ is similar. $\square$

The following inequalities generalize identities satisfied by the functions μ and $\varphi_K(r)$ (cf. [**LV**, pp. 60, 61], [**AVV**, Theorem 10.5 (3), (7)]).

Theorem 6.5. *For $a \in (0, \frac{1}{2}]$, $r \in (0,1)$, $K \in (0,\infty)$,*

$$\frac{1}{2}\mu_a(r) \le \mu_a\left(\frac{2\sqrt{r}}{1+r}\right) \le \mu_a(r). \tag{1}$$

$$\mu_a(r') \le \mu_a\left(\frac{1-r}{1+r}\right) \le 2\mu_a(r'). \tag{2}$$

$$\varphi_K^a(r) \le \varphi_K^a\left(\frac{2\sqrt{r}}{1+r}\right) \le \varphi_{2K}^a(r). \tag{3}$$

$$\varphi_{K/2}(r') \le \varphi_K^a\left(\frac{1-r}{1+r}\right) \le \varphi_K^a(r'). \tag{4}$$

Proof. Inequalities (1) and (2) were obtained in [**QVu2**, Theorem 1.14 (1)], and (3), (4) follow from (1), (2), respectively if we divide by K and apply μ_a^{-1}. $\square$

Theorem 6.6. *For each $a \in (0, \frac{1}{2}]$, $r \in (0,1)$, let $f : (0,\infty) \to \mathbf{R}$ be defined by*

$$f(K) \equiv \frac{K\log r - \log\varphi_{1/K}^a(r)}{K-1}.$$

Then f is increasing, with $f(0+)=0$,

$$\lim_{K\to 1} f(K)=\log r+m_a(r), \quad \textit{and} \quad \lim_{K\to\infty} f(K)=\log r+\mu_a(r).$$

In particular, for $K\in(1,\infty)$, $r\in(0,1)$, $a\in(0,\frac{1}{2}]$,

$$\text{(1)} \quad 1<e^{(K-1)(m_a(r)+\log r)}<\frac{r^K}{\varphi^a_{1/K}(r)}<e^{(K-1)(\mu_a(r)+\log r)}<e^{(K-1)R(a)/2}$$

and

$$\text{(2)} \qquad 1<\frac{\varphi^a_K(r)}{r^{1/K}}<e^{(1-1/K)(\log r+m_a(r))}<e^{(1-1/K)R(a)/2}.$$

Proof. Since $\mu_a^{-1}(0+)=1$, it is clear that $f(0+)=0$. Now let $t=\varphi^a_{1/K}(r)$, and write $f(K)=g(K)/h(K)$, where $g(K)=K\log r-\log\varphi^a_{1/K}(r)$ and $h(K)=K-1$. Then $g(1)=h(1)=0$, and

$$\frac{g'(K)}{h'(K)}=\log r+\frac{4}{\pi^2}(t'\mathcal{K}_a(t))^2\mu_a(r),$$

which is increasing in K on $(0,\infty)$ by Lemma 5.4 (1) and the fact that t is decreasing in K. Hence f is increasing, by Lemma 5.1.

Next, by l'Hôpital's Rule,

$$\lim_{K\to 1} f(K)=\log r+\frac{2}{\pi\sin(\pi a)}r'^2\mathcal{K}_a\mathcal{K}'_a=\log r+m_a(r),$$

while

$$\lim_{K\to\infty} f(K)=\log r+\frac{4}{\pi^2}\mathcal{K}_a(0)^2\mu_a(r)=\log r+\mu_a(r).$$

Inequalities (1), (2) follow from the above argument and Theorems 5.5 (3), (2). □

The next theorem is a generalization of a result due to Hübner and He (cf. [**AVV**, Theorem 10.9 (1)]).

Theorem 6.7. *For each $a\in(0,\frac{1}{2}]$ and $K\in(1,\infty)$, let f,g be defined on $(0,1]$ by*

$$f(r)=r^{-1/K}\varphi^a_K(r) \quad \textit{and} \quad g(r)=r^{-K}\varphi^a_{1/K}(r).$$

Then f is decreasing and g is increasing, with $f((0,1])=[1,e^{(1-1/K)R(a)/2})$ and $g((0,1])=(e^{(1-K)R(a)/2},1]$.

Proof. First, let $s=\varphi^a_K(r)$. Then

$$\frac{f'(r)}{f(r)}=\frac{1}{Kr}\left(\left(\frac{s'\mathcal{K}_a(s)}{r'\mathcal{K}_a(r)}\right)^2-1\right),$$

which is negative by Lemma 6.2 (1), (2). Clearly $f(1) = 1$, while

$$\lim_{r\to 0}\log(r^{-1/K}s) = \lim_{r\to 0}\left[(\mu_a(s)+\log s) - \frac{1}{K}(\mu_a(r)+\log r)\right]$$
$$= \frac{R(a)}{2}\left(1-\frac{1}{K}\right),$$

by (1.5) and Theorem 5.5 (2).

Next, let $t = \varphi^a_{1/K}(r)$. Then $r = \varphi^a_K(t)$. Hence $g(r) = (t^{-1/K}\varphi^a_K(t))^{-K}$, so that the assertion about $g(r)$ follows from the properties of $f(r)$ already proved. □

Theorem 6.8. *For each $a \in (0, \frac{1}{2}]$ and $K \in (1,\infty)$, the function $f(r) \equiv \varphi^a_K(r)$ is increasing and concave from $(0,1)$ onto $(0,1)$, and $g(r) \equiv \varphi^a_{1/K}(r)$ is increasing and convex from $(0,1)$ onto $(0,1)$.*

Proof. By Theorem 4.1 (7), $f'(r) = (1/K)ss'^2\mathcal{K}_a(s)^2/(rr'^2\mathcal{K}_a(r)^2)$, which is positive and decreasing by Theorem 6.6 and Lemma 6.2 (1), (2). The proof for $g(r)$ is similar. □

Corollary 6.9. *For each $a \in (0, \frac{1}{2}]$, $K \in (1,\infty)$,*

(1) $$\lim_{r\to 0+}\frac{\partial\varphi^a_K(r)}{\partial r} = \infty, \quad \lim_{r\to 0+}\frac{\partial\varphi^a_{1/K}(r)}{\partial r} = 0,$$

(2) $$\lim_{r\to 1-}\frac{\partial\varphi^a_K(r)}{\partial r} = 0, \quad \lim_{r\to 1-}\frac{\partial\varphi^a_{1/K}(r)}{\partial r} = \infty.$$

Proof. Let $s = \varphi^a_K(r)$. From Theorems 4.1 (7) and 6.7 we have

$$\lim_{r\to 0+}\frac{\partial\varphi^a_K(r)}{\partial r} = \frac{1}{K}\lim_{r\to 0+}\frac{s}{r^{1/K}}\lim_{r\to 0+}r^{-1+1/K} = \infty$$

and

$$\lim_{r\to 1-}\frac{\partial\varphi^a_K(r)}{\partial r} = K\lim_{r\to 1-}\left(\frac{s'}{r'^K}\right)^2\lim_{r\to 1-}r'^{2(K-1)} = 0.$$

The proof for $\varphi^a_{1/K}(r)$ is similar. □

6.10. Proof of Theorem 1.16. Let $y = g(x) = p(\varphi^a_K(q(x))) = p(s)$, where $s = \varphi^a_K(r)$, $r = q(x)$. Then $q(y) = s$, so that $q'(y)dy/dx = ds/dx = (ds/dr)(dr/dx)$. Thus

$$q'(y)g'(x) = \frac{1}{K}\frac{ss'^2\mathcal{K}_a(s)^2}{rr'^2\mathcal{K}_a(r)^2}\frac{dr}{dx}$$

by Theorem 4.1 (7). Now

$$r^2 = q(x)^2 = \frac{e^x}{e^x+1} = 1 - \frac{1}{e^x+1},$$

giving

$$2r\frac{dr}{dx} = r^2 r'^2.$$

Thus $q'(x) = rr'^2/2$ and $q'(y) = ss'^2/2$. Hence, we get

$$g'(x) = \frac{1}{K}\frac{\mathcal{K}_a(s)^2}{\mathcal{K}_a(r)^2},$$

which increases from $1/K$ to K, by Lemma 6.2 (1). The proof for $h(x)$ is similar. $\square$

Remark 6.11. Let p, q, g, h be the functions defined in Theorem 1.16, and let $s = \varphi_K^a(1/\sqrt{2})$. Then, by Theorem 1.16 and integration, we obtain

$$\frac{x}{K} + 2\log\frac{s}{s'} \le g(x) \le Kx + 2\log\frac{s}{s'},$$
$$\frac{x}{K} - 2\log\frac{s}{s'} \le h(x) \le Kx - 2\log\frac{s}{s'}.$$

In the next result we obtain further bounds for the functions g and h.

Theorem 6.12. *Let $p : (0,1) \to (-\infty,\infty)$ and $q : (-\infty,\infty) \to (0,1)$ be given by $p(x) = 2\log(x/x')$ and $q(x) = p^{-1}(x) = \sqrt{e^x/(e^x+1)}$, respectively, and for $a \in (0, \frac{1}{2}]$, $K \in (1,\infty)$, let $g, h : (-\infty,\infty) \to (-\infty,\infty)$ be defined by $g(x) = p(\varphi_K^a(q(x)))$ and $h(x) = p(\varphi_{1/K}^a(q(x)))$. Then*

$$g(x) \ge \begin{cases} Kx, & x \ge 0, \\ \dfrac{x}{K}, & x < 0, \end{cases} \qquad \text{and} \qquad h(x) \le \begin{cases} \dfrac{x}{K}, & x \ge 0, \\ Kx, & x < 0. \end{cases}$$

Proof. We only give the proof for g, since the proof for h is similar. First, if $x > 0$, then

$$g(x) \ge Kx \Leftrightarrow \varphi_K^a(q(x)) \ge q(Kx)$$

$$\Leftrightarrow \mu_a^{-1}\left(\frac{1}{K}\mu_a\left(\sqrt{\frac{e^x}{e^x+1}}\right)\right) \ge \sqrt{\frac{e^{Kx}}{e^{Kx}+1}}$$

$$\Leftrightarrow \mu_a\left(\sqrt{\frac{e^x}{e^x+1}}\right) \le K\mu_a\left(\sqrt{\frac{e^{Kx}}{e^{Kx}+1}}\right).$$

This will be true if $f(K) \equiv K\mu_a(\sqrt{e^{Kx}/(e^{Kx}+1)})$ is increasing on $[1,\infty)$.

Now, with $r = \sqrt{e^{Kx}/(e^{Kx}+1)}$, $r^2 = e^{Kx}/(e^{Kx}+1)$, $r'^2 = 1/(e^{Kx}+1)$, we have

$$2r\frac{dr}{dK} = \frac{xe^{Kx}}{(e^{Kx}+1)^2} = xr^2r'^2, \quad \frac{dr}{dK} = \frac{xrr'^2}{2}.$$

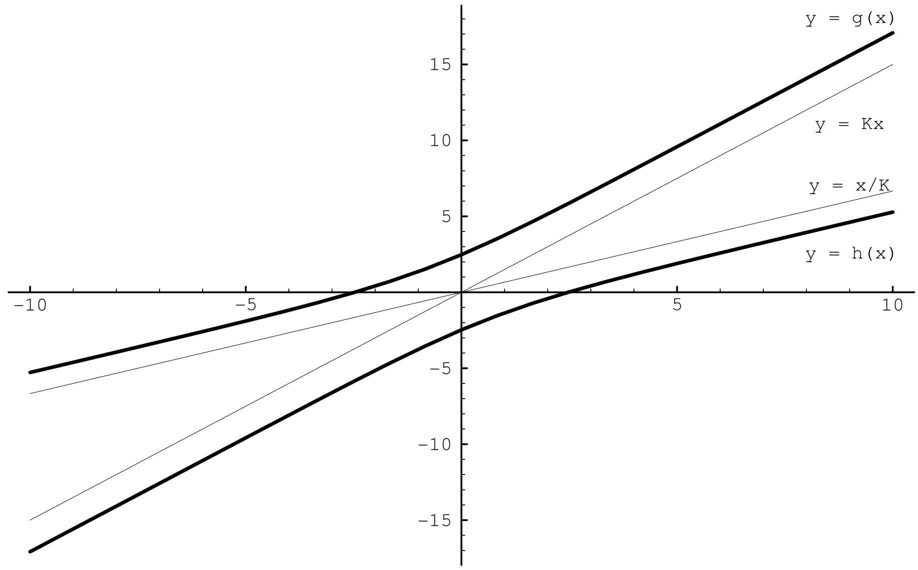

Figure 1. The graphs of the functions g, h, and those of the lines $Kx, x/K$, in Theorem 6.12, where $a = 0.25$ and $K = 1.5$.

Hence

$$\begin{aligned} f'(K) &= \mu_a(r) - \frac{K\pi^2}{4rr'^2\mathcal{K}_a(r)^2}\frac{dr}{dK} \\ &= \mu_a(r) - \frac{\pi^2 Kx}{8\mathcal{K}_a(r)^2} = \frac{\pi}{2}\left[\frac{\mathcal{K}'_a(r)}{\mathcal{K}_a(r)\sin(\pi a)} - \frac{\pi Kx}{4\mathcal{K}_a(r)^2}\right], \end{aligned}$$

and

$$\begin{aligned} &\frac{\pi}{2\mathcal{K}_a(r)^2\sin(\pi a)}[\mathcal{K}_a(r)\mathcal{K}'_a(r) - \frac{\pi}{4}Kx\sin(\pi a)] > 0 \\ &\Leftrightarrow \mathcal{K}_a(r)\mathcal{K}'_a(r) > \frac{\pi}{4}(Kx)\sin(\pi a) \\ &\Leftrightarrow \mathcal{K}_a(r)\mathcal{K}'_a(r) > \frac{\pi}{2}\left(\log\frac{r}{r'}\right)\sin(\pi a), \end{aligned}$$

for $r \in (1/\sqrt{2}, 1)$. Now from (4.9) and Theorem 5.5 (3),

$$\begin{aligned} \mathcal{K}_a(r)\mathcal{K}'_a(r) &= \frac{\pi}{2}(m_a(r) + m_a(r'))\sin(\pi a) \\ &> \frac{\pi}{2}\left(\log\frac{1}{rr'}\right)\sin(\pi a) > \frac{\pi}{2}\left(\log\frac{r}{r'}\right)\sin(\pi a), \end{aligned}$$

which proves the first inequality for $g(x)$.

Next, $g(-x) \geq -x/K, x > 0$

$$\Leftrightarrow \varphi_K^a\left(\sqrt{\frac{e^{-x}}{e^{-x}+1}}\right) \geq \sqrt{\frac{e^{-x/K}}{e^{-x/K}+1}}, \quad x > 0$$

$$\Leftrightarrow \mu_a^{-1}\left(\frac{1}{K}\mu_a\left(\sqrt{\frac{1}{e^x+1}}\right)\right) \geq \sqrt{\frac{1}{e^{x/K}+1}}$$

$$\Leftrightarrow \mu_a\left(\frac{1}{\sqrt{e^x+1}}\right) \leq K\mu_a\left(\sqrt{\frac{1}{e^{x/K}+1}}\right).$$

This will be true if $F(K) \equiv K\mu_a(1/\sqrt{e^{x/K}+1})$ is increasing on $[1,\infty)$. Let $1/\sqrt{e^{x/K}+1} = t$, so that $t \in (0, 1/\sqrt{2})$. Now $t^2 = 1/(e^{x/K}+1)$, $t'^2 = e^{x/K}/(e^{x/K}+1)$, and

$$\begin{aligned} F'(K) &= \mu_a(t) - \frac{K\pi^2}{4tt'^2\mathcal{K}_a(t)}\frac{dt}{dK} \\ &= \frac{\pi}{2\sin(\pi a)}\frac{\mathcal{K}_a'(t)}{\mathcal{K}_a(t)} - \frac{K\pi^2}{4tt'^2\mathcal{K}_a(t)^2}\frac{dt}{dK}. \end{aligned}$$

Now

$$2t\frac{dt}{dK} = \frac{x}{K^2}\frac{e^{x/K}}{(e^{x/K}+1)^2} = \frac{x}{K^2}t^2t'^2, \quad \frac{dt}{dK} = \frac{x}{K^2}\frac{tt'^2}{2}.$$

Hence

$$\begin{aligned} F'(K) &= \frac{\pi}{2\sin(\pi a)}\frac{\mathcal{K}_a'(t)}{\mathcal{K}_a(t)} - \frac{\pi^2 x}{8K\mathcal{K}_a(t)^2} \\ &= \frac{\pi}{2(\sin(\pi a))\mathcal{K}_a(t)^2}\left[\mathcal{K}_a(t)\mathcal{K}_a'(t) - \frac{\pi(\sin(\pi a))x}{4K}\right] \\ &= \frac{\pi}{2(\sin(\pi a))\mathcal{K}_a(t)^2}\left[\mathcal{K}_a(t)\mathcal{K}_a'(t) - \frac{\pi}{2}\left(\log\frac{t'}{t}\right)\sin(\pi a)\right]. \end{aligned}$$

Hence $F'(K) > 0$ if and only if $\mathcal{K}_a(t)\mathcal{K}_a'(t) > (\pi/2)(\sin(\pi a))\log(t'/t)$. Now, by (4.9) and Theorem 5.5 (3), we have

$$\begin{aligned} \mathcal{K}_a(t)\mathcal{K}_a'(t) &= \frac{\pi\sin(\pi a)}{2}(m_a(t) + m_a(t')) \\ &> \frac{\pi}{2}\left(\log\frac{1}{tt'}\right)\sin(\pi a) > \frac{\pi}{2}\left(\log\frac{t'}{t}\right)\sin(\pi a). \end{aligned}$$

$\square$

Theorem 6.13. *For each $a \in (0, \frac{1}{2}]$, $K \in (1,\infty)$, let f and g be defined on $(0,1)$ by*

$$f(r) = \frac{\arcsin(\varphi_K^a(r))}{\arcsin(r^{1/K})}, \ g(r) = \frac{\arcsin(\varphi_{1/K}^a(r))}{\arcsin(r^K)}.$$

Then f is decreasing and g is increasing, with $f((0,1]) = [1, e^{(1-1/K)R(a)/2})$ and $g((0,1]) = (e^{(1-K)R(a)/2}, 1]$.

Proof. To prove the monotoneity for f, let $s = \varphi_K^a(r)$ and $f(r) = f_1(r)/f_2(r)$, where $f_1(r) = \arcsin(s)$ and $f_2(r) = \arcsin(r^{1/K})$. Then $f_1(0) = f_2(0) = 0$, and

$$\frac{f_1'(r)}{f_2'(r)} = \frac{s}{r^{1/K}} \left(\frac{1-r^{2/K}}{1-r^2}\right)^{1/2} \frac{s'\mathcal{K}_a(s)^2}{r'\mathcal{K}_a(r)^2},$$

which is decreasing, by Lemma 6.2 (2) and Theorem 6.7. The limiting values follow from l'Hôpital's Rule and Theorem 6.7. The proof for g is similar. □

Theorem 6.14. *For each $a \in (0, \frac{1}{2}]$, $K \in (1,\infty)$, let f and g be defined on $(0,\infty)$ by*

$$f(x) = \operatorname{arth} \varphi_K^a(\operatorname{th} x) \quad \text{and} \quad g(x) = \operatorname{arth} \varphi_{1/K}^a(\operatorname{th} x).$$

Then f and g are increasing automorphisms; f is concave and g is convex. In particular,

$$f(x+y) \le f(x) + f(y) \le 2f\left(\frac{x+y}{2}\right), \tag{6.15}$$

and

$$g(x+y) \ge g(x) + g(y) \ge 2g\left(\frac{x+y}{2}\right), \tag{6.16}$$

for all $x, y \in (0,\infty)$.

Proof. For f, let $r = \operatorname{th} x$ and $s = \varphi_K^a(r)$. Then

$$f'(x) = \frac{1}{s'^2} \ K \ \frac{ss'^2 \ \mathcal{K}_a'(s)^2}{rr'^2 \ \mathcal{K}_a'(r)^2} \ r'^2 = K \frac{s\mathcal{K}_a'(s)^2}{r\mathcal{K}_a'(r)^2},$$

which is positive and decreasing, by Lemma 6.2 (3). Hence f is increasing and concave, so $f(x)/x$ is decreasing by Lemma 5.1. The double inequality (6.15) follows from [**AVV**, Lemma 1.24] and the concavity of f. The proof for g is similar. □

Remark 6.17 (Cf. [**AVV**, Theorem 10.12])**.** The inequalities in Theorem 6.14 can be simplified to

$$\varphi_K^a\left(\frac{r+s}{1+rs}\right) \le \frac{\varphi_K^a(r) + \varphi_K^a(s)}{1 + \varphi_K^a(r)\varphi_K^a(s)}$$

and

$$\frac{\varphi_K^a(r)+\varphi_K^a(s)}{1+\varphi_K^a(r)\varphi_K^a(s)+\varphi_{1/K}^a(r')\varphi_{1/K}^a(s')} \le \varphi_K^a\left(\frac{r+s}{1+rs+r's'}\right),$$

for $a \in (0, \frac{1}{2}], r, s, \in (0,1)$, and $K \in (1,\infty)$. The inequalities are reversed if we replace K by $1/K$.

Theorem 6.18. *For $a \in (0, \frac{1}{2}],\ r \in (0,1),\ K \in (1,\infty)$,*

$$\mathrm{th}(K \text{ arth } r) < \varphi_K^a(r).$$

The inequality is reversed if we replace K by $1/K$.

Proof. Let $s = \varphi_K^a(r)$. Then $s > r$, and it follows from (1.5) and Theorem 5.5 (5) that

$$\frac{1}{K}\mu_a(r)\text{arth } s = \mu_a(s)\,\text{arth } s > \mu_a(r)\,\text{arth } r,$$

and the result follows if we solve for s.

The reverse inequality follows if we let $\varphi_K^a(r) = x$ in the inequality just proved and solve for r. □

The next result generalizes Theorem 10.24 in [**AVV**].

Theorem 6.19. *For each $a \in (0, \frac{1}{2}]$ and $K \in (1,\infty)$, the function*

$$f(x) \equiv \frac{\log(\eta_K^a(x)/\eta_K^a(1))}{\log x}$$

is increasing from $(0,\infty)$ onto $(1/K, K)$, while

$$g(x) \equiv \frac{\log(\eta_{1/K}^a(x)/\eta_{1/K}^a(1))}{\log x}$$

is decreasing from $(0,\infty)$ onto $(1/K, K)$. In particular,

$$\eta_K^a(1)\min\{x^K, x^{1/K}\} \le \eta_K^a(x) \le \eta_K^a(1)\max\{x^K, x^{1/K}\},$$
$$\eta_{1/K}^a(1)\min\{x^K, x^{1/K}\} \le \eta_{1/K}^a(x) \le \eta_{1/K}^a(1)\max\{x^K, x^{1/K}\}$$

for all $x \in (0,\infty),\ a \in (0, \frac{1}{2}],\ K \in [1,\infty)$.

Proof. Let $r = \sqrt{x/(x+1)},\ s = \varphi_K^a(r)$, and $f(x) = G(x)/H(x)$, where $G(x) = \log(\eta_K^a(x)/\eta_K^a(1))$ and $H(x) = \log x$. Then $G(1) = H(1) = 0$ and

$$\frac{G'(x)}{H'(x)} = \frac{1}{K}\left(\frac{\mathcal{K}_a(s)}{\mathcal{K}_a(r)}\right)^2,$$

which is increasing from $(0,\infty)$ onto $(1/K, K)$ by Lemma 6.2 (1). Hence f is increasing by Lemma 5.1, while $f(0+) = 1/K$ and $\lim_{x\to\infty} f(x) = K$ follow from l'Hôpital's Rule. The proof for $g(x)$ is similar. □

Theorem 6.20. *For $a \in (0, \frac{1}{2}]$, $x \in (0, \infty)$, $K \in [1, \infty)$, we have*

$$1 \leq \frac{\eta_K^a(x)}{x^{1/K}(x+1)^{K-1/K}} \leq e^{R(a)(K-1/K)},$$

where $R(a)$ is as in (3.4)*.*

Proof. By (1.13) and Lemma 6.1 (1) we have $\eta_K^a(x) = (\varphi_K^a(r)/\varphi_{1/K}^a(r'))^2$, where $r = \sqrt{x/(x+1)}$ and $r' = 1/\sqrt{x+1}$. Next,

$$r^{1/K} \leq \varphi_K^a(r) \leq e^{(1-1/K)R(a)/2}\, r^{1/K}$$

and

$$e^{(1-K)R(a)/2}\, r'^{\,K} \leq \varphi_{1/K}^a(r') \leq r'^{\,K}$$

by Theorem 6.6. Hence the result follows when we divide and substitute for r in terms of x. □

Theorem 6.21. *For each $a \in (0, \frac{1}{2}]$, $K \in (1, \infty)$, let f, g be defined on $(0,1)$ by*

$$f(r) = \varphi_K^a(r)\ \mathrm{ch}\left(\frac{1}{K}\mathrm{arch}\left(\frac{1}{r}\right)\right),$$

and

$$g(r) = \varphi_{1/K}^a(r)\ \mathrm{ch}\left(K\mathrm{arch}\left(\frac{1}{r}\right)\right).$$

Then f is decreasing and g is increasing, with $f((0,1)) = (1, e^{(1-1/K)R(a)/4})$ and $g((0,1)) = (e^{(1-K)R(a)/4}, 1)$.

Proof. To prove the assertion for f, we first let $s = \varphi_K^a(r)$ and $1/t = \mathrm{ch}((1/K)\ \mathrm{arch}\ (1/r))$. Then $s' = \varphi_{1/K}^a(r')$ and $t' = \mathrm{th}((1/K)\ \mathrm{arth}\ (r'))$. Now, by differentiation and simplification,

$$Ktrr'f'(r) = st'\left[\frac{s'}{t'}\ \frac{s'\mathcal{K}_a(s)^2}{r'\mathcal{K}_a(r)^2} - 1\right],$$

which is negative, by Lemma 6.2 (2) and Theorem 6.18. Hence f is decreasing. The limiting values follow from the expression

$$\mathrm{ch}\left(\frac{1}{K}\ \mathrm{arch}\ \left(\frac{1}{r}\right)\right) = \frac{(1+r')^{1/K} + (1-r')^{1/K}}{2r^{1/K}},$$

and Theorem 6.7. The proof for g is similar. □

Theorem 6.22. *Let $\mathcal{M}$ be a set with a multiplication operation and let $\{f_K :\ K \in \mathcal{M}\}$ be a collection of increasing functions $f_K : (0,1) \to (0,1)$, such that $f_K \circ f_L = f_{KL}$ and such that $f_K(x)/x$ is decreasing (respectively, increasing). Then*

$$f_K(x)f_L(y) \leq f_{KL}(xy),$$

(respectively, $f_K(x)f_L(y) \geq f_{KL}(xy)$).

Proof. Let $f_K(x)/x$ be decreasing for each $K \in \mathcal{M}$. Since $xy \le x$, we have $f_K(x)/x \le f_K(xy)/(xy)$. Thus $f_K(x)y \le f_K(xy)$. Hence

$$f_K(x)f_L(y) \le f_K(xf_L(y)) \le f_K(f_L(xy)) = f_{KL}(xy).$$

The other case is similar. □

Corollary 6.23. *For $a \in (0, \frac{1}{2}]$, $x, y \in (0,1)$, $K, L \in [1,\infty)$,*

$$\varphi_K^a(x)\varphi_L^a(y) \le \varphi_{KL}^a(xy), \tag{1}$$

$$\varphi_{1/K}^a(x)\varphi_{1/L}^a(y) \ge \varphi_{1/(KL)}^a(xy). \tag{2}$$

7. Dependence on *a*.

In this section we study the generalized elliptic integrals as functions of the parameter a.

Lemma 7.1. *For each nonnegative integer n, the function $f_n(x) \equiv -(-x,\ n+1)(x, n+1)$ is positive and increasing on $[0, \frac{1}{2}]$. The function $g_n(x) \equiv (x, n+1)(1-x, n+1)$ is positive, increasing on $[0, \frac{1}{2}]$ and decreasing on $[\frac{1}{2}, 1]$. For $n \ge 1$, the function f_n is positive and decreasing on $[1/\sqrt{2}, 1]$. But f_0 is positive and increasing on $[0,1]$.*

Proof. We have

$$f_n(x) = x^2(1-x^2)(2^2-x^2)\cdots(n^2-x^2).$$

On $[1/\sqrt{2}, 1]$, $x^2(1-x^2)$ is decreasing, hence so is f_n if $n \ge 1$. Next, on $(0, \frac{1}{2})$,

$$\frac{f_n'(x)}{2f_n(x)} = \frac{1}{x} - \sum_{k=1}^{n} \frac{x}{k^2-x^2},$$

which is clearly decreasing. Hence

$$\frac{f_n'(x)}{2f_n(x)} > 2 - 2\sum_{k=1}^{n} \frac{1}{4k^2-1} = 2 - \left(1 - \frac{1}{2n+1}\right) > 0.$$

Clearly, $f_0(x) = x^2$, which is increasing on $[0,1]$. Since $g_n(x) = g_n(1-x)$, we need only prove the second assertion on $[0, \frac{1}{2}]$. Now

$$\frac{g_n'(x)}{g_n(x)} = \frac{1}{x} - \frac{1}{n+1-x} - 2\sum_{k=1}^{n} \frac{x}{k^2-x^2},$$

which is clearly decreasing. Thus for $x \in (0, \frac{1}{2})$,

$$\frac{g_n'(x)}{g_n(x)} > 2 - \frac{2}{2n+1} - \sum_{k=1}^{n} \frac{4}{4k^2-1} = 0.$$

□

Theorem 7.2. *For each $r \in (0,1)$, let f, g be defined on $[0,1]$, by $f(a) = F(a-1, 1-a; 1; r^2)$ and $g(a) = F(a, 1-a; 1; r^2)$.*

(1) *If $\frac{1}{2} \le a < b \le 1$, then all coefficients are positive in the Taylor series for $f(b) - f(a)$ in powers of r^2.*
(2) *If $0 \le a < b \le 1 - 1/\sqrt{2}$, then all coefficients are negative in the Taylor series for $f(b) - f(a) - (b-a)(2-a-b)r^2$ in powers of r^2.*
(3) *If $0 \le a < b \le \frac{1}{2}$ (respectively, $\frac{1}{2} \le a < b \le 1$), then all coefficients are positive (respectively, negative) in the Taylor series for $g(b) - g(a)$ in powers of r^2.*

Proof. (1) In this case,

$$f(b) - f(a) = \sum_{n=1}^{\infty} [(1-b,n)(b-1,n) - (1-a,n)(a-1,n)] \frac{r^{2n}}{(n!)^2}.$$

Now $-(1-a,n)(a-1,n) - (-(1-b,n)(b-1,n)) > 0$, by Lemma 7.1.

(2) In this case, $1 \ge 1-a > 1-b \ge 1/\sqrt{2}$. Hence

$$(1-b,n)(b-1,n) - (1-a,n)(a-1,n) < 0,$$

for all $n \ge 2$, by Lemma 7.1.

(3) We write

$$g(b) - g(a) = \sum_{n=1}^{\infty} [(b,n)(1-b,n) - (a,n)(1-a,n)] \frac{r^{2n}}{(n!)^2}.$$

Now, $(b,n)(1-b,n) - (a,n)(1-a,n)$ is positive or negative according as $0 \le a < b \le \frac{1}{2}$, or $\frac{1}{2} \le a < b \le 1$, by Lemma 7.1. □

Corollary 7.3. *For each $r \in (0,1)$, the function*

(1) *$f(a) \equiv \mathcal{E}_{1-a}(r)$ is decreasing from $[0, \frac{1}{2}]$ onto $[\mathcal{E}(r), \pi/2]$.*
(2) *$g(a) \equiv \mathcal{K}_a(r)$ is increasing from $[0, \frac{1}{2}]$ onto $[\pi/2, \mathcal{K}(r)]$.*
(3) *$h(a) \equiv \mathcal{E}_a(r)$ is increasing from $[0, \frac{1}{2}]$ onto $[\pi r'^2/2, \mathcal{E}(r)]$.*

Proof. Parts (1) and (2) follow from (1.9), (1.10), and Theorem 7.2. Next, by Theorem 4.1 (4), $(d/dr)(\mathcal{E}_a(r) - r'^2\mathcal{K}_a(r)) = 2ar\mathcal{K}_a(r)$, which is increasing in a by (2). Hence, if $0 < a < b \le \frac{1}{2}$,

$$\int_0^r \frac{d}{dt}(\mathcal{E}_a(t) - t'^2\mathcal{K}_a(t))dt < \int_0^r \frac{d}{dt}(\mathcal{E}_b(t) - t'^2\mathcal{K}_b(t))dt,$$

so that

$$\mathcal{E}_a(r) - r'^2\mathcal{K}_a(r) < \mathcal{E}_b(r) - r'^2\mathcal{K}_b(r)$$

by Lemma 5.2 (1). Thus

$$\mathcal{E}_a(r) < \mathcal{E}_b(r) + r'^2(\mathcal{K}_a(r) - \mathcal{K}_b(r)) < \mathcal{E}_b(r)$$

by (2). □

Theorem 7.4. (1) *For each fixed* $r \in (0,1)$*, the function* $f(a,r) \equiv \mu_a(r)$ *is decreasing in* a *from* $(0, \frac{1}{2}]$ *onto* $[\mu(r), \infty)$*.*

(2) *For each fixed* $x \in (0,\infty)$*, the function* $g(a,x) \equiv \mu_a^{-1}(x)$ *is decreasing in* a *from* $(0, \frac{1}{2}]$ *onto* $[\mu^{-1}(x), 1)$*.*

(3) *For each fixed* $r \in (0,1)$ *and* $K \in (1,\infty)$*, the function* $h(a,r) \equiv \varphi_K^a(r)$ *is decreasing in* a *from* $(0, \frac{1}{2}]$ *onto* $[\varphi_K(r), 1)$*. Moreover, the function* $h_1(a,r) \equiv \varphi_{1/K}^a(r)$ *is increasing in* a *from* $(0, \frac{1}{2}]$ *onto* $(0, \varphi_{1/K}(r)]$*.*

Proof. (1) Part (1) was obtained in [**QVu2**, Theorem 1.22]. However, we give here a different proof for the monotoneity of f in a. It follows from Theorem 4.1 (5) and Corollary 7.3 (2) that $d\mu_a/dr$ is increasing in a on $(0, \frac{1}{2}]$. Hence, for $0 < a < b \le \frac{1}{2}$,

$$\mu_a(1-) - \mu_a(r) = \int_r^1 \frac{d\mu_a(t)}{dt} dt < \int_r^1 \frac{d\mu_b(t)}{dt} dt = \mu_b(1-) - \mu_b(r).$$

This implies that $\mu_a(r) > \mu_b(r)$, since $\mu_a(1-) = 0 = \mu_b(1-)$.

(2) Let $r = g(a,x) = \mu_a^{-1}(x)$. Then $x = \mu_a(r) = f(a,r)$, and

$$0 = \frac{dx}{da} = \frac{\partial f}{\partial a} + \frac{\partial f}{\partial r}\frac{dr}{da}.$$

Since $\partial f/\partial a < 0$ by part (1) (see also [**QVu2**, 3.1]), it follows from Theorem 4.1 (5) that $dr/da = -(\partial f/\partial a)/(\partial f/\partial r) < 0$, and the monotoneity of g in a follows. Clearly, $g(\frac{1}{2}, x) = \mu^{-1}(x)$. Fix $x \in (0,\infty)$. Suppose $g(0+, x) = r_0 \in (0,1)$. Choose $\epsilon > 0$ such that $(r_0 - \epsilon, r_0 + \epsilon) \subset (0,1)$. Then there exists $\delta \in (0,1)$ such that $g(a,x) < r_0 + \epsilon$ for all $a \in (0,\delta)$. Hence $x = \mu_a(g(a,x)) > \mu_a(r_0 + \epsilon)$. If we now let $a \to 0+$, we get $x = \infty$, a contradiction. Thus,

$$g(0+, x) = 1. \tag{7.5}$$

(3) Put $s = h(a,r)$, $a_n = (a,n)(1-a,n)$, and $b_n = P(a,n) - P(a,0)$ for $n = 0,1,2,\ldots$, where $P(a,n) = \psi(a+n) - \psi(1-a+n)$, and let $Q(a,r) = 2\mathcal{K}_a(r)/\pi = F(a, 1-a; 1; r^2)$. Then, by (1.3), (1.5), and (1.2),

$$\frac{Q(a,s')}{Q(a,s)} = \frac{1}{K}\frac{Q(a,r')}{Q(a,r)}, \tag{7.6}$$

and

$$\frac{\partial Q(a,r)}{\partial r} = 2\sum_{n=1}^{\infty} \frac{n a_n}{(n!)^2} r^{2n-1}, \qquad \frac{\partial Q(a,r)}{\partial a} = \sum_{n=1}^{\infty} \frac{a_n b_n}{(n!)^2} r^{2n}. \tag{7.7}$$

By (7.7), it follows from (7.6) and logarithmic differentiation with respect to a that

$$\frac{1}{Q(a,s')}\left[\frac{\partial Q(a,s')}{\partial a} - \frac{s}{s'}\frac{\partial Q(a,s')}{\partial s'}\frac{ds}{da}\right]$$
$$- \frac{1}{Q(a,s)}\left[\frac{\partial Q(a,s)}{\partial a} + \frac{\partial Q(a,s)}{\partial s}\frac{ds}{da}\right]$$

$$= \frac{1}{Q(a,r')}\frac{\partial Q(a,r')}{\partial a} - \frac{1}{Q(a,r)}\frac{\partial Q(a,r)}{\partial a},$$

and hence, by simplification,

$$\begin{aligned} (7.8) \qquad & \left[\frac{s}{s'Q(a,s')}\frac{\partial Q(a,s')}{\partial s'} + \frac{1}{Q(a,s)}\frac{\partial Q(a,s)}{\partial s}\right]\frac{ds}{da} \\ & = [Q_1(a,s') - Q_1(a,r')] + [Q_1(a,r) - Q_1(a,s)], \end{aligned}$$

where

$$(7.9) \qquad Q_1(a,x) = \frac{1}{Q(a,x)}\frac{\partial Q(a,x)}{\partial a} = \frac{\sum_{n=1}^{\infty}\alpha_n x^{2n}}{\sum_{n=0}^{\infty}\beta_n x^{2n}} = \frac{\sum_{n=0}^{\infty}\alpha_n x^{2n}}{\sum_{n=0}^{\infty}\beta_n x^{2n}},$$

and where $\alpha_n = a_n b_n (n!)^{-2}$ and $\beta_n = a_n (n!)^{-2}$. Since

$$\frac{\alpha_n}{\beta_n} = b_n = \psi(1-a) - \psi(a) - (1-2a)\sum_{k=0}^{\infty}\frac{1}{(a+n+k)(1-a+n+k)}$$

(cf. [**Ah**, p. 198]), which is clearly increasing in n, it follows from [**PV**, Lemma 2.1] that $Q_1(a,x)$ is increasing in x on $(0,1)$. Since $\partial Q(a,x)/\partial x > 0$ by Theorem 4.1 (1), and since $s > r$ ($s < r$, respectively) for $K > 1$ ($K < 1$, respectively), the monotoneity properties of h and h_1 in a follow from (7.8).

In case $K = 1$, we have $\varphi_K^a(r) = r$ for all $a \in (0,1)$. Next, supppose $K > 1$. Take $0 < L = \varphi_K^{0+}(r) < 1$, and choose $\epsilon > 0$ such that $(L-\epsilon, L+\epsilon) \subset (0,1)$. Then there exists $\delta \in (0,1)$ such that $a \in (0,\delta)$ implies that $\varphi_K^a(r) < L + \epsilon = t$, say. Hence, $1/K > \mu_a(t)/\mu_a(r)$. If we let $a \to 0+$, we get $1/K \geq 1$, a contradiction. Hence $L = 1$. Finally, if $0 < K < 1$, then the assertion follows from the result for the case $K > 1$ and Lemma 6.1 (1). □

Acknowledgments. This work was carried out during the visits of the third and fourth authors to Michigan State University, USA, and that of the third author to the University of Helsinki, Finland, and the Hangzhou Institute of Electronics Engineering, China. The authors would like to thank these institutions and the University of Auckland, the Academy of Finland, and the Finnish Mathematical Association for hospitality and support.

We are indebted to R. Askey for discussions and advice concerning Section 3, and to F. Garvan for bringing reference [**Mi**] to our attention.

References

[AS] M. Abramowitz and I.A. Stegun, editors, *Handbook of Mathematical Functions with Formulas, Graphs and Mathematical Tables*, Dover, New York, 1965.

[Ah] L.V. Ahlfors, *Complex Analysis*, 2nd ed., McGraw-Hill, New York, 1966.

[ABRVV] G.D. Anderson, R.W. Barnard, K.C. Richards, M.K. Vamanamurthy and M. Vuorinen, *Inequalities for zero-balanced hypergeometric functions*, Trans. Amer. Math. Soc., **347** (1995), 1713-1723.

[AVV] G.D. Anderson, M.K. Vamanamurthy and M. Vuorinen, *Conformal Invariants, Inequalities, and Quasiconformal Maps*, John Wiley & Sons, New York, 1997.

[Ask] R. Askey, *Ramanujan and hypergeometric and basic hypergeometric series*, Ramanujan Internat. Symposium on Analysis, December 26-28, 1987, ed. by N.K. Thakare, 1-83, Pune, India, Russian Math. Surveys, **451** (1990), 37-86.

[BPV] R. Balasubramanian, S. Ponnusamy and M. Vuorinen, *Functional inequalities for the quotients of hypergeometric functions*, J. Math. Anal. Appl., **218** (1998), 256-268.

[Ba] Bateman Manuscript Project (ed. A. Erdélyi, W. Magnus, F. Oberhettinger and F.G. Tricomi), *Higher Transcendental Functions*, Vol. I, McGraw-Hill, New York, 1953.

[Be1] B.C. Berndt, *Ramanujan's Notebooks*, Vol. I, Springer-Verlag, Berlin-Heidelberg-New York, 1985.

[Be2] ———, *Ramanujan's Notebooks*, Vol. II, Springer-Verlag, Berlin-Heidelberg-New York, 1989.

[Be3] ———, *Ramanujan's Notebooks*, Vol. III, Springer-Verlag, Berlin-Heidelberg-New York, 1991.

[BBG] B.C. Berndt, S. Bhargava and F.G. Garvan, *Ramanujan's theories of elliptic functions to alternative bases*, Trans. Amer. Math. Soc., **347** (1995), 4163-4244.

[BB1] J.M. Borwein and P.B. Borwein, *Pi and the AGM*, John Wiley & Sons, New York, 1987.

[BB2] ———, *A cubic counterpart of Jacobi's identity and the AGM*, Trans. Amer. Math. Soc., **323** (1991), 691-701.

[Bo] F. Bowman, *Introduction to Elliptic Functions with Applications*, Dover, New York, 1961.

[BF] P.F. Byrd and M.D. Friedman, *Handbook of Elliptic Integrals for Engineers and Physicists*, 2nd ed., Die Grundlehren Math. Wiss., **67**, Springer-Verlag, Berlin-Göttingen-Heidelberg-New York, 1971.

[C] I. Chavel, *Riemannian Geometry – A Modern Introduction*, Cambridge Tracts in Math., **108**, Cambridge Univ. Press, 1993.

[Di] A.L. Dixon, *Generalizations of Legendre's formula* $KE' - (K - E)K' = \frac{1}{2}\pi$, Proc. London Math. Soc., **3**(**2**) (1905), 206-224.

[Du] P.L. Duren, *The Legendre relation for elliptic integrals*, in Paul Halmos: Celebrating 50 Years of Mathematics, ed. by J.H. Ewing and F.W. Gehring, Springer-Verlag, New York, (1991), 305-315.

[El] E.B. Elliott, *A formula including Legendre's* $EK' + KE' - KK' = \frac{1}{2}\pi$, Messenger of Math., **33** (1904), 31-40.

[Ga] F.G. Garvan, *Ramanujan's theories of elliptic functions to alternative bases – a symbolic excursion*, J. Symbolic Comput., **20** (1995), 517-536.

[LV] O. Lehto and K.I. Virtanen, *Quasiconformal Mappings in the Plane*, 2nd ed., Die Grundlehren der Math. Wiss., **126**, Springer-Verlag, New York-Berlin, 1973.

[Mark] A.I. Markushevich, *Theory of Functions of a Complex Variable*, Vol. II, Prentice-Hall, Englewood Cliffs, NJ, 1965.

[Mart] G.J. Martin, *The distortion theorem for quasiconformal mappings, Schottky's theorem and holomorphic motions*, Proc. Amer. Math. Soc., **125** (1997), 1095-1103.

[Mi] W. Miller, Jr., *Symmetry and separation of variables*, With a foreword by Richard Askey; Encyclopedia of Mathematics and its Applications, Vol. 4, Reading, Massachusetts, Addison-Wesley Publishing Company, **XXX** (1977).

[PV] S. Ponnusamy and M. Vuorinen, *Asymptotic expansions and inequalities for hypergeometric functions*, Mathematika, **44** (1997), 278-301.

[QVu1] S.-L. Qiu and M. Vuorinen, *Landen inequalities for hypergeometric functions*, Nagoya Math. J., **154** (1999), 31-56.

[QVu2] ———, *Duplication inequalities for the ratios of hypergeometric functions*, Forum Math., **11** (1999), to appear.

[QVu3] ———, *Infinite products and the normalized quotients of hypergeometric functions*, SIAM J. Math. Anal., **30** (1999), 1057-1075.

[R] E.D. Rainville, *Special Functions*, MacMillan, New York, 1960.

[Var] V.S. Varadarajan, *Linear meromorphic differential equations: A modern point of view*, Bull. Amer. Math. Soc., **33** (1996), 1-42.

[WW] E.T. Whittaker and G.N. Watson, *A Course of Modern Analysis*, 4th ed., Cambridge Univ. Press, London, l958.

Received April 23, 1998.

Michigan State University
East Lansing, MI 48824
E-mail address: anderson@math.msu.edu

Hangzhou Institute of Electronics Engineering
Hangzhou 310037
P.R. China

University of Auckland
Auckland
New Zealand
E-mail address: vamanamu@math.auckland.nz

University of Helsinki
FIN-00014
Finland
E-mail address: vuorinen@csc.fi

PACIFIC JOURNAL OF MATHEMATICS
Vol. 192, No. 1, 2000

PARTIAL REPRESENTATIONS AND AMENABLE FELL BUNDLES OVER FREE GROUPS

Ruy Exel

We show that a Fell bundle $\mathbb{B} = \{B_t\}_{t\in\mathbb{F}}$, over an arbitrary free group $\mathbb{F}$, is amenable, whenever it is orthogonal (in the sense that $B_x^* B_y = 0$, if x and y are distinct generators of $\mathbb{F}$) and semi-saturated (in the sense that B_{ts} coincides with the closed linear span of $B_t B_s$, when the multiplication "ts" involves no cancelation).

In this work we continue the study of the phenomena of amenability for Fell bundles over discrete groups, initiated in [**E3**]. By definition, a Fell bundle is said to be amenable if the left regular representation of its cross-sectional C^*-algebra is faithful. This property is also equivalent to the faithfulness of the standard conditional expectation. The reader is referred to [**E3**] for more information, but we also offer a very brief survey containing some of the most relevant definitions, in our section on preliminaries below.

The starting point for our work is Theorem 6.7 of [**E3**], where it is shown that a certain grading of the Cuntz–Krieger algebra gives rise to an amenable Fell bundle over a free group. Our main goal is to further pursue the argument leading to this result, in order to obtain a large class of amenable Fell bundles. We find that the crucial properties implying the amenability of a Fell bundle, over a free group $\mathbb{F}$, are *orthogonality* and *semi-saturatedness*. A Fell bundle $\mathbb{B} = \{B_t\}_{t\in\mathbb{F}}$ is said to be *orthogonal* if the fibers B_x and B_y, corresponding to two distinct generators x and y of $\mathbb{F}$, are orthogonal in the sense that $B_x^* B_y = 0$. On the other hand, $\mathbb{B}$ is said to be *semi-saturated* when each fiber B_t is "built up" from the fibers corresponding to the generators appearing in the reduced decomposition of t. More precisely, if $t = x_1 x_2 \cdots x_n$ is in reduced form, then one requires that $B_t = B_{x_1} B_{x_2} \cdots B_{x_n}$ (meaning closed linear span). This property makes sense for any group G, which, like the free group, is equipped with a length function $|\cdot|$. A Fell bundle over such a group is said to be semi-saturated if $B_{ts} = B_t B_s$ (closed linear span), whenever t and s satisfy $|ts| = |t| + |s|$.

Our main result, Theorem 6.3, states, precisely, that any Fell bundle over $\mathbb{F}$, which is orthogonal, semi-saturated, and has separable fibers, must be amenable.

To arrive at this conclusion we first restrict ourselves to a very special case of Fell bundles, namely those which are associated to a partial representation of $\mathbb{F}$ (see below for definitions). For these bundles, we prove an even stronger result, which is that they satisfy the approximation property of [**E3**]. This property implies amenability and also some other interesting facts related to induced ideals of the cross-sectional C^*-algebra (see [**E3**, 4.10]).

The proof of the approximation property for these restricted bundles is a direct generalization of [**E3**, 6.6], where we proved that, for every semi-saturated partial representation σ of $\mathbb{F}_n$ (see below for definitions), such that $\sum_{i=1}^n \sigma(g_i)\sigma(g_i)^* = 1$, the associated Fell bundle satisfies the approximation property. Here, $\{g_1, \ldots, g_n\}$ are the generators of the free group $\mathbb{F}_n$.

Our generalization of this result, namely Theorem 3.7, below, amounts to replacing the hypothesis that $\sum_{i=1}^n \sigma(g_i)\sigma(g_i)^* = 1$, by the weaker requirement that this sum is no larger than 1, or, equivalently, that the $\sigma(g_i)\sigma(g_i)^*$ are pairwise orthogonal projections.

Arriving at this generalization turns out to require a considerable understanding of the various idempotents accompanying a partial representation of $\mathbb{F}$, and underlines the richness of ideas surrounding the concept of partial representations. In addition, the new hypothesis, that is, the orthogonality of these projections, is easily generalizable to free groups with infinitely many generators. With the same ease, based on a simple inductive limit argument, we extend Theorem 3.7 to the infinitely generated case, obtaining Theorem 4.1, below.

In Section 5, armed with this partial result, we study Fell bundles which satisfy, in addition to the hypothesis of our main theorem, a stability property. Employing a fundamental result of Brown, Green and Rieffel [**BGR**], we are able to show that, for such bundles, there is a hidden partial representation of $\mathbb{F}$ which sends us back to the previously studied situation. We finally remove the extra stability hypothesis by means of a simple stabilization argument.

It does not seem outlandish to expect that all amenable Fell bundles satisfy some form of the approximation property. However, having no definite evidence that this is so, we must be cautious in distinguishing these properties. Accordingly, we must stress that our main result falls short of proving the approximation property for the most general situation treated, that is, of orthogonal semi-saturated bundles. In this case, all we obtain is amenability, leaving that stronger property as an open question.

The author would like to express his thanks to the members of the Mathematics Department at the University of Newcastle – Australia, where part of this work was developed. Stimulating conversations with M. Laca and I. Raeburn were essential to the formulation of the question we have studied here.

1. Preliminaries.

For the reader's convenience, and also to fix our notation, we shall begin by briefly discussing some basic facts about partial group representations, Fell bundles and the rich way in which these concepts are interrelated. The reader is referred to [**FD**], [**E2**], and [**E3**] for more information on these subjects.

Let G be a group, fixed throughout this section. Also, let H be a Hilbert space and denote by $\mathcal{B}(H)$ the algebra of all bounded linear operators on H.

Definition 1.1. A *partial representation* of G on H is, by definition, a map $\sigma : G \to \mathcal{B}(H)$ such that

i) $\sigma(t)\sigma(s)\sigma(s^{-1}) = \sigma(ts)\sigma(s^{-1})$,
ii) $\sigma(t^{-1}) = \sigma(t)^*$,
iii) $\sigma(e) = I$,

for all $t, s \in G$, where e denotes the unit group element and I is the identity operator on H.

Let σ be a partial representation of G on H. It is an easy consequence of the definition that each $\sigma(t)$ is a partially isometric operator and hence that

$$\mathrm{e}(t) := \sigma(t)\sigma(t)^*$$

is a projection (that is, a self-adjoint idempotent). It is not hard to show (see [**E2**]) that these projections commute among themselves, and satisfy the commutation relation

$$\sigma(t)\mathrm{e}(s) = \mathrm{e}(ts)\sigma(t), \tag{1.2}$$

for all $t, s \in G$.

There is a special kind of partial representations worth considering, whenever G is equipped with a "length" function, that is, a non-negative real valued function $|\cdot| : G \to \mathbf{R}_+$ satisfying $|e| = 0$ and the triangular inequality $|ts| \leq |t| + |s|$.

Definition 1.3. A partial representation σ of G is said to be *semi-saturated* (with respect to a given length function $|\cdot|$ on G) if $\sigma(t)\sigma(s) = \sigma(ts)$ whenever t and s satisfy $|ts| = |t| + |s|$.

The concept of partial representations is closely related to that of Fell bundles (also known as C^*-algebraic bundles [**FD**]) as we shall now see.

Definition 1.4. Given a partial representation σ of G, for each t in G, let B_t^σ be the closed linear subspace of $\mathcal{B}(H)$ spanned by the set of operators of the form

$$\mathrm{e}(r_1)\mathrm{e}(r_2)\cdots\mathrm{e}(r_k)\sigma(t),$$

where $k \in \mathbf{N}$, and $r_1, r_2, \ldots, r_k$ are arbitrary elements of G.

Using the axioms of partial representations and 1.2, it is an easy exercise to show (see [**E3**, Section 6]) that, for all $t, s \in G$, we have

$$B_t^\sigma B_s^\sigma \subseteq B_{ts}^\sigma \quad \text{and} \quad (B_t^\sigma)^* = B_{t^{-1}}^\sigma.$$

Therefore, the collection

$$\mathbb{B}^\sigma := \{B_t^\sigma\}_{t\in G}$$

is seen to form a Fell bundle over G.

Definition 1.5. A Fell bundle over a discrete group G is a collection $\mathbb{B} = \{B_t\}_{t\in G}$ of closed subspaces of $\mathcal{B}(H)$, such that $B_t B_s \subseteq B_{ts}$ and $(B_t)^* = B_{t^{-1}}$, for all t and s in G.

As is the case with C^*-algebras, which can be defined, concretely, as a norm-closed *-subalgebra of $\mathcal{B}(H)$, as well as a certain abstract mathematical object, defined via a set of axioms, Fell bundles may also be seen under a dual point of view, specially if one restricts attention to the case of discrete groups. The above definition of Fell bundles is the one we adopt here, referring the reader to [**FD**, VIII.16.2] for the abstract version and to [**FD**, VIII.16.4] for the equivalence of these. Nevertheless, it should be said that the point of view one usually adopts in the study of Fell bundles stresses that each B_t should be viewed as a Banach space in its own, and that for each t and s in G, one has certain algebraic operations

$$\cdot : B_t \times B_s \to B_{ts}$$

and

$$* : B_t \to B_{t^{-1}},$$

which, in our case, are induced by the multiplication and involution on $\mathcal{B}(H)$, respectively. If G is not discrete, then one should also take into account a topology on the disjoint union $\dot{\bigcup}_{t\in G} B_t$, which is compatible with the other ingredients present in the situation. See [**FD**] for details. Since we will only deal with Fell bundles over discrete groups, we need not worry about this topology.

Definition 1.6. A Fell bundle $\mathbb{B}$, over of G, is said to be *semi-saturated* (with respect to a given length function $|\cdot|$ on G) if $B_{ts} = B_t B_s$ (closed linear span), whenever t and s satisfy $|ts| = |t| + |s|$.

When σ is semi-saturated, one can prove, as in [**E3**, 6.2], that the Fell bundle $\mathbb{B}^\sigma$ is also semi-saturated.

Definition 1.7. Given any Fell bundle $\mathbb{B} = \{B_t\}_{t\in G}$, with G discrete, one defines its l_1 cross-sectional algebra [**FD**, VIII.5], denoted $l_1(\mathbb{B})$, to be the Banach *-algebra consisting of the l_1 cross-sections of $\mathbb{B}$, under the multiplication

$$fg(t) = \sum_{s\in G} f(s)g(s^{-1}t), \qquad \text{for } t \in G,\ f, g \in l_1(\mathbb{B}),$$

involution

$$f^*(t) = (f(t^{-1}))^*, \quad \text{for } t \in G,\ f \in l_1(\mathbb{B}),$$

and norm

$$\|f\| = \sum_{s \in G} \|f(s)\|, \quad \text{for } f \in l_1(\mathbb{B}).$$

The cross-sectional C^*-algebra of $\mathbb{B}$ [**FD**, VIII.17.2], denoted $C^*(\mathbb{B})$, is defined to be the enveloping C^*-algebra of $l_1(\mathbb{B})$.

There is also a *reduced* cross-sectional C^*-algebra, indicated by $C^*_r(\mathbb{B})$, which is defined to be the closure of $l_1(\mathbb{B})$ in a certain *regular representation* (acting on the right-B_e-Hilbert-bimodule formed by the l_2 cross-sections). See [**E3**, 2.3] for a precise definition.

Both $C^*(\mathbb{B})$ and $C^*_r(\mathbb{B})$ contain a copy of the algebraic direct sum $\bigoplus_{t \in G} B_t$, as a dense subalgebra, making them into G-graded C^*-algebras in the sense of [**FD**, VIII.16.11] (see also [**E3**, 3.1]). In both cases, the projections onto the factors extend to bounded linear maps on the whole algebra, and, in particular, for B_e, that projection gives a conditional expectation [**E3**, 3.3].

This conditional expectation, say E, is faithful in the case of $C^*_r(\mathbb{B})$, in the sense that

$$E(x^*x) = 0 \ \Rightarrow \ x = 0$$

for every $x \in C^*_r(\mathbb{B})$ [**E3**, 2.12]. However, the same cannot be said with respect to $C^*(\mathbb{B})$. In fact, there always exists an epimorphism

$$\Lambda : C^*(\mathbb{B}) \to C^*_r(\mathbb{B})$$

which restricts to the identity map on $\bigoplus_{t \in G} B_t$ (see the discussion following [**E3**, 2.2] as well as [**E3**, 3.3]). The kernel of Λ coincides with the *degeneracy* ideal for E, namely

$$\mathcal{D} = \{x \in C^*(\mathbb{B}) : E(x^*x) = 0\},$$

where we are also denoting the conditional expectation for $C^*(\mathbb{B})$ by E, by abuse of language (see [**E3**, 3.6]). This can be used to give an alternate definition of $C^*_r(\mathbb{B})$, as the quotient of $C^*(\mathbb{B})$ by that ideal.

The crucial property of Fell bundles with which we will be concerned throughout this work is that of amenability. This property is inspired, first of all, in the corresponding concept for groups [**G**], but also in the work of Anantharaman-Delaroche [**A**] and Nica [**Ni**]. In the context we are interested, that is, for Fell bundles, it first appeared in [**E3**], for discrete groups, and was subsequently generalized by Ng for the non-discrete case [**Ng**]. See also [**EN**].

Definition 1.8. A Fell bundle $\mathbb{B}$, over a discrete group G, is said to be amenable, if Λ is an isomorphism.

According to the characterization of the kernel of Λ, as in our discussion above, we see that:

Proposition 1.9. *A necessary and sufficient condition for $\mathbb{B}$ to be amenable is that the conditional expectation E of $C^*(\mathbb{B})$ be faithful.*

Any bundle is amenable when the base group G is amenable [**E3**, 4.7], whereas the typical example of non-amenable bundle is the *group bundle* [**FD**, VIII.2.7] over a non-amenable group G. That is, the bundle $\mathbf{C} \times G$, with the operations (abstractly) defined by

$$(z,t)(w,s) = (zw, ts) \quad \text{and} \quad (z,t)^* = (\bar{z}, t^{-1})$$

for $t, s \in G$ and $z, w \in \mathbf{C}$. In this case $C^*(\mathbb{B})$ is the full group C^*-algebra of G, while $C^*_r(\mathbb{B})$ is its reduced algebra. It is well known [**P**] that Λ is an isomorphism, in this case, if and only if G is amenable.

A Fell bundle may be amenable even if its base group G is not. One example of this situation is given by [**E3**, 6.7]. It consists of a Fell bundle over the non-amenable free group which is, itself, amenable. This example is particularly interesting, since its cross-sectional C^*-algebra is isomorphic to the Cuntz–Krieger algebra $\mathcal{O}_A$.

Definition 1.10. We say that $\mathbb{B}$ has the *approximation property* [**E3**, 4.5] if there exists a net $\{a_i\}_{i\in I}$ of finitely supported functions $a_i : G \to B_e$, which is uniformly bounded in the sense that there exists a constant $M > 0$ such that

$$\left\| \sum_{t\in G} a_i(t)^* a_i(t) \right\| \leq M,$$

for all i, and such that for all b_t in each B_t one has that

$$b_t = \lim_{i\to\infty} \sum_{r\in G} a_i(tr)^* b_t a_i(r).$$

The relevance of the approximation property is that:

Theorem 1.11. *If a Fell bundle $\mathbb{B}$ has the approximation property, then it is amenable.*

Proof. See [**E3**, 4.6]. □

For later use, it will be convenient to have certain equivalent forms of the approximation property, which we now study.

Lemma 1.12. *Let $\mathbb{B}$ be a Fell bundle over G, and let $a : G \to B_e$ be a finitely supported function. Then, for each t in G, the map*

$$b_t \in B_t \mapsto \sum_{r\in G} a(tr)^* b_t a(r) \in B_t$$

is bounded, with norm no bigger than $\left\|\sum_{r\in G} a(r)^* a(r)\right\|$.

Proof. Recall that $\|\sum_{i=1}^n x_i^* y_i\| \leq \|\sum_{i=1}^n x_i^* x_i\|^{\frac{1}{2}} \|\sum_{i=1}^n y_i^* y_i\|^{\frac{1}{2}}$, whenever $x_1, \ldots, x_n$ and $y_1, \ldots, y_n$ are elements of a C^*-algebra. Therefore, letting $M = \left\|\sum_{r\in G} a(r)^* a(r)\right\|$, we have, for all b_t in B_t, that

$$\left\|\sum_{r\in G} a(tr)^* b_t a(r)\right\| \leq \left\|\sum_{r\in G} a(tr)^* a(tr)\right\|^{\frac{1}{2}} \left\|\sum_{r\in G} a(r)^* b_t^* b_t a(r)\right\|^{\frac{1}{2}}$$
$$\leq M^{\frac{1}{2}} \|b_t\| \left\|\sum_{r\in G} a(r)^* a(r)\right\|^{\frac{1}{2}} = M\|b_t\|.$$

□

Proposition 1.13. *Let $\mathbb{B} = \{B_t\}_{t\in G}$ be a Fell bundle over the discrete group G. Also, suppose we are given a dense subset D_t of B_t, for each t in G. Then the following are equivalent:*

i) *$\mathbb{B}$ satisfies the approximation property.*
ii) *There exists a net $\{a_i\}_{i\in I}$ satisfying all of the properties of* 1.10, *except that the condition involving the limit is only assumed for b_t belonging to D_t.*
iii) *There exists a constant $M > 0$ such that, for all finite sets $\{b_{t_1}, b_{t_2}, \ldots, b_{t_n}\}$, with $b_{t_k} \in D_{t_k}$, and any $\varepsilon > 0$, there exists a finitely supported function $a : G \to B_e$ such that*

$$\left\|\sum_{t\in G} a(t)^* a(t)\right\| \leq M$$

and

$$\left\|b_{t_k} - \sum_{r\in G} a(t_k r)^* b_{t_k} a(r)\right\| < \varepsilon,$$

for $k = 1, \ldots, n$.

Proof. The implications (i) ⇒ (ii) and (ii) ⇒ (iii) are obvious. We shall than prove that (iii) ⇒ (ii) and (ii) ⇒ (i). With respect to our first task, consider the set of pairs (X, ε), where X is any finite subset of the disjoint union $\dot{\bigcup}_{t\in G} D_t$, and ε is a positive real. If these pairs are ordered by saying that $(X_1, \varepsilon_1) \leq (X_2, \varepsilon_2)$ if and only if $X_1 \subseteq X_2$ and $\varepsilon_1 \geq \varepsilon_2$, we clearly get a directed set. For each such (X, ε), let $a_{X,\varepsilon}$ be chosen such as to satisfy the conditions of (iii) with respect to the set X and ε. It is then clear that the net $\{a_{X,\varepsilon}\}_{(X,\varepsilon)}$ provides the required net.

As for (ii) ⇒ (i), since the maps of 1.12 are bounded, the convergence referred to in (ii) is easily seen to hold throughout B_t. □

2. Free groups and orthogonal partial representations.

This section is devoted to introducing a certain class of partial representations of the free group. Let $\mathbb{F}$ denote the free group on a possibly infinite set $\mathcal{S}$ (whose elements we call the generators). Each t in $\mathbb{F}$ has a unique decomposition (called its reduced decomposition, or reduced form)

$$t = x_1 x_2 \cdots x_k,$$

where $x_i \in \mathcal{S} \cup \mathcal{S}^{-1}$ and $x_{i+1} \neq x_i^{-1}$ for all i. In this case, we set $|t| = k$ and it is not hard to see that this gives, in fact, a length function for $\mathbb{F}$. It is with respect to this length function that we will speak of semi-saturated partial representations of $\mathbb{F}$.

Definition 2.1. A partial representation σ of $\mathbb{F}$ is said to be *orthogonal* if $\sigma(x)^*\sigma(y) = 0$ whenever $x, y \in \mathcal{S}$ are generators with $x \neq y$.

A partial representation of a group may not be determined by its values on a set of generators. For example, if we set

$$\sigma(t) = \begin{cases} 1 & \text{if } |t| \text{ is even} \\ 0 & \text{if } |t| \text{ is odd} \end{cases}$$

then σ is a partial representation of $\mathbb{F}$ (on a one dimensional Hilbert space), which coincides, on the generators, with the partial representation

$$\sigma'(t) = \begin{cases} 1 & \text{if } t = e \\ 0 & \text{otherwise.} \end{cases}$$

However, if σ is semi-saturated, then

$$\sigma(t) = \sigma(x_1)\sigma(x_2)\cdots\sigma(x_k),$$

whenever $t = x_1 x_2 \cdots x_k$ is in reduced form. Therefore, the values of σ on the generators end up characterizing σ completely. For this reason the fact that σ is orthogonal often says little, unless one supposes that σ is semi-saturated as well.

We shall denote by $\mathcal{W}$ the sub-semigroup of $\mathbb{F}$ generated by $\mathcal{S}$, that is, the set of all products of elements from $\mathcal{S}$ (as opposed to $\mathcal{S} \cup \mathcal{S}^{-1}$). By convention, $\mathcal{W}$ also includes the identity group element. The elements of $\mathcal{W}$ are called the *positive* elements and will usually be denoted by letters taken from the beginning of the Greek alphabet. For each natural number k we will denote by $\mathcal{W}_k$ the set of positive elements of length k.

Note that, if σ is a semi-saturated partial representation of $\mathbb{F}$, and $\alpha, \beta \in \mathcal{W}$, then $\sigma(\alpha)\sigma(\beta) = \sigma(\alpha\beta)$, since $|\alpha\beta| = |\alpha| + |\beta|$. This property will be useful in many situations, below.

Proposition 2.2. *Let σ be an orthogonal, semi-saturated partial representation of $\mathbb{F}$. Then $\sigma(t) = 0$ for all elements t in $\mathbb{F}$ which are not of the form $\mu\nu^{-1}$, with μ and ν positive.*

Proof. Let $t = x_1 x_2 \cdots x_k$, with $x_i \in \mathcal{S} \cup \mathcal{S}^{-1}$, be in reduced form. Then, since σ is semi-saturated, $\sigma(t) = \sigma(x_1)\sigma(x_2)\cdots\sigma(x_k)$. Now, because σ is orthogonal, if $x_i \in \mathcal{S}^{-1}$ and $x_{i+1} \in \mathcal{S}$ then $\sigma(x_i)\sigma(x_{i+1}) = 0$. So, in order to have $\sigma(t)$ nonzero, all elements from $\mathcal{S}$ must be to the left of the elements from $\mathcal{S}^{-1}$ in the decomposition of t. That is, t is of the form described in the statement. $\square$

Proposition 2.3. *Let σ be an orthogonal, semi-saturated partial representation of $\mathbb{F}$, and let $\alpha, \beta \in \mathcal{W}$. If $|\alpha| = |\beta|$, but $\alpha \neq \beta$, then $\sigma(\alpha)^*\sigma(\beta) = 0$.*

Proof. Let $m = |\alpha| = |\beta|$. If $m = 1$ then α and β are in $\mathcal{S}$ and the conclusion is a consequence of the orthogonality assumption. If $m > 1$ write $\alpha = x\tilde{\alpha}$ and $\beta = y\tilde{\beta}$ with $\tilde{\alpha}, \tilde{\beta} \in \mathcal{W}$ and $x, y \in \mathcal{S}$.

Assume, by way of contradiction, that $\sigma(\alpha)^*\sigma(\beta) \neq 0$. Then

$$0 \neq \sigma(x\tilde{\alpha})^*\sigma(y\tilde{\beta}) = \sigma(\tilde{\alpha})^*\sigma(x)^*\sigma(y)\sigma(\tilde{\beta}).$$

So, in particular, $\sigma(x)^*\sigma(y) \neq 0$, which implies that $x = y$.

We therefore have, using 1.2,

$$0 \neq \sigma(\tilde{\alpha})^*\sigma(x)^*\sigma(x)\sigma(\tilde{\beta}) = \sigma(\tilde{\alpha})^*\mathrm{e}(x^{-1})\sigma(\tilde{\beta}) = \sigma(\tilde{\alpha})^*\sigma(\tilde{\beta})\mathrm{e}(\tilde{\beta}^{-1}x^{-1}),$$

which implies that $\sigma(\tilde{\alpha})^*\sigma(\tilde{\beta}) \neq 0$, and hence, by induction, that $\tilde{\alpha} = \tilde{\beta}$. So $\alpha = \beta$. $\square$

The concept of orthogonality also applies to Fell bundles over free groups.

Definition 2.4. A Fell bundle $\mathbb{B} = \{B_t\}_{t\in\mathbb{F}}$ over $\mathbb{F}$ is said to be *orthogonal* if $B_x^* B_y = \{0\}$ whenever $x, y \in \mathcal{S}$ are generators with $x \neq y$.

The parallel between this concept and its homonym 2.1 is illustrated by our next:

Proposition 2.5. *If σ is an orthogonal partial representation of $\mathbb{F}$, then $\mathbb{B}^\sigma$ is an orthogonal Fell bundle.*

Proof. Left to the reader. $\square$

3. The finitely generated case.

We now start the main technical section of the present work. Here we shall prove the approximation property for Fell bundles arising from certain partial representations of free groups. Even though our long range objective is to treat arbitrary free groups, we shall temporarily restrict our attention to finitely generated free groups. So we make the following:

Standing hypothesis 3.1. For the duration of this section, the set $\mathcal{S}$, of generators of $\mathbb{F}$ will be assumed to be finite and σ will be a fixed orthogonal, semi-saturated partial representation of $\mathbb{F}$.

Recall that $\mathrm{e}(t)$ denotes the final projection $\sigma(t)\sigma(t)^*$ of the partial isometry $\sigma(t)$. In addition to $\mathrm{e}(t)$, the following operators will play a crucial role:

$$P_k = \sum_{\alpha\in\mathcal{W}_k} \mathrm{e}(\alpha), \quad k \geq 1$$

$$Q_0 = 1 - P_1,$$

$$f(t) = \sigma(t)Q_0\sigma(t)^*, \quad t \in \mathbb{F}$$

$$Q_k = \sum_{\alpha\in\mathcal{W}_k} f(\alpha), \quad k \geq 1.$$

The only place where the finiteness hypothesis in 3.1 will be explicitly used is in the observation that these sums are finite sums.

Proposition 3.2. *The following relations hold among the operators defined above:*

i) *P_1 and Q_0 are projections.*
ii) *Each $f(t)$ is a projection.*
iii) *If $t, s \in \mathbb{F}$ then $\sigma(t)f(s) = f(ts)\sigma(t)$.*
iv) *For every t one has $f(t) \leq \mathrm{e}(t)$.*
v) *If $\alpha, \beta \in \mathcal{W}$ are such that $|\alpha| = |\beta|$ but $\alpha \neq \beta$, then $\mathrm{e}(\alpha) \perp \mathrm{e}(\beta)$, $f(\alpha) \perp f(\beta)$, and $\mathrm{e}(\alpha) \perp f(\beta)$.*
vi) *For all $k \geq 1$, both P_k and Q_k are projections and $Q_k = P_k - P_{k+1}$.*
vii) *For every n, we have that $Q_0 + Q_1 + \cdots + Q_{n-1} + P_n = 1$.*
viii) *If α and β are distinct positive elements of $\mathbb{F}$ then, regardless of their length, we have that $f(\alpha) \perp f(\beta)$.*

Proof. For $x, y \in \mathcal{W}_1 = \mathcal{S}$, with $x \neq y$, we have, by the orthogonality assumption, that $e(x)e(y) = \sigma(x)\sigma(x)^*\sigma(y)\sigma(y)^* = 0$. Hence P_1 is a sum of pairwise orthogonal projections, and thus, itself a projection. Therefore Q_0 is also a projection.

Speaking of (ii) we have

$$f(t)^2 = \sigma(t)Q_0\sigma(t)^*\sigma(t)Q_0\sigma(t)^* = \sigma(t)Q_0\mathrm{e}(t^{-1})Q_0\sigma(t)^*.$$

Taking into account that the final and initial projections associated to the partial isometries in a partial representation all commute with each other [**E2**], we see that the above equals

$$\sigma(t)Q_0\sigma(t)^*\sigma(t)\sigma(t)^* = \sigma(t)Q_0\sigma(t)^* = f(t).$$

To prove (iii), let $t, s \in \mathbb{F}$. Then

$$\begin{aligned}\sigma(t)f(s) &= \sigma(t)\sigma(s)Q_0\sigma(s)^* = \sigma(t)\sigma(s)\sigma(s)^*\sigma(s)Q_0\sigma(s)^* \\ &= \sigma(ts)\mathrm{e}(s^{-1})Q_0\sigma(s)^* = \sigma(ts)\sigma(ts)^*\sigma(ts)Q_0\mathrm{e}(s^{-1})\sigma(s)^* \\ &= \sigma(ts)Q_0\sigma(ts)^*\sigma(ts)\sigma(s)^* = \sigma(ts)Q_0\sigma(ts)^*\sigma(t) = f(ts)\sigma(t).\end{aligned}$$

As for (iv)

$$\mathrm{e}(t)f(t) = \sigma(t)\sigma(t)^*\sigma(t)Q_0\sigma(t)^* = f(t).$$

Given α and β as in (v) we have, by 2.3, that

$$\mathrm{e}(\alpha)\mathrm{e}(\beta) = \sigma(\alpha)\sigma(\alpha)^*\sigma(\beta)\sigma(\beta)^* = 0$$

which, when combined with (iv) above, yields the other statements of (v).

That P_k is a projection follows from the fact that the summands in its definition are pairwise orthogonal projections. The same reasoning applies to Q_k. Now

$$\begin{aligned}Q_k &= \sum_{\alpha\in\mathcal{W}_k} \sigma(\alpha)(1-P_1)\sigma(\alpha)^* \\ &= \sum_{\alpha\in\mathcal{W}_k} \sigma(\alpha)\sigma(\alpha)^* - \sum_{\alpha\in\mathcal{W}_k}\sum_{x\in\mathcal{S}}\sigma(\alpha)\sigma(x)\sigma(x)^*\sigma(\alpha)^* \\ &= P_k - \sum_{\alpha\in\mathcal{W}_k}\sum_{x\in\mathcal{S}}\sigma(\alpha x)\sigma(\alpha x)^* \\ &= P_k - \sum_{\beta\in\mathcal{W}_{k+1}} \sigma(\beta)\sigma(\beta)^* = P_k - P_{k+1}.\end{aligned}$$

To prove (vii), we just note that

$$\begin{aligned}&Q_0 + Q_1 + \cdots + Q_{n-1} + P_n \\ &= 1 - P_1 + P_1 - P_2 + \cdots + P_{n-1} - P_n + P_n = 1.\end{aligned}$$

Finally, let $\alpha \neq \beta$ be positive and let $k = |\alpha|$ and $l = |\beta|$. If $k = l$ then we have already seen in (v), that $f(\alpha) \perp f(\beta)$. On the other hand, if $k \neq l$, then

$$f(\alpha) \leq Q_k \perp Q_l \geq f(\beta),$$

where the orthogonality of Q_k and Q_l follows from (vii). This implies, again, that $f(\alpha) \perp f(\beta)$. □

Recall that $\mathbb{B}^\sigma = \{B_t^\sigma\}_{t\in G}$ denotes the Fell bundle associated to σ, and consider, for each integer $n \geq 1$, the map $b_n : \mathcal{W} \to B_e^\sigma$ given by

$$b_n(\alpha) = \begin{cases} f(\alpha) & \text{if } |\alpha| < n \\ \mathrm{e}(\alpha) & \text{if } |\alpha| = n \\ 0 & \text{if } |\alpha| > n. \end{cases}$$

Lemma 3.3. *For every $n \geq 1$ we have* $\displaystyle\sum_{\alpha \in \mathcal{W}} b_n(\alpha) = 1.$

Proof. We have

$$\sum_{\alpha \in \mathcal{W}} b_n(\alpha) = \sum_{k=0}^{n} \sum_{\alpha \in \mathcal{W}_k} b_n(\alpha) = \sum_{k=0}^{n-1} \sum_{\alpha \in \mathcal{W}_k} f(\alpha) + \sum_{\alpha \in \mathcal{W}_n} \mathrm{e}(\alpha) = \sum_{k=0}^{n-1} Q_k + P_n = 1.$$

□

The last relevant definition is that of another sequence of maps $a_n : \mathcal{W} \to B_e^\sigma$, this time given by

$$a_n(\alpha) = \left(\frac{1}{n} \sum_{k=1}^{n} b_k(\alpha) \right)^{\frac{1}{2}}.$$

Note that $a_n(\alpha) = 0$ for $|\alpha| > n$. We shall also think of the a_n as functions defined on the whole of $\mathbb{F}$, by setting $a_n(t) = 0$ when t is not positive.

Lemma 3.4. *For every $n \geq 1$ we have* $\displaystyle\sum_{t \in \mathbb{F}} a_n(t)^* a_n(t) = 1.$

Proof. As already observed, $a_n(t)$ vanishes unless t is positive. In addition, $a_n(\alpha)$ is self-adjoint, so we must compute

$$\sum_{\alpha \in \mathcal{W}} a_n(\alpha)^2 = \sum_{\alpha \in \mathcal{W}} \frac{1}{n} \sum_{k=1}^{n} b_k(\alpha) = \frac{1}{n} \sum_{k=1}^{n} \sum_{\alpha \in \mathcal{W}} b_k(\alpha) = 1,$$

where we have used 3.3 in order to conclude the last step above. □

The square root appearing in the definition of a_n can be explicitly computed if we note that, for $|\alpha| \leq n$, we have

$$\sum_{k=1}^{n} b_k(\alpha) = \left(\sum_{k=|\alpha|+1}^{n} f(\alpha) \right) + \mathrm{e}(\alpha) = (n - |\alpha|) f(\alpha) + \mathrm{e}(\alpha)$$
$$= (n - |\alpha| + 1) f(\alpha) + \Big(\mathrm{e}(\alpha) - f(\alpha) \Big),$$

and that the expression above consists of a linear combination of orthogonal projections, namely $f(\alpha)$ and $\mathrm{e}(\alpha) - f(\alpha)$ (see 3.2.iv). It follows that $a_n(\alpha)$ is given, explicitly, by

$$a_n(\alpha) = \left(\frac{n - |\alpha| + 1}{n} \right)^{\frac{1}{2}} f(\alpha) + \left(\frac{1}{n} \right)^{\frac{1}{2}} \Big(\mathrm{e}(\alpha) - f(\alpha) \Big). \tag{3.5}$$

The following is the main technical point in showing the approximation property for $\mathbb{B}^\sigma$:

Lemma 3.6. *For every t in $\mathbb{F}$ we have* $\displaystyle\sigma(t) = \lim_{n \to \infty} \sum_{r \in \mathbb{F}} a_n(tr)^* \sigma(t) a_n(r).$

Proof. By 2.2 we may assume that $t = \mu\nu^{-1}$, where μ and ν are in $\mathcal{W}$. We may also suppose that $|t| = |\mu| + |\nu|$, that is, no cancellation takes place when μ and ν^{-1} are multiplied together.

In addition, since $a_n(r) = 0$ unless r is positive and $|r| \leq n$, each sum above is actually a finite sum. In fact, the nonzero summands in it are among those for which both r and tr are positive of length no bigger than n.

Since $tr = \mu\nu^{-1}r$, if both r and tr are to be positive, we must have $r = \nu\beta$, for some $\beta \in \mathcal{W}$, and then $|tr| = |\mu\nu^{-1}\nu\beta| = |\mu| + |\beta|$.

Also, in order to have $|r|$ and $|tr|$ no larger than n, we will need $|r| = |\nu| + |\beta| \leq n$, as well as $|\mu| + |\beta| \leq n$, which are equivalent to $|\beta| \leq m$, where

$$m = \min\{n - |\nu|, n - |\mu|\}.$$

Summarizing, for every n, we have

$$\sum_{r \in \mathbb{F}} a_n(tr)^* \sigma(t) a_n(r) = \sum_{|\beta| \leq m} a_n(\mu\beta)\sigma(t)a_n(\nu\beta),$$

where we have also taken into account that each $a_n(t)$ is self-adjoint.

Substituting the expression for a_n, obtained in 3.5, in the above sum, we conclude that each individual summand equals

$$\begin{aligned}
&\left[\left(\frac{n - |\mu\beta| + 1}{n}\right)^{\frac{1}{2}} f(\mu\beta) + \left(\frac{1}{n}\right)^{\frac{1}{2}} \Big(\mathrm{e}(\mu\beta) - f(\mu\beta)\Big)\right] \\
&\qquad \cdot \sigma(t) \left[\left(\frac{n - |\nu\beta| + 1}{n}\right)^{\frac{1}{2}} f(\nu\beta) + \left(\frac{1}{n}\right)^{\frac{1}{2}} \Big(\mathrm{e}(\nu\beta) - f(\nu\beta)\Big)\right] \\
&= \left(\frac{n - |\mu\beta| + 1}{n}\right)^{\frac{1}{2}} \left(\frac{n - |\nu\beta| + 1}{n}\right)^{\frac{1}{2}} f(\mu\beta)\ \sigma(t)\ f(\nu\beta) \\
&\quad + \left(\frac{n - |\mu\beta| + 1}{n}\right)^{\frac{1}{2}} \left(\frac{1}{n}\right)^{\frac{1}{2}} f(\mu\beta)\ \sigma(t)\ \Big(\mathrm{e}(\nu\beta) - f(\nu\beta)\Big) \\
&\quad + \left(\frac{1}{n}\right)^{\frac{1}{2}} \left(\frac{n - |\nu\beta| + 1}{n}\right)^{\frac{1}{2}} \Big(\mathrm{e}(\mu\beta) - f(\mu\beta)\Big)\ \sigma(t)\ f(\nu\beta) \\
&\quad + \left(\frac{1}{n}\right)^{\frac{1}{2}} \left(\frac{1}{n}\right)^{\frac{1}{2}} \Big(\mathrm{e}(\mu\beta) - f(\mu\beta)\Big)\ \sigma(t)\ \Big(\mathrm{e}(\nu\beta) - f(\nu\beta)\Big).
\end{aligned}$$

Let us indicate the four summands after the last equal sign above by (i), (ii), (iii), and (iv), in that order. In regards to (i), note that, employing 3.2.iii, we have

$$\begin{aligned}
f(\mu\beta)\sigma(t)f(\nu\beta) &= f(\mu\beta)\sigma(\mu)\sigma(\nu^{-1})f(\nu\beta) \\
&= \sigma(\mu)f(\beta)f(\beta)\sigma(\nu^{-1}) = \sigma(\mu)f(\beta)\sigma(\nu)^*.
\end{aligned}$$

Referring to (ii) we have

$$f(\mu\beta)\sigma(t)\Big(\mathrm{e}(\nu\beta) - f(\nu\beta)\Big) = f(\mu\beta)\Big(\mathrm{e}(\mu\beta) - f(\mu\beta)\Big)\sigma(t) = 0,$$

because of 3.2.iv. Similarly one proves that (iii) vanishes as well. As for (iv)

$$\begin{aligned}
&\Big(\mathrm{e}(\mu\beta) - f(\mu\beta)\Big)\sigma(t)\Big(\mathrm{e}(\nu\beta) - f(\nu\beta)\Big)\\
&= \Big(\mathrm{e}(\mu\beta) - f(\mu\beta)\Big)\sigma(\mu)\sigma(\nu^{-1})\Big(\mathrm{e}(\nu\beta) - f(\nu\beta)\Big)\\
&= \sigma(\mu)\Big(\mathrm{e}(\beta) - f(\beta)\Big)\sigma(\nu)^*.
\end{aligned}$$

So,

$$\begin{aligned}
&a_n(\mu\beta)\sigma(t)a_n(\nu\beta)\\
&= \sigma(\mu)\Big[\left(\frac{n-|\mu\beta|+1}{n}\right)^{\frac{1}{2}}\left(\frac{n-|\nu\beta|+1}{n}\right)^{\frac{1}{2}} f(\beta)\\
&\qquad + \frac{1}{n}\Big(\mathrm{e}(\beta) - f(\beta)\Big)\Big]\sigma(\nu)^*,
\end{aligned}$$

and the conclusion will follow once we prove that the term between brackets above, summed over $|\beta| \le m$, converges, in norm, to the identity operator, as $n \to \infty$. We now set to do precisely this. Speaking of the identity operator, recall from 3.4 and 3.5 that,

$$1 = \sum_{t\in\mathbb{F}} a_m(t)^2 = \sum_{|\beta|\le m} \frac{m-|\beta|+1}{m} f(\beta) + \frac{1}{m}\Big(\mathrm{e}(\beta) - f(\beta)\Big).$$

Using this expression for the identity operator, we must then prove the vanishing of the following limit

$$\begin{aligned}
&\lim_{n\to\infty}\Bigg\| \sum_{|\beta|\le m} \left(\frac{n-|\mu\beta|+1}{n}\right)^{\frac{1}{2}}\left(\frac{n-|\nu\beta|+1}{n}\right)^{\frac{1}{2}} f(\beta)\\
&\quad + \frac{1}{n}\Big(\mathrm{e}(\beta) - f(\beta)\Big) - \frac{m-|\beta|+1}{m} f(\beta) - \frac{1}{m}\Big(\mathrm{e}(\beta) - f(\beta)\Big)\Bigg\|\\
&= \lim_{n\to\infty}\Bigg\| \sum_{|\beta|\le m} \left(\frac{1}{n}(n-|\mu\beta|+1)^{\frac{1}{2}}(n-|\nu\beta|+1)^{\frac{1}{2}} - \frac{m-|\beta|+1}{m}\right) f(\beta)\\
&\quad + \left(\frac{1}{n} - \frac{1}{m}\right)\Big(\mathrm{e}(\beta) - f(\beta)\Big)\Bigg\|\\
&\le \lim_{n\to\infty}\Bigg\| \sum_{|\beta|\le m} \left(\frac{1}{n}(n-|\mu\beta|+1)^{\frac{1}{2}}(n-|\nu\beta|+1)^{\frac{1}{2}} - \frac{m-|\beta|+1}{m}\right) f(\beta)\Bigg\|
\end{aligned}$$

$$+ \lim_{n\to\infty} \left\| \sum_{|\beta|\le m} \left(\frac{1}{n} - \frac{1}{m}\right) \Big(\mathrm{e}(\beta) - f(\beta)\Big) \right\|.$$

The two limits will now be shown to equal zero. We should point out that, with respect to the the first one, we are facing a linear combination of pairwise orthogonal projections by 3.2.viii. The same, however, is not true for the second.

Using this observation, we see that the norm, in the first case, equals

$$\max_{|\beta|\le m} \left| \frac{1}{n}(n - |\mu\beta| + 1)^{\frac{1}{2}}(n - |\nu\beta| + 1)^{\frac{1}{2}} - \frac{m - |\beta| + 1}{m} \right|.$$

In order to show that this goes to zero as $n \to \infty$, let us assume, without loss of generality, that $|\mu| \ge |\nu|$, and hence that $m = n - |\mu|$.

In addition, it is easy to see that, for every pair of positive reals x and y, one has that $|x - y| \le |x^2 - y^2|^{\frac{1}{2}}$. So, the task facing us can be replaced by

$$\lim_{n\to\infty} \max_{|\beta|\le m} \left| \frac{(n - |\mu\beta| + 1)(n - |\nu\beta| + 1)}{n^2} - \frac{(n - |\mu\beta| + 1)^2}{(n - |\mu|)^2} \right| \stackrel{?}{=} 0.$$

The term between the single bars is no bigger than

$$\begin{aligned} &\left| \frac{(n - |\mu\beta| + 1)(n - |\nu\beta| + 1)}{n^2} - \frac{(n - |\mu\beta| + 1)^2}{n^2} \right| \\ &\quad + (n - |\mu\beta| + 1)^2 \left| \frac{1}{n^2} - \frac{1}{(n - |\mu|)^2} \right| \\ &\le \frac{n - |\mu\beta| + 1}{n^2} \Big| |\mu| - |\nu| \Big| + (n - |\mu\beta| + 1)^2 \left| \frac{-2n|\mu| + |\mu|^2}{n^2(n - |\mu|)^2} \right| \\ &\le \frac{n + 1}{n^2} \Big| |\mu| - |\nu| \Big| + (n + 1)^2 \left| \frac{-2n|\mu| + |\mu|^2}{n^2(n - |\mu|)^2} \right|, \end{aligned}$$

which is now easily seen to go to zero, uniformly on β, as $n \to \infty$.

To conclude, we need only show the vanishing of

$$\begin{aligned} &\lim_{n\to\infty} \left\| \sum_{|\beta|\le m} \left(\frac{1}{n} - \frac{1}{m}\right) \Big(\mathrm{e}(\beta) - f(\beta)\Big) \right\| \\ &= \lim_{n\to\infty} \left| \frac{1}{n} - \frac{1}{m} \right| \left\| \sum_{k=0}^{m} \sum_{\beta\in\mathcal{W}_k} \mathrm{e}(\beta) - f(\beta) \right\| \\ &= \lim_{n\to\infty} \left| \frac{1}{n} - \frac{1}{m} \right| \left\| \sum_{k=0}^{m} P_k - Q_k \right\| \\ &= \lim_{n\to\infty} \left| \frac{1}{n} - \frac{1}{m} \right| \left\| \sum_{k=0}^{m} P_{k+1} \right\| \end{aligned}$$

$$\leq \lim_{n\to\infty} \left|\frac{1}{n} - \frac{1}{m}\right| (m+1).$$

Now, recalling our assumption that $m = n - |\mu|$, the limit above equals

$$\lim_{n\to\infty} \left|\frac{1}{n} - \frac{1}{n-|\mu|}\right| (n-|\mu|+1) = \lim_{n\to\infty} \frac{|\mu|(n-|\mu|+1)}{n(n-|\mu|)} = 0.$$

□

We are now prepared to face one of our main goals.

Theorem 3.7. *Let σ be an orthogonal, semi-saturated partial representation of a finitely generated free group $\mathbb{F}$. Then the Fell bundle $\mathbb{B}^\sigma$ satisfies the approximation property and hence is amenable. Moreover, the constant M referred to in* 1.10 *may be taken to be* 1.

Proof. Let $\{a_n\}_{n\in\mathbf{N}}$ be defined as above. Then $\left\|\sum_{t\in\mathbb{F}} a_n(t)^* a_n(t)\right\| = 1$, for all n, by 3.4, and employing 1.13.ii, it is now enough to show that

$$b_t = \lim_{n\to\infty} \sum_{r\in\mathbb{F}} a_n(tr)^* b_t a_n(r),$$

for all b_t of the form $b_t = \mathrm{e}(r_1)\mathrm{e}(r_2)\cdots\mathrm{e}(r_k)\sigma(t)$, where $k \in \mathbf{N}$, and r_1, $r_2, \ldots, r_k$ are arbitrary elements of $\mathbb{F}$. This is because the linear combinations of the elements of this form are dense in B_t^σ, by Definition 1.4.

We have already observed that the projections associated to the partial isometries in a partial representation form a commutative set. Since $a_n(\alpha)$ is given by a linear combination of such projections, by 3.5 (and $a_n(t) = 0$ when $t \notin \mathcal{W}$), it is clear that $a_n(t)$ commutes with the $\mathrm{e}(r_j)$. Therefore, by 3.6,

$$\begin{aligned}
&\lim_{n\to\infty} \sum_{r\in\mathbb{F}} a_n(tr)^* b_t a_n(r) \\
&= \mathrm{e}(r_1)\mathrm{e}(r_2)\cdots\mathrm{e}(r_k) \lim_{n\to\infty} \sum_{r\in\mathbb{F}} a_n(tr)^* \sigma(t) a_n(r) \\
&= \mathrm{e}(r_1)\mathrm{e}(r_2)\cdots\mathrm{e}(r_k)\sigma(t) = b_t.
\end{aligned}$$

This concludes the proof of the approximation property and hence of the amenability of $\mathbb{B}^\sigma$, by 1.11. □

4. Arbitrary free groups.

We will now extend the results of the previous section, by dropping the finiteness hypothesis of 3.1, and hence including infinitely generated free groups in our study. The strategy will be to adapt the work done above to the general case using an inductive limit argument.

Let, therefore, $\mathbb{F}$ be the free group on a set $\mathcal{S}$, no longer assumed to be finite, or even countable. Also let σ be an orthogonal, semi-saturated partial representation of $\mathbb{F}$, considered fixed throughout this section.

For each finite subset X of $\mathcal{S}$, let $\mathbb{F}_X$ denote the subgroup of $\mathbb{F}$ generated by X. It is quite obvious that $\mathbb{F}_X$ is again a free group, and that $\mathbb{F}$ is the union of the increasing net $\{\mathbb{F}_X\}_X$. The length functions we've been considering are compatible in the sense that the one for $\mathbb{F}$ restricts to the one for $\mathbb{F}_X$. Therefore the restriction of σ to $\mathbb{F}_X$ is also semi-saturated, and obviously also orthogonal. Let $\mathbb{B}^X$ denote the Fell bundle for $\sigma|_{\mathbb{F}_X}$, as in 1.4.

It is clear that, for each t in $\mathbb{F}$, one has that B_t^σ is the closure of the union of the B_t^X, as X ranges in the collection of finite subsets $X \subseteq \mathcal{S}$, such that $t \in \mathbb{F}_X$.

Theorem 4.1. *Let σ be an orthogonal, semi-saturated partial representation of an arbitrary free group $\mathbb{F}$. Then the Fell bundle $\mathbb{B}^\sigma$ satisfies the approximation property and hence is amenable.*

Proof. For each t in $\mathbb{F}$, let D_t be the union of the B_t^X, as described above, which is dense in B_t^σ. We will now prove 1.13.iii, with respect to this choice of D_t. Let $M = 1$. Then, given a finite set $\{b_{t_1}, b_{t_2}, \ldots, b_{t_n}\}$, with $b_{t_k} \in D_{t_k}$, and any $\varepsilon > 0$, there clearly exists a single finite $X \subseteq \mathcal{S}$, such that every $b_{t_k} \in B_{t_k}^X$. Now, by 3.7 we conclude that a finitely supported map $a : \mathbb{F}_X \to B_e^X \subseteq B_e^\sigma$ exists, satisfying $\left\| \sum_{t \in \mathbb{F}_X} a(t)^* a(t) \right\| \leq 1$, and $\left\| b_{t_k} - \sum_{r \in \mathbb{F}_X} a(t_k r)^* b_{t_k} a(r) \right\| < \varepsilon$, for all $k = 1, \ldots, n$. If we extend a to the whole of $\mathbb{F}$ by declaring it zero outside $\mathbb{F}_X$, then these two sums may be taken for $t \in \mathbb{F}$, as opposed to $t \in \mathbb{F}_X$, without changing the result. We conclude that 1.13.iii holds, and hence that $\mathbb{B}^\sigma$ satisfies the approximation property . ∎

5. Stable Fell bundles.

We shall now treat Fell bundles over arbitrary free groups, which are not necessarily presented in terms of a partial representation. This section does not yet contain our strongest result, because we shall be working under the assumption that the unit fiber algebra of $\mathbb{B}$, that is B_e, is a stable C^*-algebra, in the sense that it is isomorphic to $B_e \otimes \mathcal{K}$, where $\mathcal{K}$ is the algebra of compact operators on a separable, infinite dimensional Hilbert space. In the next and final section we will then remove this stability hypothesis.

The method we shall adopt here will be to construct a partial representation which is closely related to $\mathbb{B}$. In fact this method is, essentially, the one we have used in [**E1**] to obtain the classification of Fell bundles in terms of twisted partial actions (see [**E1**, 7.3]). However, we will refrain from utilizing the full machinery of [**E1**] for two reasons. First, by proceeding more or less from scratch, and using the special features of the free group, we will be able to make the presentation somewhat more elementary and self contained. Secondly, the classification theorem mentioned includes a 2-cocycle, which will not show up here, again because of the special properties of the group we are dealing with.

We begin with some simple facts about partial isometries and projections on a Hilbert space H.

Lemma 5.1. *Let p be an operator on H, such that $p^2 = p$ and $\|p\| \le 1$. Then $p = p^*$.*

Proof. Let $\xi \in p(H)^\perp$. Then, for every $\lambda \in \mathbf{R}$ we have $\|p(\xi + \lambda p(\xi))\| \le \|\xi + \lambda p(\xi)\|$, which implies, after a short calculation, that $(1+2\lambda)\|p(\xi)\|^2 \le \|\xi\|^2$, and hence that $p(\xi) = 0$. This says that p vanishes on $p(H)^\perp$ and, since p is the identity on $p(H)$, then it must be the orthogonal projection onto $p(H)$. Hence $p = p^*$. □

Lemma 5.2. *Let p and q be projections (self-adjoint idempotents) in $\mathcal{B}(H)$. Then pq is idempotent if and only if p and q commute.*

Proof. If pq is idempotent, then, since $\|pq\| \le 1$, we have, by 5.1, that $pq = (pq)^* = qp$. The converse is trivial. □

Lemma 5.3. *Let u and v be partial isometries in $\mathcal{B}(H)$. Then uv is a partial isometry if and only if u^*u and vv^* commute (see also* [**S**]*).*

Proof. We have that uv is a partial isometry, if and only if

$$uv(uv)^*uv = uv \iff uvv^*u^*uv = uv$$
$$\iff u^*uvv^*u^*uvv^* = u^*uvv^* \iff (u^*uvv^*)^2 = u^*uvv^*,$$

which, by 5.2, is equivalent to the commutativity of u^*u and vv^*. □

Proposition 5.4. *Let $U = \{u_x\}_{x \in \mathcal{S}}$ be a family of partial isometries on a Hilbert space H and let $\mathcal{I}$ be the multiplicative sub-semigroup of $\mathcal{B}(H)$ generated by $U \cup U^*$. Denote by $\mathbb{F}$ the free group on $\mathcal{S}$. Then, the following are equivalent:*

i) *There exists a semi-saturated partial representation σ of $\mathbb{F}$ such that $\sigma(x) = u_x$ for every $x \in \mathcal{S}$.*
ii) *There exists a partial representation σ of $\mathbb{F}$ such that $\sigma(x) = u_x$ for every $x \in \mathcal{S}$.*
iii) *Every u in $\mathcal{I}$ is a partial isometry.*
iv) *For any $u, v \in \mathcal{I}$ we have that uu^* and vv^* commute.*

Proof. (i) ⇒ (ii): Obvious.

(ii) ⇒ (iii): Recall that an operator u is a partial isometry, if and only if $uu^*u = u$. So, let $u = u_1 \cdots u_n$, with $u_i \in U \cup U^*$, and take $t_i \in \mathcal{S} \cup \mathcal{S}^{-1}$

such that $\sigma(t_i) = u_i$. Then $u = \sigma(t_i)\cdots\sigma(t_n)$ and, by induction on n,

$$\begin{aligned} uu^*u &= \sigma(t_1)\cdots\sigma(t_{n-1})\mathrm{e}(t_n)\sigma(t_{n-1})^*\cdots\sigma(t_1)^*\sigma(t_1)\cdots\sigma(t_n) \\ &= \mathrm{e}(t_1\cdots t_n)\sigma(t_1)\cdots\sigma(t_{n-1})\sigma(t_{n-1})^*\cdots\sigma(t_1)^*\sigma(t_1)\cdots\sigma(t_n) \\ &= \mathrm{e}(t_1\cdots t_n)\sigma(t_1)\cdots\sigma(t_{n-1})\sigma(t_n) \\ &= \sigma(t_1)\cdots\sigma(t_{n-1})\mathrm{e}(t_n)\sigma(t_n) = \sigma(t_1)\cdots\sigma(t_{n-1})\sigma(t_n) = u. \end{aligned}$$

(iii) $\Rightarrow$ (iv): If $u, v \in \mathcal{I}$, then $u^*v \in \mathcal{I}$ and, by (iii), it is a partial isometry. Hence, using 5.3, we have that uu^* and vv^* commute.

(iv) $\Rightarrow$ (i): Define, for all $x \in \mathcal{S}$, $\sigma(x) = u_x$ and $\sigma(x^{-1}) = u_x^*$. Now, if $t = x_1\cdots x_n$, with $x_i \in \mathcal{S}\cup\mathcal{S}^{-1}$, is in reduced form, put $\sigma(t) = \sigma(x_1)\cdots\sigma(x_n)$. It is then obvious that $\sigma(t)\sigma(s) = \sigma(ts)$ whenever t and s satisfy $|ts| = |t|+|s|$.

We claim that σ is a partial representation of $\mathbb{F}$. The crucial point is to prove that $\sigma(t)\sigma(s)\sigma(s)^* = \sigma(ts)\sigma(s)^*$ for $t, s \in \mathbb{F}$. To do this we use induction on $|t|+|s|$. If either $|t|$ or $|s|$ is zero, there is nothing to prove. So, write $t = \tilde{t}x$ and $s = y\tilde{s}$, where $x, y \in \mathcal{S}\cup\mathcal{S}^{-1}$ and, moreover, $|t| = |\tilde{t}|+1$ and $|s| = |\tilde{s}|+1$.

In case $x^{-1} \neq y$ we have $|ts| = |t|+|s|$ and hence $\sigma(ts) = \sigma(t)\sigma(s)$. If, on the other hand, $x^{-1} = y$, we have

$$\sigma(t)\sigma(s)\sigma(s)^* = \sigma(\tilde{t}x)\sigma(x^{-1}\tilde{s})\sigma(\tilde{s}^{-1}x) = \sigma(\tilde{t})\sigma(x)\sigma(x)^*\sigma(\tilde{s})\sigma(\tilde{s})^*\sigma(x).$$

By (iv) and the induction hypothesis we conclude that the above equals

$$\begin{aligned} \sigma(\tilde{t})\sigma(\tilde{s})\sigma(\tilde{s})^*\sigma(x)\sigma(x)^*\sigma(x) &= \sigma(\tilde{t}\tilde{s})\sigma(\tilde{s})^*\sigma(x) \\ &= \sigma(ts)\sigma(\tilde{s}^{-1}x) = \sigma(ts)\sigma(s)^*. \end{aligned}$$

□

It would be interesting to find a condition about a set U of partial isometries, which is equivalent to the ones above, but which refers exclusively to the u_x's, themselves, rather than to arbitrary products of them. A related observation is that the Cuntz-Krieger [**CK**] relations were shown to imply the conditions above [**E3**, 5.2].

In our next result, we will use two important concepts from [**E1**], which we briefly summarize here. By definition, a TRO (for ternary ring of operators), is a closed linear subspace $E \subseteq \mathcal{B}(H)$, such that $EE^*E \subseteq E$ (see also [**Z**]). We adopt the convention that the product of two or more sets, as above, is supposed to mean the *closed linear span* of the set of products.

Given a TRO E, we say that a partial isometry u is associated to E [**E1**, 5.4], and write $u \sim E$, if $u^*E = E^*E$ and $uE^* = EE^*$. If, in addition, the range of the final projection uu^* coincides with EH (equivalently, if the range of the initial projection u^*u coincides with E^*H), then we say that u is *strictly* associated to E [**E1**, 5.5], and write $u \overset{s}{\sim} E$.

It is a consequence of [**BGR**, 3.3 and 3.4], that whenever E is separable and stable, in the sense that E^*E and EE^* are stable C^*-algebras [**E1**, 4.11], then a partial isometry strictly associated to E always exists (see also [**E1**, 5.3 and 5.2]).

The reason why TROs are relevant here is that any fiber of a Fell bundle is a TRO, as one may easily see. But, because of the separability requirement, we shall now restrict to the case of separable bundles, according to the following:

Definition 5.5. A Fell bundle $\mathbb{B}$, over a discrete group G, is said to be *separable* if each B_t is a separable Banach space.

Theorem 5.6. *Let $\mathbb{B} = \{B_t\}_{t\in\mathbb{F}}$ be a semi-saturated, separable Fell bundle over the arbitrary free group $\mathbb{F}$, represented on a Hilbert space H. Suppose that B_e is stable. Then, there exists a semi-saturated partial representation σ of $\mathbb{F}$ on H, such that $\sigma(t) \overset{s}{\sim} B_t$. In addition, if $\mathbb{B}$ is orthogonal, then σ is necessarily orthogonal, as well.*

Proof. By [**E1**, 4.12], each B_t is stable and hence, by [**E1**, 5.2], there exists a partial isometry $u_t \overset{s}{\sim} B_t$. Let $U = \{u_x\}_{x\in\mathcal{S}}$, where, as before, $\mathcal{S}$ is the set of generators of $\mathbb{F}$. We claim that U satisfies 5.4.iii. In fact, we shall prove that, given $x_1, \ldots, x_n$ in $\mathcal{S}\cup\mathcal{S}^*$, then $B_{x_1}\cdots B_{x_n}$ is a TRO and that $u_{x_1}\cdots u_{x_n}$ is a partial isometry strictly associated to it. Proceeding by induction, we may assume that $E = B_{x_1}\cdots B_{x_{n-1}}$ is a TRO and that $u = u_{x_1}\cdots u_{x_{n-1}} \overset{s}{\sim} E$. Now, observe that E^*E and $B_{x_n}B_{x_n}^*$ are ideals in B_e, and hence they commute, as sets, that is, $E^*EB_{x_n}B_{x_n}^* = B_{x_n}B_{x_n}^*E^*E$. So, by [**E1**, 6.4], it follows that EB_{x_n} is a TRO, and that $uu_{x_n} \overset{s}{\sim} EB_{x_n}$. This proves our claim. So, let σ be a semi-saturated partial representation of $\mathbb{F}$ satisfying 5.4.i. It remains to show that $\sigma(t) \overset{s}{\sim} B_t$, but this follows from the conclusion just above, once we write $t = x_1 \ldots x_n$ in reduced form.

Assume, now, that $\mathbb{B}$ is orthogonal. Then, given $x \neq y$, in $\mathcal{S}$, we have, again by [**E1**, 6.4], that $\sigma(x)^*\sigma(y) \overset{s}{\sim} B_x^*B_y = \{0\}$. Therefore $\sigma(x)^*\sigma(y) = 0$. ☐

Theorem 5.7. *Let $\mathbb{B}$ be an orthogonal, semi-saturated, separable Fell bundle over $\mathbb{F}$, and suppose that B_e is stable. Then $\mathbb{B}$ is amenable* (*see below for the non-stable case*).

Proof. Let H be the space where $\mathbb{B}$ acts, and let σ be the orthogonal, semi-saturated partial representation of $\mathbb{F}$ on H, provided by 5.6. As usual, we denote by $\mathrm{e}(t)$ the final projection of $\sigma(t)$.

Let $\mathbb{B}^\sigma$ be the Fell bundle associated to σ as in 1.4. By 4.1, we know that $\mathbb{B}^\sigma$ satisfies the approximation property. Let, therefore, $\{a_i\}_{i\in I}$ be a net of maps satisfying the conditions of 1.10, with respect to $\mathbb{B}^\sigma$.

We claim that $\mathrm{e}(t)$ commutes with B_e. In fact, because $\sigma(t) \overset{s}{\sim} B_t$, we have that $\mathrm{e}(t)$ is the orthogonal projection onto B_tH. Now, observing that B_e leaves the latter invariant, we obtain the conclusion. It follows that the C^*-algebra generated by all the $\mathrm{e}(t)$, namely B_e^σ, is contained in the commutant of B_e.

Let $t \in \mathbb{F}$ and pick $b_t \in B_t$. Define $c_t = b_t\sigma(t)^*$. Then $c_t \in B_t\sigma(t)^* = B_tB_t^* \subseteq B_e$. On the other hand, note that, since the range of b_t^* is contained in $B_{t^{-1}}H$, which is also the range of $\mathrm{e}(t^{-1})$, we have that $\mathrm{e}(t^{-1})b_t^* = b_t^*$, or simply $b_t\mathrm{e}(t^{-1}) = b_t$. This implies that $c_t\sigma(t) = b_t\sigma(t)^*\sigma(t) = b_t\mathrm{e}(t^{-1}) = b_t$.

Since each $a_i(r) \in B_e^\sigma$, we have that it commutes with c_t and hence

$$\lim_{i\to\infty}\sum_{r\in\mathbb{F}} a_i(tr)^*b_ta_i(r) = c_t\lim_{i\to\infty}\sum_{r\in\mathbb{F}} a_i(tr)^*\sigma(t)a_i(r) = c_t\sigma(t) = b_t.$$

We therefore see that the net $\{a_i\}$ satisfies the properties of 1.10 with respect to $\mathbb{B}$, except that there is no reason to expect that the values of the maps a_i lie in B_e. Therefore this falls short of proving the approximation property for $\mathbb{B}$ and hence we cannot use 1.11 to conclude the amenability of $\mathbb{B}$.

Fortunately, what we do have is enough to fit the hypothesis of a slight generalization of the results of [**E3**] leading to 1.11, which goes as follows: Let $C^*(\mathbb{B})$ be faithfully represented on a Hilbert space K. Since $C^*(\mathbb{B})$ contains $\bigoplus_{t\in G} B_t$, we may then identify each B_t with its image under that representation, and then think of B_t as a space of operators on K. In other words, this provides a faithful representation of $\mathbb{B}$ as operators on K, and hence we might as well assume that $H = K$, which we do, from now on. Under this assumption we have that the sub-C^*-algebra of $\mathcal{B}(H)$ generated by $\bigcup_t B_t$ is isomorphic to $C^*(\mathbb{B})$.

For each t, consider the space C_t of operators on $H \otimes l_2(\mathbb{F})$, given by $C_t = B_t \otimes \lambda_t$, where λ is the left regular representation of $\mathbb{F}$. It is easy to see that the C_t form a Fell bundle, which is again isomorphic to $\mathbb{B}$. However, the C^*-algebra generated by $\bigcup_t C_t$ is now isomorphic to $C_r^*(\mathbb{B})$, a fact that follows from [**E3**, 3.7].

Recall that the reasoning at the beginning of the present proof provided a net $\{a_i\}_{i\in I}$ of maps $a_i : G \to \mathcal{B}(H)$, such that $\sup_i \left\|\sum_{t\in G} a_i(t)^*a_i(t)\right\| < \infty$, and that $b_t = \lim_{i\to\infty}\sum_{r\in G} a_i(tr)^*b_ta_i(r)$, for all b_t in each B_t. Following the argument used in [**E3**, 4.3], let, for each i,

$$V_i : H \to H \otimes l_2(\mathbb{F}),$$

be given by the formula $V_i(\xi) = \sum_{t\in\mathbb{F}} a_i(t)\xi \otimes \delta_t$, where $\{\delta_t\}$ is the standard orthonormal base of $l_2(\mathbb{F})$. One then proves, as in [**E3**, 4.3], that $\|V_i\| \leq \|\sum_{t\in\mathbb{F}} a_i(t)^*a_i(t)\|^{1/2}$, and hence that the V_i are uniformly bounded.

Now, define the completely positive maps

$$\Psi_i : \mathcal{B}(H \otimes l_2(\mathbb{F})) \to \mathcal{B}(H)$$

by $\Psi_i(T) = V_i^* T V_i$, for each T in $\mathcal{B}(H \otimes l_2(\mathbb{F}))$. Again as in [**E3**, 4.3], one has that, for every b_t in B_t,

$$\Psi_i(b_t \otimes \lambda_t) = \sum_{r \in \mathbb{F}} a_i(tr)^* b_t a_i(r).$$

The somewhat annoying fact that $a_i(t)$ may not belong to B_e forbids us to say that Ψ_i is a map from $C_r^*(\mathbb{B})$ into $C^*(\mathbb{B})$, as stated in [**E3**, 4.3]. This, however, is not a cause for despair.

Consider the canonical map $\Lambda : C^*(\mathbb{B}) \to C_r^*(\mathbb{B})$, described in section 1. Under the current representation of $C_r^*(\mathbb{B})$ on $H \otimes l_2(\mathbb{F})$, we have that $\Lambda(b_t) = b_t \otimes \lambda_t$ for all b_t. Now, consider the composition of maps

$$C^*(\mathbb{B}) \xrightarrow{\Lambda} C_r^*(\mathbb{B}) \xrightarrow{\Psi_i} \mathcal{B}(H).$$

For b_t in B_t, we have

$$\lim_i \Psi_i(\Lambda(b_t)) = \lim_i \Psi_i(b_t \otimes \lambda_t) = \lim_i \sum_{r \in \mathbb{F}} a_i(tr)^* b_t a_i(r) = b_t.$$

Now, by the uniform boundedness of these maps we then conclude that $\lim_i \Psi_i(\Lambda(x)) = x$ for all $x \in C^*(\mathbb{B})$. This implies that Λ is injective and hence that $\mathbb{B}$ is amenable, as required. $\square$

6. The general case.

In this section we will prove our most general result, which is the amenability of orthogonal, semi-saturated, separable bundles. This amounts to dropping the stability hypothesis of the previous section, which we do by a "stabilization argument". Ideally one should develop the whole theory of tensor products for Fell bundles but we feel this is not the place to do it. Instead, we perform our tensor products in a way which is enough for our purposes, albeit in a somewhat crude manner.

Let $\mathbb{B}$ be any Fell bundle over a discrete group G, acting on the Hilbert space H. As in the proof of 5.7, we may assume that the sub-C^*-algebra of $\mathcal{B}(H)$ generated by $\bigcup_t B_t$ is isomorphic to $C^*(\mathbb{B})$.

For each $t \in G$, consider the the subset of $\mathcal{B}(H \otimes l_2)$ (where l_2 is the usual infinite dimensional separable Hilbert space), denoted by $B_t \otimes \mathcal{K}$, and defined by

$$B_t \otimes \mathcal{K} = \overline{\operatorname{span}}\{b_t \otimes k : b_t \in B_t, k \in \mathcal{K}\}.$$

Here $\mathcal{K}$ is the algebra of compact operators on l_2. It is elementary to show that $\mathbb{B} \otimes \mathcal{K}$, as defined by $\mathbb{B} \otimes \mathcal{K} = \{B_t \otimes \mathcal{K}\}_{t \in G}$, is a Fell bundle in its own right. We shall say that $\mathbb{B} \otimes \mathcal{K}$ is the *stabilization* of $\mathbb{B}$.

Proposition 6.1. *Let $\mathbb{B}$ be a Fell bundle. Then $C^*(\mathbb{B} \otimes \mathcal{K})$ is isomorphic to $C^*(\mathbb{B}) \otimes \mathcal{K}$.*

Proof. Let us temporarily use the notation $\tilde{\mathbb{B}}$ for $\mathbb{B} \otimes \mathcal{K}$ and $\tilde{B}_t$ for $B_t \otimes \mathcal{K}$. Observe that $\mathcal{K}$ may be viewed, in a canonical way, as a subalgebra of the multiplier algebra $\mathcal{M}(\tilde{B}_e)$, which, in turn, may be viewed as a subalgebra of $\mathcal{M}(C^*(\tilde{\mathbb{B}}))$ [**FD**, VIII.5.8 and 1.15]. Since one clearly has $\mathcal{K}C^*(\tilde{\mathbb{B}}) = C^*(\tilde{\mathbb{B}})$, one may now show that $C^*(\tilde{\mathbb{B}})$ is isomorphic to $A \otimes \mathcal{K}$, where $A = pC^*(\tilde{\mathbb{B}})p$, and p is any minimal projection in $\mathcal{K}$.

Using the universal property [**FD**, VIII.16.12] of cross-sectional C^*-algebras, one may show that the assignment $b_t \in B_t \mapsto b_t \otimes p \in C^*(\tilde{\mathbb{B}})$ extends to a surjective *-homomorphism $\phi : C^*(\mathbb{B}) \to A$.

We claim that ϕ is injective. In fact, let H be the space where $\mathbb{B}$ acts, so that $\tilde{\mathbb{B}}$ sits inside of $\mathcal{B}(H \otimes l_2)$. The universal property, this time applied to $\tilde{\mathbb{B}}$, implies that there exists a *-representation

$$\pi : C^*(\tilde{\mathbb{B}}) \to \mathcal{B}(H \otimes l_2)$$

which, restricted to each $\tilde{B}_t$, coincides with the inclusion of $\tilde{B}_t$ in $\mathcal{B}(H \otimes l_2)$. It is then easy to see that $\pi\phi$ maps $C^*(\mathbb{B})$ onto the closed linear span of $\bigcup_t B_t \otimes p$, within $\mathcal{B}(H \otimes l_2)$, which is isomorphic to $C^*(\mathbb{B}) \otimes p$ (see the second paragraph of this section).

Since $\pi\phi$ sends each b_t to $b_t \otimes p$, we see that $\pi\phi$ is the canonical isomorphism between $C^*(\mathbb{B})$ and $C^*(\mathbb{B}) \otimes p$. This shows that ϕ is injective and hence an isomorphism onto A, concluding the proof. $\square$

Proposition 6.2. *If $\mathbb{B}$ is a Fell bundle such that $\mathbb{B} \otimes \mathcal{K}$ is amenable, then $\mathbb{B}$, itself, is amenable.*

Proof. Consider the diagram

$$\begin{array}{ccc} C^*(\mathbb{B}) & \stackrel{\phi}{\to} & C^*(\mathbb{B} \otimes \mathcal{K}) \\ E \downarrow & & \downarrow \tilde{E} \\ B_e & \to & B_e \otimes \mathcal{K} \end{array}$$

where ϕ is the injective map described in the proof of 6.1, the vertical arrows represent the corresponding conditional expectations, and the unlabeled horizontal arrow maps each b_e to $b_e \otimes p$. It is easy to see that this is a commutative diagram. Recall that a Fell bundle is amenable if and only if the conditional expectation on its cross-sectional algebra is faithful, as seen in 1.9. So, let us show that $\mathbb{B}$ possesses this property. If $x \in C^*(\mathbb{B})$ is such that $E(x^*x) = 0$, then we have that $\tilde{E}(\phi(x^*x)) = 0$, whence $\phi(x^*x) = 0$ and, finally, $x = 0$ because ϕ is injective. $\square$

The following is our main result:

Theorem 6.3. *Let $\mathbb{B}$ be an orthogonal, semi-saturated, separable Fell bundle over $\mathbb{F}$. Then $\mathbb{B}$ is amenable.*

Proof. All of the properties assumed on $\mathbb{B}$ are clearly inherited by $\mathbb{B} \otimes \mathcal{K}$. In addition, the latter possesses a stable unit fiber algebra and hence, by 5.7, it is amenable. The conclusion then follows from 6.2. □

Some of the most useful facts about amenable Fell bundles (see e.g. [**E3**, 4.10]) require not only that the bundle be amenable, but that the approximation property holds. This leads one to ask whether the result above could be strengthened by replacing amenability with the approximation property. We do not have a satisfactory answer to this question but then, again, we do not know of any example of an amenable Fell bundle which does not satisfy the approximation property.

References

[A] C. Anantharaman-Delaroche, *Systemes dynamiques non commutatifs et moyenabilité*, Math. Ann., **279** (1987), 297-315.

[BGR] L.G. Brown, P. Green and M.A. Rieffel, *Stable isomorphism and strong Morita equivalence of C^*-algebras*, Pacific J. Math., **71** (1977), 349-363.

[CK] J. Cuntz and W. Krieger, *A class of C*-algebras and topological Markov chains*, Inventiones Math., **56** (1980), 251-268.

[E1] R. Exel, *Twisted partial actions, a classification of regular C^*-algebraic bundles*, Proc. London Math. Soc., **74** (1997), 417-443.

[E2] ———, *Partial actions of groups and actions of inverse semigroups*, Proc. Amer. Math. Soc., **126** (1998), 3481-3494.

[E3] ———, *Amenability for Fell bundles*, J. Reine Angew. Math., **492** (1997), 41-73.

[EN] R. Exel and C.K. Ng, *Approximation property of C^*-algebraic bundles*, preprint, Oxford University, 1999; (math.OA/9906070).

[FD] J.M.G. Fell and R.S. Doran, *Representations of *-algebras, locally compact groups, and Banach *-algebraic bundles*, Pure and Applied Mathematics, Academic Press, Vol. 125 and 126, 1988.

[G] F. P. Greenleaf, *Invariant means on topological groups*, van Nostrand-Reinhold, Mathematical Studies, **16** (1969).

[Ng] C.-K. Ng, *Reduced Cross-sectional C^*-algebras of C^*-algebraic bundles and Coactions*, Oxford University, preprint, 1996.

[Ni] A. Nica, *C^*-algebras generated by isometries and Wiener-Hopf operators*, J. Operator Theory, **27** (1991), 1-37.

[P] G. K. Pedersen, *C^*-Algebras and their automorphism groups*, Acad. Press, 1979.

[S] N. Sieben, *C^*-crossed products by partial actions and actions of inverse semigroups*, Arizona State University, Masters Thesis, 1994.

[Z] H. Zettl, *A characterization of ternary rings of operators*, Adv. Math., **48** (1983), 117-143.

Received April 20, 1998. The author was partially supported by CNPq – Brazil.

Universidade Federal de Santa Catarina
88010-970 Florianópolis SC
Brazil
E-mail address: exel@mtm.ufsc.br

PACIFIC JOURNAL OF MATHEMATICS
Vol. 192, No. 1, 2000

DISCS AND THE MORERA PROPERTY

Josip Globevnik and Edgar Lee Stout

In previous work, we have considered the problem of showing that a continuous function on a real hypersurface Γ in $\mathbb{C}^N$ satisfies the tangential Cauchy-Riemann equations provided that its slices satisfy conditions of Morera type. For instance, these results imply that if $\Omega \subset \mathbb{C}^N$ is a bounded convex domain with smooth boundary, strictly convex at $z_0 \in bD$, if L_0 is a complex line tangent to $b\Omega$ at z_0 and if f is a continuous function on $b\Omega$ such that $\int_{L\cap b\Omega} f\omega = 0$ for all complex lines L close to L_0 which meet Ω and for all $(1,0)$ forms with constant coefficients, then f is a CR function in a neighbourhood of z_0. This fails to hold if L_0 is a complex line that meets Ω even under much stronger assumption of holomorphic extendibility along complex lines. Indeed, let $\mathbb{B}$ be the open unit ball in $\mathbb{C}^2$, and define a function f on $b\mathbb{B}\setminus\{z=0\}$ by $f(z,w) = 1/\overline{z}$. It is easy to verify that for each complex line L close to the z-axis, $f|L\cap b\mathbb{B}$ has a continuous extension to $L\cap\overline{\mathbb{B}}$ which is holomorphic on $L\cap\mathbb{B}$, yet there is no open set in $b\mathbb{B}$ on which f is a CR function. So to conclude that f is a CR function one has to assume the holomorphic extension property for a larger family of analytic discs.

1. Introduction and the main result.

Let Δ be the open unit disc in $\mathbb{C}$. An *analytic disc attached to a manifold* $M \subset \mathbb{C}^N$ is a continuous map $\varphi\colon \overline{\Delta} \to \mathbb{C}^N$, holomorphic on Δ and such that $\varphi(b\Delta) \subset M$. A.E. Tumanov proved that if f is a function of class $\mathcal{C}^1$ on a generic submanifold M of $\mathbb{C}^N$, and φ_0 is an analytic disc attached to M such that for all analytic discs φ, attached to M and close to φ_0, the function $\zeta \to f(\varphi(\zeta))$ has a continuous extension from $b\Delta$ to $\overline{\Delta}$ which is holomorphic on Δ, then f is a CR function in a neighbourhood of $\varphi_0(b\Delta)$ [**Tu1**].

In the present paper we consider a similar problem with conditions of Morera type for continuous functions on real hypersurfaces. Let $\Omega \subset \mathbb{C}^N$ be a bounded domain with $\mathcal{C}^2$ boundary. A subset $D \subset \Omega$ is called a *transversely embedded analytic disc* if $D = V \cap \Omega$ where V is a one dimensional complex submanifold of an open neighbourhood of $\overline{\Omega}$ which intersects $b\Omega$ transversely such that $V \cap \Omega$ is biholomorphically equivalent to Δ. Then $V \cap b\Omega$ is a simple closed curve which bounds D in V and which we denote by bD. Clearly $D = \varphi(\Delta)$ where $\varphi\colon \overline{\Delta} \to \mathbb{C}^N$ is a one to one continuous map which is holomorphic and regular on Δ. We say that φ *parametrizes* D. Let D_0 be a transversely embedded analytic disc and let φ_0 be a parametrization of D_0. A family $\mathcal{D}$ of such discs is called a *neighbourhood of* D_0 if there is an $\varepsilon > 0$ such that $\mathcal{D}$ contains each transversely embedded analytic disc D which can be parametrized by a map φ_D satisfying $|\varphi_D(\zeta) - \varphi_0(\zeta)| < \varepsilon$ $(\zeta \in \Delta)$.

As we shall see, the formulation of our principal result depends in an essential way on the linear structure of $\mathbb{C}^N$ so that it has no obvious analogue in the setting of domains in complex manifolds. However, in the final part of the paper we deduce from the principal result certain consequences that do hold on domains in Stein manifolds.

In formulating our main results it is convenient to use the notation that $\mathbb{C}^{1;0}[dz]$ denotes the space of all $(1,0)-$forms on $\mathbb{C}^N$ with constant coefficients and $\mathbb{C}^{1;1}[dz]$ the space of all $(1,0)-$forms with coefficients that are polynomials of degree not more than one. These spaces are invariant under the action of the group of affine automorphisms of $\mathbb{C}^N$ but under the action of no larger subgroup of the group $\mathbf{Aut}(\mathbb{C}^N)$.

The following is the principal result of the paper:

Theorem 1.1. *Let $\Omega \subset \mathbb{C}^N$ be a bounded domain with boundary of class $\mathcal{C}^2$. Let $D_0 \subset \Omega$ be a transversely embedded analytic disc and let f be a continuous function in a neighbourhood of bD_0 in $b\Omega$. Let w_0 be a point of D_0, z_0 a point of bD_0. Assume that*

$$\int_{bD} f\omega = 0 \tag{1.1}$$

for every $\omega \in \mathbb{C}^{1;1}[dz]$ and for every transversely embedded analytic disc D belonging to a neighbourhood $\mathcal{D}$ of D_0 such that $w_0 \in D$. If Ω is strictly pseudoconvex at z_0 then the function f is a CR function in a neighbourhood of z_0. If Ω is strictly convex at z_0, it suffices to assume the vanishing of the integrals (1.1) *only when $\omega \in \mathbb{C}^{1;0}[dz]$.*

In the following N is a fixed positive integer at least two. By $\mathbb{B}$ we denote the open unit ball in $\mathbb{C}^N$ or in $\mathbb{R}^n$, depending on the context. Similarly, $\mathbb{B}(z,r)$ is the ball of radius r centered at the point z.

We describe the idea of the proof in the special case when $\Omega \subset \mathbb{C}^2$ is a convex domain and $D_0 = \Lambda_0 \cap b\Omega$ where Λ_0 is a complex line which meets Ω. Let L_0 be the complex line tangent to $b\Omega$ at z_0. To prove that f is a CR function in a neighbourhood of z_0 it is, by [**GlS**, Th. 3.2.1], enough to prove that

$$\int_{L\cap b\Omega} f\omega = 0 \tag{1.2}$$

for all $\omega \in \mathbb{C}^{1;0}[dz]$ and for all complex lines L which meet Ω and belong to a neighbourhood $\mathcal{L}$ of L_0 in the space of complex lines. Fix such a line L and let $z \in L \cap b\Omega$. By the strict convexity of Ω at z_0, z is close to z_0 provided that $\mathcal{L}$ is sufficiently small. Let Λ be the complex line passing through w_0 and z. If $\mathcal{L}$ is sufficiently small then $\Lambda \cap \Omega$ belongs to $\mathcal{D}$. For convenience assume that $z = 0$. Let ℓ and h be linear functions on $\mathbb{C}^2$ such that $L = \{z \in \mathbb{C}^2 : \ell(z) = 0\}$, $\Lambda = \{z \in \mathbb{C}^2 : h(z) = 0\}$. We show that there are $\alpha \in \mathbb{R}$ and $\tau > 0$ such that if $V_t = \{z \in \mathbb{C}^2 : \ell(z)h(z) = t^2 e^{i\alpha}\}$, $0 < t < \tau$, and if E is a unitary map such that $|E - I| < \tau$ then $E(V_t) \cap \Omega$ is a transversely embedded analytic disc which belongs to $\mathcal{D}$ provided that $\mathcal{L}$ and τ are small enough. For each t, $0 < t < \tau$, we choose E_t in such a way that $w_0 \in E_t(V_t)$ and that $E_t \to I$ as $t \to 0$. We show that if $t \to 0$ then $E_t(V_t) \cap b\Omega$ converges to $[\Lambda \cap b\Omega] \cup [L \cap b\Omega]$ in such a way that

$$\int_{E_t(V_t)\cap b\Omega} f\omega \to \int_{\Lambda\cap b\Omega} f\omega + \int_{L\cap b\Omega} f\omega$$

for each smooth 1-form ω on $\mathbb{C}^2$. Now, for each $\omega \in \mathbb{C}^{1;0}[dz]$ the terms on the left and the first term on the right vanish by the assumption so (1.2) holds.

In Section 2 we analyze carefully the intersections of the varieties $V_t = \{(z, w) \in \mathbb{C}^2 : \ zw = t^2, \ t > 0\}$ with certain smooth perturbations of the real hyperplane $\{(z, w) \in \mathbb{C}^2 : \ Re(z + w) = 0\}$. In Section 3 we use the results of this analysis to show the existence of α and τ with the properties above and then prove the theorem.

2. Intersections in $\mathbb{C}^2$.

To begin with, we study how certain small smooth perturbations of the real hyperplane

$$H = \{(z, w) \in \mathbb{C}^2 : Re(z + w) = 0\}$$

intersect the variety $V_1 = \{(z, w) \in \mathbb{C}^2 : zw = 1\}$.

The first step in this direction is to compute explicitly the intersection $V_1 \cap H$ and to show that this intersection is transverse.

The transversality assertion goes as follows. Denote by Λ the (unique) complex line in H that passes through the origin. Thus, $\Lambda = \{(z, w) \in \mathbb{C}^2 : w = -z\}$. At a point $(z, \frac{1}{z})$ of V_1, the tangent space to V_1 is the complex line $\{\zeta(1, -1/z^2) : \zeta \in \mathbb{C}\}$. This coincides with Λ if and only if $z^2 = 1$, i.e., if and only if $z = \pm 1$. It follows that V_1 meets the real hyperplane H transversely.

The determination of $V_1 \cap H$ uses the identification of $\mathbb{C}^2$ with $\mathbb{R}^4$ obtained by setting $z = x + iy$ and $w = u + iv$. The point (z, w) lies in $V_1 \cap H$ if and only if $x + u = 0, xu - yv = 1$, and $uy + xv = 0$. These equations imply that $u = 0$, $x = 0$, and $yv = -1$. Thus

$$V_1 \cap H = \{(x + iy, u + iv) : x = u = 0, \ yv = -1\},$$

a hyperbola contained in the (real) $(y,v)-$plane in $\mathbb{C}^2$. When studying the intersection of V_1 with perturbations of H, it is useful to view this hyperbola $yv=-1$ as the union of two graphs in a new coordinate system with the lines $y=\pm v$ as the coordinate axes. For this purpose introduce new coordinates in $\mathbb{R}^4$ by means of the real orthogonal transformation

$$X=\frac{-x+u}{\sqrt{2}},\ Y=\frac{-y+v}{\sqrt{2}},\ U=\frac{x+u}{\sqrt{2}},\ V=\frac{y+v}{\sqrt{2}}.$$

The inverse of this is the orthogonal transformation given by

$$x=\frac{-X+U}{\sqrt{2}},\ y=\frac{-Y+V}{\sqrt{2}},\ u=\frac{X+U}{\sqrt{2}},\ v=\frac{Y+V}{\sqrt{2}}.$$

With respect to the $(X,Y,U,V)-$coordinate system, the equation of H is $U=0$, and

$$V_1\cap H=\left\{(0,\sqrt{V^2+2},0,V):V\in\mathbb{R}\right\}\cup\left\{(0,-\sqrt{V^2+2},0,V):V\in\mathbb{R}\right\}.$$

The perturbations of H that we need to consider are graphs $Gr(\varphi)$ of the form

$$Gr(\varphi)=\{(X,Y,U,V):U=\varphi(X,Y,V):X,Y,V\in\mathbb{R}\}$$

where φ is a function of class $\mathcal{C}^1$ on $\mathbb{R}^3$ such that

$$|\varphi(p)|<\eta\ (p\in\mathbb{R}^3,|p|<\rho)\text{ and }|(D\varphi)(p)|<\eta\ (p\in\mathbb{R}^3) \tag{2.1}$$

with suitably chosen ρ and η. Note that in (2.1) the condition $|\varphi(p)|<\eta$ is assumed to hold only for $p\in\mathbb{R}^3$ in a fixed neighborhood of the origin; it need not hold on large subsets of $\mathbb{R}^3$. Nevertheless, it turns out that if η is small enough, then $Gr(\varphi)\cap H$ is similar to $V_1\cap H$ and, in particular, it is a union of two graphs:

Lemma 2.1. *There are $\eta>0$, $\rho<\infty$, and $M<\infty$ such that whenever φ is a function of class $\mathcal{C}^1$ on $\mathbb{R}^3$ that satisfies* (2.1), *then*

(a) *$Gr(\varphi)$ is transverse to V_1, whence $Gr(\varphi)\cap V_1$ is a closed submanifold of $\mathbb{R}^4$ of class $\mathcal{C}^1$,*

(b) *$Gr(\varphi)\cap V_1=\Lambda_1\cup\Lambda_2$ where Λ_1 and Λ_2 are disjoint curves of class $\mathcal{C}^1$ each of which is of the form*

$$\Lambda_j=\{(X_j(V),Y_j(V),U_j(V),V):V\in\mathbb{R}\}$$

where X_j,Y_j, and U_j are $\mathcal{C}^1$ functions on $\mathbb{R}$ the first derivatives of which are bounded by M, and

(c) *$Y_1(V)>0$ and $Y_2(V)<0$ $\ (V\in\mathbb{R})$.*

Remark 1. If $\varphi\equiv0$, then $Gr(\varphi)=H$, $U_j=X_j=0$, and Y_1,Y_2 are the functions $V\mapsto\pm\sqrt{V^2+2}$.

Remark 2. Having proved (a), to prove (b) it is enough to prove that there is a $\beta > 0$ such that for each $s \in \mathbb{R}$, $Gr(\varphi) \cap V_1$ intersects the hyperplane

$$E_s = \{(X, Y, U, s) : \ X, Y, U \in \mathbb{R}\}$$

at precisely two points and, moreover, that if $p \in Gr(\varphi) \cap V_1 \cap E_s$, then the angle between the hyperplane E_s and the (tangent line to) the curve $Gr(\varphi) \cap V_1$ at p is bounded below by β.

The proof of Lemma 2.1 requires some preliminaries.

Denote by L_1 and L_2 the $z-$ and $w-$axes in $\mathbb{C}^2$, respectively. Far from the origin, V_1 is a slight perturbation of $L_1 \cup L_2$, so to understand $Gr(\varphi) \cap V_1$ far from the origin, we first understand how $Gr(\varphi)$ intersects L, a small perturbation of L_1 or L_2.

Lemma 2.2. *There are δ, $\beta > 0$ such that if φ is a $\mathcal{C}^1$ function on $\mathbb{R}^3$ with*

$$|(D\varphi)(p)| < \delta \quad (p \in \mathbb{R}^3) \tag{2.2}$$

and if f, g are functions of class $\mathcal{C}^1$ on $\mathbb{R}^2$ that satisfy

$$||f||_{\mathcal{C}^1(\mathbb{R}^2)} < \delta, \quad \|g\|_{\mathcal{C}^1(\mathbb{R}^2)} < \delta \tag{2.3}$$

then with $L = \{(x, y, f(x, y), g(x, y)) : \ x, y \in \mathbb{R}\}$ we have that

(a) *$Gr(\varphi)$ is transverse to L,*
(b) *for each s, $Gr(\varphi) \cap L \cap E_s$ consists of a single point, and*
(c) *if $p \in Gr(\varphi) \cap E_s$, then the angle between E_s and $Gr(\varphi) \cap L$ at p is bounded below by β.*

Proof. The $z-$axis $L_1 = \{(x, y, 0, 0) : x, y \in \mathbb{R}\}$ intersects H transversely, and the line $L_1 \cap H$ intersects E_s transversely at an angle that does not depend on s. Further, (2.3) implies that at each point p of L, the tangent space T_pL is arbitrarily close to L_1, uniformly with respect to $p \in L$ provided that δ is small enough, and (2.2) implies that each point $p \in Gr(\varphi)$, $T_pGr(\varphi)$ is arbitrarily close to H, uniformly with respect to p, provided that δ is small enough. This shows that $Gr(\varphi)$ is transverse to L if δ is small enough, which proves (a). It also shows that the tangent line to $L \cap Gr(\varphi)$ at a point p is arbitrarily close to $L_1 \cap H$ when δ is small enough, which proves that there is a $\beta > 0$ satisfying (c) when δ is small enough.

If $s \in \mathbb{R}$, then $L \cap E_s = \{(x, y, f(x, y), g(x, y)) : y + g(x, y) = s\sqrt{2}\}$. If δ is small enough, then a simple one-variable argument shows that given $s, x \in \mathbb{R}$, there is a unique $y = y(x, s) = s\sqrt{2} + \psi(x, s)$ such that $y + g(x, y) = s\sqrt{2}$, and, moreover, that $\psi(x, s)$ is arbitrarily small, uniformly with respect to $x, s \in \mathbb{R}$ when δ is sufficiently small. By the implicit mapping theorem ψ is of class $\mathcal{C}^1$. By differentiating the equality $\psi(x, s) + g(x, s\sqrt{2} + \psi(x, s)) = 0$, we see that $\frac{\partial \psi}{\partial x}$ and $\frac{\partial \psi}{\partial s}$ are arbitrarily small, uniformly in $x, s \in \mathbb{R}$ when δ is small. If we put $t = -x/\sqrt{2}$ and pass to $X, Y, U, V-$coordinates, we can write $L \cap E_s = F_s(\mathbb{R})$ where $F_s(t) = (t + A(t,s), -s + B(t,s), -t + A(t,s), s)$ with $\|A\|_{\mathcal{C}^1}, \|B\|_{\mathcal{C}^1}$ arbitrarily small provided that δ is small enough. The set $E_s \cap Gr(\varphi)$ has two (connected) components $\Omega_s^- = \{(X,Y,U,s) : U < \varphi(X,Y,s)\}$ and $\Omega_s^+ = \{(X,Y,U,s) : U > \varphi(X,Y,s)\}$. If $p \in Gr(\varphi) \cap E_s$, let n_p be the unit normal vector to $Gr(\varphi) \cap E_s$ at p that points in the direction of Ω_s^+. By (2.2), n_p is arbitrarily close to $(0,0,1,0)$, uniformly with respect to p if δ is small enough. Further, for sufficiently small δ, $\dot{F}(t)/|\dot{F}(t)|$ is arbitrarily close to $\left(\frac{1}{\sqrt{2}}, 0, -\frac{1}{\sqrt{2}}, 0\right)$, uniformly with respect to s and t. Thus, granted that δ is small enough, if $\langle , \rangle$ denotes the real inner product on $\mathbb{R}^4$, then

$$\left\langle \frac{\dot{F}_s(t)}{|\dot{F}_s(t)|}, n_p \right\rangle < 0 \quad (t \in \mathbb{R}, p \in Gr(\varphi) \cap E_s),$$

which shows that only way that $F_s(t)$ can meet $Gr(\varphi)$ as t increases is when it passes from Ω_s^+ to Ω_s^-, which implies that $Gr(\varphi) \cap E_s \cap L$ contains at most one point.

To see that $Gr(\varphi) \cap E_s \cap L \neq \emptyset$ we must show that $\varphi(t + A(t,s), -s + B(t,s), s) = -t + A(t,s)$ for at least one t. For small δ this follows from the one-variable fact that if p, q are two functions of class $\mathcal{C}^1$ on $\mathbb{R}$ such that $p'(t) < \mu < \nu < q'(t)$, $t \in \mathbb{R}$, for two constants μ and ν, then their graphs intersect.

Lemma 2.3. *If $R > 0$, then the δ of Lemma 2.2 can be chosen so small that if (2.2) and (2.3) are satisfied, and if, in addition, φ satisfies*

$$|\varphi(X,Y,V)| < \delta \quad (|X|, |Y|, |V| < 2R) \tag{2.4}$$

then $|s| \le R$ and $(X,Y,U,s) \in Gr(\varphi) \cap L \cap E_s$ imply that $|X| < R$ and $|Y| < 2R$.

Proof. Choose $\omega \in (0, \min\{1, R\})$ such that $2\omega/(1-\omega) < R - \omega$. Let A, B be as in the proof of Lemma 2.2. Choose $\delta < \omega$ so small that Lemma 2.2 holds, that $|A(t,s)| < \omega, |B(t,s)| < \omega$ $(t, s \in \mathbb{R})$, and that φ satisfies

$$\left| \frac{d}{dt} \varphi(t + A(t,s), -s + B(t,s), s) \right| < \omega \quad (t, s \in \mathbb{R}). \tag{2.5}$$

Assume also that φ satisfies (2.4). Recall that $Gr(\varphi) \cap L \cap E_s = \{(\lambda + A(\lambda,s), -s + B(\lambda,s), -\lambda + A(\lambda,s), s)\}$ where $-\lambda + A(\lambda,s) = \varphi(\lambda + A(\lambda,s), -s + B(\lambda,s), s)$. It follows that

$$\begin{aligned} & -\lambda + A(\lambda,s) \\ & \quad = \int_0^\lambda \frac{d}{dt} \varphi(t + A(t,s), -s + B(t,s), s) dt + \varphi(A(0,s), -s + B(0,s), s). \end{aligned}$$

Since $|A(0,s)|, |B(0,s)| < R$, it follows by (2.4) that $|s| \le R$ implies that

$$|\varphi(A(0,s), -s + B(0,s), s)| < \delta < \omega$$

whence (2.5) implies that $|\lambda| < 2\omega/(1-\omega) < R-\omega$ so $|-\lambda + A(\lambda, s)| < R$. Clearly $|-s+B(t,s)| < R+R = 2R$. This completes the proof.

To prove Lemma 2.1, we also need the following quantitative form of the inverse function theorem:

Lemma 2.4. *Let $\mathcal{U}$ be an open subset of $\mathbb{R}^n$ that contains the origin, and let $F: \mathcal{U} \to R^n$ be a map of class $\mathcal{C}^1$ such that $F(0) = 0$ and $(DF)(0)$ is nonsingular. If $r > 0$ is so small that*

$$|(DF)(x) - (DF)(0)| < \frac{1}{8|(DF)(0)^{-1}|} \quad (|x| < r), \tag{2.6),}$$

then given a $\mathcal{C}^1$ map $\varphi : \mathcal{U} \to \mathbb{R}^n$ such that

$$|(D\varphi)(x)| < \frac{1}{16|(DF)(0)^{-1}|} \quad (|x| < r) \tag{2.7}$$

and

$$|\varphi(0)| < \frac{r}{4|(DF)(0)^{-1}|}, \tag{2.8}$$

there is a unique x with $|x| < r$ such that $F(x) + \varphi(x) = 0$.

Remark 3. The assumptions imply that if $|x| < r$, then $D(F+\varphi)(x)$ is nonsingular. Indeed, if $|x| < r$, then $DF(0)^{-1}D(F+\varphi)(x)$ is invertible:

$$\begin{aligned}
&|DF(0)^{-1}D(F+\varphi)(x) - I| \\
&= |DF(0)^{-1}\{D(F+\varphi)(x) - DF(0)\}| \\
&\le |DF(0)^{-1}|\{|DF(x) - DF(0)| + |D\varphi(x)|\} \\
&\le |DF(0)^{-1}|\left\{\frac{1}{8}\frac{1}{|DF(0)^{-1}|} + \frac{1}{16}\frac{1}{|DF(0)^{-1}|}\right\} = 3/16 < 1.
\end{aligned}$$

The proof of Lemma 2.4 is in the Appendix to this section.

Proof of Lemma 2.1. For each $t > 0$, set $D_t = \{(X, Y, U, V) : |V| < t\}$, and let L_1, L_2 be the coordinate axes in $\mathbb{C}^2$. Sufficiently far away from the origin, V_1 is an arbitrarily small perturbation of $L_1 \cup L_2$, so by Lemmas 2.2 and 2.3, there are R, $2 < R < \infty$, $\delta > 0$, and $\gamma > 0$ such that if φ satisfies (2.2) and (2.4), then:
(i) $Gr(\varphi)$ is transverse to V_1 at each point of $\mathbb{C}^2 \setminus D_R$,
(ii) for each s, $|s| \ge R$, the set $Gr(\varphi) \cap V_1 \cap E_s$ consists of precisely two points and at each of these points the angle between (the tangent line to) $Gr(\varphi) \cap V_1$ and E_s is at least γ, and
(iii) if $|s| \le R$ and if $(X, Y, U, s) \in Gr(\varphi) \cap V_1$, then $|X| < R$ and $|Y| < 2R$.

Let the ρ of condition (2.1) satisfy $\rho > 8R$. To complete the proof, it suffices to show that there are η, $0 < \eta < \delta$, and $\gamma' > 0$ such that if φ satisfies (2.1), then $G(\varphi)$ is transverse to V_1 at each point of $\overline{D}_R$ and that for each s, $|s| < R$, the set $\{|X| < R, |Y| < 2R\} \cap V_1 \cap Gr(\varphi) \cap E_s$ consists of precisely two points at each of which the angle between (the tangent line to) $V_1 \cap Gr(\varphi)$ and E_s is at least γ'.

The equations of V_1 in the (X, Y, U, V)-coordinates are $-X^2 + Y^2 + U^2 - V^2 = 2$ and $XY - UV = 0$, so given $s \in \mathbb{R}$, we find $E_s \cap V_1 \cap Gr(\varphi)$ by solving $\tilde{F}(X, Y, U, V) = (0, 0, 0, s)$ where

$$\tilde{F}(X, Y, U, V) = (-X^2 + Y^2 + U^2 - V^2 - 2, XY - UV, U - \varphi(X, Y, V), V).$$

Let

$$F(X, Y, U, V) = (-X^2 + Y^2 + U^2 - V^2 - 2, XY - UV, U, V).$$

Then $F^{-1}(0, 0, 0, s) = V_1 \cap H \cap E_s = \{(0, \sqrt{2 + s^2}, 0, s), (0, -\sqrt{2 + s^2}, 0, s)\}$. We have $\det[DF(0, \pm\sqrt{2 + s^2}, 0, s)] = -2(2 + s^2)\,(s \in \mathbb{R})$, so $(DF)(p)$ is invertible at each point of $V_1 \cap H$. Clearly $(DF)(p)$ and $(DF)(p)^{-1}$ depend continuously on $p = (0, \pm\sqrt{2 + s^2}, 0, s)$ on each branch of $V_1 \cap H$, i.e., they depend continuously on s. By compactness Lemma 2.4 now implies that

(**A**) there are r, $0 < r < \sqrt{R^2 + 2} - R$ and $\beta > 0$ such that whenever $p = (0, \pm\sqrt{s^2 + 2}, 0, s)$ with $|s| \leq R$ and φ satisfies $|\varphi| < \beta$ and $|D\varphi| < \beta$ on $(p + r\mathbb{B}) \cap H$, then $Gr(\varphi) \cap V_1 \cap E_s \cap (p + r\mathbb{B})$ consists of precisely one point, at which $D\tilde{F}$ is nonsingular.

The nonsingularity of $D\tilde{F}$ at p means that $Gr(\varphi)$ is transverse to V_1 at p and that the angle between $Gr(\varphi) \cap V_1$ and E_s at p is positive. Since this angle depends continuously on s, it follows that for $|s| \leq R$ it is bounded below by a positive constant.

The equation of $V_1 \cap H$ is $X = U = 0$, $Y = \pm\sqrt{2 + V^2}$. Since $R > 2$ and $r < \sqrt{R^2 + 2} - R$, it follows that $|X'| < R$ and $|Y'| < 2R$ whenever $|V| \leq R, p = (X, Y, 0, V) \in V_1 \cap H$, and $p' = (X', Y', U', V)$ satisfies $|p' - p| < r$. Let $\mathcal{U} = [\overline{D}_R \cap V_1 \cap H] + r\mathbb{B}$. By compactness there is some η, $0 < \eta < \min\{\delta, \beta\}$, such that $\overline{D}_R \cap V_1 \cap \{(X, Y, U, V) : |X| < R, |Y| < 2R, |U| < \eta\} \subset \mathcal{U}$. Since $\rho > 8R$, it follows that if φ satisfies (2.1) then $|\varphi| < \beta$ and $|D\varphi| < \beta$ on $\mathcal{U} \cap H$, which, by (**A**), implies that if $|s| \leq R$, then the only points (X, Y, U, V) in $E_s \cap V_1 \cap Gr(\varphi)$ that satisfy $|X| < R$, $|Y| < 2R$ are the two described in (**A**).

It remains only to prove assertion (c). For this, note that $V_1 \setminus \mathbb{B}(0, r)$ is an arbitrarily small perturbation of $(L_1 \cup L_2) \setminus \mathbb{B}(0, r)$ where the L_i are the coordinate axes, provided that r is large enough. In particular, if ρ in Lemma 2.1 is large enough and $\eta > 0$ small enough then for, say, $|V| > \rho/4$, the tangent line to Λ_i is arbitrarily close to either $L_1 \cap H = \{X = U = V = 0\} = \{X = U = 0, V = -Y\}$ or to $L_2 \cap H = \{X = U = Y = 0\} = \{X = U = 0, V = Y\}$. In particular, there are disjoint neighborhoods $\mathcal{W}_1$ of $L_1 \cap H$ and $\mathcal{W}_2$ of $L_2 \cap H$ in the space of real lines passing through the origin such that the tangent line to Λ_i for $|V| > \rho/4$ is either in $\mathcal{W}_1$ or in $\mathcal{W}_2$. With no loss of generality, assume that if $\{X = p_1V, Y = p_2V, U = p_3V : V \in \mathbb{R}\}$ belongs to $\mathcal{W}_1$, then $p_2 > \frac{1}{2}$ and if it belongs to $\mathcal{W}_2$, then $p_2 < -\frac{1}{2}$. On the other hand, provided that η is small enough, $\{V_1 \cap Gr(\varphi) : |V| < \rho/2\}$ is an arbitrarily small perturbation of $\{V_1 \cap H : |V| < \rho/2\} = \{X = U = 0, Y = -\sqrt{2+V^2}\} \cup \{X = U = 0, Y = \sqrt{2+V^2}\}$. It follows that one of the functions Y_i, say Y_1, is positive on $|V| \leq \rho/2$ and the other, Y_2, is negative there. By the preceding discussion $Y_1'(\rho/2) > \frac{1}{2}$, and $Y_1'(\rho/2) < -\frac{1}{2}$, and $Y_2'(\rho/2) < -\frac{1}{2}$, and $Y_2'(\rho/2) > \frac{1}{2}$ provided that η is small enough. Since for $|V| \geq \rho/4$, either $Y_i'(V) > \frac{1}{2}$ or $Y_i'(V) < -\frac{1}{2}$, it follows, by the continuity of Y_i' that Y_1 decreases on $(-\infty, -\rho/4)$ and increases on $(\rho/4, \infty)$ and that Y_2 increases on $(-\infty, \rho/4)$ and decreases on $(\rho/4, \infty)$. This establishes (c) and completes the proof of Lemma 2.1.

We now want to understand how a surface S passing through the origin and almost tangent to $H = \{(z, w) :\ Re(z + w) = 0\}$ intersects the varieties $V_t = \{(z, w) :\ zw = t^2\}$ for small $t > 0$ in a neighborhood of the origin. To do this, we consider a $\mathcal{C}^1$ function ψ on $\mathbb{R}^3$ that satisfies $\psi(0) = 0$ and consider how $S = Gr(\psi)$ intersects V_t.

The main fact we use is the homogeneity condition that $V_t = tV_1$ when $t > 0$. This implies that $V_t \cap Gr(\psi) = t(V_1 \cap \frac{1}{t}Gr(\psi))$. Notice that $\frac{1}{t}Gr(\psi) = Gr(\varphi)$ if $\varphi(X, Y, V) = \frac{1}{t}\psi(tX, tY, tV)$. Also $(D\varphi)(X, Y, V) = (D\psi)(tX, tY, tV)$, so $|(D\varphi)(p)| < \tau$ for all $p \in \mathbb{R}^3$ if and only $|(D\psi)(p)| < \tau$ for all $p \in \mathbb{R}^3$.

If η, ρ, and M are as in Lemma 2.1, then the conditions (2.1) for $\varphi(p) = \frac{1}{t}\psi(tp)$ are

$$\left|\frac{1}{t}\psi(tp)\right| < \eta \quad (|p| < \rho) \text{ and } |(D\psi)(p)| < \eta \quad (p \in \mathbb{R}^3). \tag{2.9}$$

Lemma 2.5. *There are $\tau > 0$ and $M < \infty$ such that if the smooth real function ψ on $\mathbb{R}^3$ satisfies*

$$\psi(0) = 0 \quad \textit{and} \quad |(D\psi)(p)| < \tau \quad (p \in \mathbb{R}^3), \tag{2.10}$$

then for every $t > 0$

(a) *$Gr(\psi)$ is transverse to V_t, and*
(b) *$Gr(\psi) \cap V_t = \Lambda_1 \cup \Lambda_2$ where Λ_1 and Λ_2 are $\mathcal{C}^1$ curves of the form*

$$\Lambda_j = \{(X_j(V), Y_j(V), U_j(V), V) :\ V \in \mathbb{R}\}$$

where the functions X_j, Y_j, and U_j are of class $\mathcal{C}^1$ on $\mathbb{R}$ and where the first derivative of each is bounded uniformly by M, and
(c) *$Y_1(V) > 0$ and $Y_2(V) < 0 \quad (V \in \mathbb{R})$.*

Moreover, given $\epsilon > 0$, there is $\delta > 0$ such that if

$$W = \{(X, Y, U, V) :\ |V| < \delta\}, \tag{2.11}$$

then for all $t > 0$

$$\text{length } (W \cap V_t \cap Gr(\psi)) < \epsilon.$$

Proof. Let η, ρ and M be as in Lemma 2.1. This lemma and the preceding discussion imply that to prove (a)-(c) it suffices to show that $\tau > 0$ in (2.10) can be chosen so small that (2.10) implies (2.9) for every $t > 0$. Let $\tau = \min\{\eta, \eta/\rho\}$, and assume that ψ satisfies (2.10). The second inequality in (2.9) is obviously satisfied. To prove the first inequality, let $|p| < \rho$. Then (2.10) implies that

$$\left|\frac{1}{t}\psi(tp)\right| = \left|\frac{1}{t}\int_0^t (D\psi)(\lambda p)\cdot p d\lambda\right| \le \frac{1}{t}\tau|p|t < \frac{\eta}{\rho}\rho = \eta.$$

Finally, if $\delta > 0$ is so small that $\delta\sqrt{3M^2+1} < \epsilon$, and if W is given by (2.11) then

$$\begin{aligned}\text{length}\,(W \cap V_t \cap Gr(\varphi)) &= \sum_{j=1}^{2}\int_{-\delta}^{\delta}\sqrt{X_j'(t)^2 + Y_j'(t)^2 + U_j'(t)^2 + 1}\,dt \\ &\le 4\delta\sqrt{3M^2+1} < \epsilon.\end{aligned}$$

This completes the proof.

Appendix to Section 2.

The proof of Lemma 2.4 depends on the following standard result:

Lemma 2.A.1. *Let $\mathcal{U}$ be a neighborhood of $0 \in \mathbb{R}^n$, and let $F : \mathcal{U} \to \mathbb{R}^n$ be a $\mathcal{C}^1$ map with $F(0) = 0$ and $(DF)(0) = I$. If $r > 0$ is so small that $|(DF)(0)(x) - (DF)(0)| < \frac{1}{2}$ when $|x| < r$, then for each y, $|y| < \frac{r}{2}$, there is precisely one x, $|x| < r$, such that $F(x) = y$.*

Proof. Let $G(x) = F(x) - x$. Then $|(DG)(x)| < \frac{1}{2}$ when $|x| < r$. Thus, if $|x|$, $|y| < r$, then $|G(x) - G(y)| = |\int_0^1 (DG)(x+t(y-x))(y-x)dt| \le \frac{1}{2}|x-y|$, so $\frac{1}{2}|x-y| \ge |G(x)-G(y)| = |F(x)-F(y)-(x-y)| \ge |x-y|-|F(x)-F(y)|$, and $|F(x) - F(y)| \ge \frac{1}{2}|x-y|$, whence F is one-to-one on $W = \{|x| < r\}$. Let $|y| < \frac{r}{2}$ and set $x_0 = 0$, $x_n = y - G(x_{n-1})$, $n = 1, 2, \ldots$. Then $|x_{n+1} - x_n| = |G(x_n) - G(x_{n-1})| \le \frac{1}{2}|x_n - x_{n-1}|$. Note that x_n, x_{n-1} do lie in $\mathcal{W} : |x_1| = |y| < \frac{r}{2}$, so $|x_1 - x_0| < \frac{r}{2} \le \frac{1}{2}|x_2 - x_1| \le \frac{1}{2}|x_1 - x_0| \le \frac{1}{2}\frac{r}{2}$, etc. Thus, $x_n \to x$, and $x = y - F(x) + x$, whence $y = F(x)$. This completes the proof of the Lemma.

Lemma 2.A.2. *Let $\mathcal{U}$ be a neighborhood of the origin in $\mathbb{R}^n$, and let $F : \mathcal{U} \to \mathbb{R}^n$ be a $\mathcal{C}^1$ map with $F(0) = 0$ and $(DF)(0)$ nonsingular. If $r > 0$ is so small that*

$$|(DF)(x) - (DF)(0)| < \frac{1}{2|(DF)(0)^{-1}|} \quad \textit{when} \quad |x| < r$$

then for each z *with* $|z| < \frac{r}{2|(DF)(0)^{-1}|}$ *there is a unique* x, $|x| < r$, *such that* $F(x) = z$.

Proof. Consider the map $x \mapsto (DF)(0)^{-1} \circ F(x) = G(x)$. We have that $G(0) = 0$ and that $|(DG)(x) - (DG)(0)| = |DF(0)^{-1}[(DF)(x) - (DF)(0)]| \leq |(DF)(0)^{-1}| \cdot |(DF)(x) - (DF)(0)| < \frac{1}{2}$ when $|x| < r$. By Lemma 2.A.1, given y, $|y| < \frac{r}{2}$, there is a unique x, $|x| < r$, such that $G(x) = y$, i.e., $F(x) = (DF)(0)(y)$. Consequently, for each $z \in \frac{r}{2}(DF)(0)(\mathbb{B})$, there is a unique x, $|x| < r$, such that $F(x) = z$. As $\frac{1}{|(DF)(0)^{-1}|}\mathbb{B} \subset (DF)(0)(\mathbb{B})$ it follows that for each z, $|z| < \frac{r}{2}\frac{1}{|DF(0)^{-1}|}$ there is precisely one x with $|x| < r$ such that $F(x) = z$. This completes the proof.

Proof of Lemma 2.4. Put $G(x) = F(x) + \varphi(x) - \varphi(0)$ so that $G(0) = 0$, and $(DG)(x) = (DF)(x) + (D\varphi)(x)$. We have

$$\begin{aligned}(DG)(0)^{-1} &= [(DF)(0) + (D\varphi)(0)]^{-1} \\ &= (DF)(0)^{-1} \sum_{k=0}^{\infty} \left[-(DF)(0)^{-1} D\varphi(0)\right]^k.\end{aligned}$$

By (2.7) $|(DF)(0)^{-1} D\varphi(0)| \leq |(DF)(0)^{-1}||(D\varphi)(0)| < \frac{1}{2}$, so

$$|(DG)(0)^{-1}| \leq 2|(DF)(0)|^{-1}. \tag{2.12}$$

Further, if $|x| < r$, then by (2.7) and (2.8)

$$\begin{aligned}|(DG)(x) - (DG)(0)| &\leq |(DF)(x) - (DF)(0)| + |(D\varphi)(0)| + |(D\varphi)(x)| \\ &< \frac{1}{8}\frac{1}{|DF(0)^{-1}|} + \frac{1}{8}\frac{1}{|DF(0)^{-1}|} \\ &< \frac{1}{4}\frac{1}{|DF(0)^{-1}|} \leq \frac{1}{2|DG(0)^{-1}|}\end{aligned}$$

so by Lemma 2.A.2, for each $|z| < \frac{r}{2|DG(0)^{-1}|}$ there is a unique x, $|x| < r$, such that $G(x) = z$. In particular, by (2.8) and (2.12), this holds for $z = -\varphi(0)$. Accordingly, there is a unique x, $|x| < r$, such that $F(x) + \varphi(x) = 0$. This completes the proof of Lemma 2.4.

3. Proof of Theorem 1.1.

We begin with two lemmas of a general character, which will be used in the main part of the proof of the theorem.

The proof of the first of these is quite short but is decidedly nontrivial, as it depends essentially on some complicated results in the theory of polynomial convexity.

Lemma 3.1. *If $\varphi : \bar{\Delta} \to \mathbb{C}^N$ is a continuous map that is holomorphic on Δ, that is one-to-one on $\bar{\Delta}$ and that carries $b\Delta$ onto a rectifiable simple closed curve, then the set $X = \varphi(\bar{\Delta})$ is polynomially convex.*

Proof. Let Σ be the image of $b\Delta$ under φ, and denote by $\widehat{\Sigma}$ the polynomially convex hull of Σ. It is a result of Alexander [**Al1**] that $\widehat{\Sigma} \setminus \Sigma$ is a closed analytic subvariety of the domain $\mathbb{C}^N \setminus \Sigma$. Moreover, by [**Al2**], this variety is irreducible. As $\varphi(\Delta)$ is a closed analytic subvariety of $\mathbb{C}^N \setminus \Sigma$ that is contained in $\widehat{\Sigma}$, it follows that, as claimed, $X = \widehat{\Sigma}$.

Lemma 3.2. *Let $P \subset \mathbb{C}^2$ be a bounded domain. Let $V_t = \{(z, w) : zw = t^2\}$, $t > 0$, and suppose that $V_t \cap P \neq \emptyset$ and that $V_t \cap bP = \Gamma$ is a simple closed curve. Then $V_t \cap P$ is biholomorphically equivalent to a disc.*

Proof. The projection $\pi : \mathbb{C}^2 \to \mathbb{C}$ given by $\pi(z, w) = z$ carries V_t biholomorphically onto $\mathbb{C} \setminus \{0\}$. Let Q be the bounded component of $\mathbb{C} \setminus \pi(\Gamma)$. Since P is bounded, it follows that if $z \in \mathbb{C} \setminus \bar{Q}$, $z \neq 0$, then $(z, \frac{t^2}{z}) \in \mathbb{C}^2 \setminus \bar{P}$. We show that $0 \notin Q$. Otherwise, there are points $z_n \in Q$ with $z_n \to 0$ such that $(z_n, \frac{t^2}{z_n}) \in P$ for all n, contradicting the boundedness of P. Thus, $0 \notin Q$, and $\{(z, \frac{t^2}{z}) : z \in Q\} = P \cap V_t$. As Q is biholomorphically equivalent to a disc, the same is true of $P \cap V_t$. This completes the proof.

We now begin the proof of Theorem 1.1 itself. Initially we work in $\mathbb{C}^2$. We shall deal first with the case that Ω is strictly convex at z_o.

By the assumption of convexity, there are a ball $\mathbb{B}(z_o, r)$ centered at z_0 and a function ϱ on $\mathbb{B}(z_0, r)$ with nonvanishing gradient and positive definite Hessian such that $\mathbb{B}(z_0, r) \cap b\Omega = \{w \in \mathbb{B}(z_0, r) : \varrho(w) = 0\}$ and $\mathbb{B}(z_0, r) \cap \Omega = \{w \in \mathbb{B}(z_o, r) : \varrho(w) < 0\}$. Denote by $\Lambda(z_0, b\Omega)$ the complex line passing through z_0 and tangent to $b\Omega$ at z_0. There is a neighborhood $\mathcal{L}$ of $\Lambda(z_0, b\Omega)$ in the space of all complex lines in $\mathbb{C}^2$ such that if $\Lambda \in \mathcal{L}$ meets $\mathbb{B}(z_0, r) \cap \Omega$, then Λ meets $\mathbb{B}(z_0, r) \cap b\Omega$ transversely in a closed curve that bounds a convex domain in Λ. We shall show, after passing to a smaller $\mathcal{L}$ if necessary, that

$$\int_{\Lambda \cap \mathbb{B}(z_0, r) \cap b\Omega} f\omega = 0 \quad (\Lambda \in \mathcal{L}) \tag{3.1}$$

for all $\omega \in \mathbb{C}^{1;0}[dz]$. By [**GlS**, Th. 3.2.1] this implies that f is a CR-function in a neighborhood of z_0.

Given $\Lambda \in \mathcal{L}$ which meets $\mathbb{B}(0, r) \cap \Omega$ we shall show that the neighbourhood $\mathcal{D}$ contains transversely embedded analytic discs D_n, $n \in \mathbb{N}$, and D, such that as $n \to \infty$, bD_n converges to $bD \cup [\Lambda \cap \mathbb{B}(z_0, r) \cap b\Omega]$ in the sense that for every continuous 1-form γ on $\mathbb{C}^2$,

$$\lim_{n \to \infty} \int_{bD_n} \gamma = \int_{bD} \gamma + \int_{\Lambda \cap \mathbb{B}(z_0, r) \cap b\Omega} \gamma. \tag{3.2}$$

If $\omega \in \mathbb{C}^{1;0}[dz]$, then by hypothesis $\int_{bD_n} f\omega = 0, (n \in \mathbb{N})$, and $\int_{bD} f\omega = 0$. Applying (3.2) with $\gamma = f\omega$ gives (3.1).

To construct such sequences, first consider the following special case. We again use the notation that $V_t = \{(z,w) \in \mathbb{C}^2 : zw = t^2\}$ $(t > 0)$.

Lemma 3.3. *Let $\Omega \subset \mathbb{C}^2$ be a bounded domain with boundary of class $\mathcal{C}^2$. Let $0 \in b\Omega$, and let* $\mathrm{Re}(z+w) = 0$ *be the tangent plane to $b\Omega$ at 0. Denote by L_1 and L_2 the coordinate axes of $\mathbb{C}^2$, which we assume to intersect $b\Omega$ transversely and in such a way that $L_j \cap b\Omega = \Lambda_j$ is a simple closed curve that bounds a domain $D_j \subset L_j$, $j = 1, 2$. There is then $\tau > 0$ such that if $0 < t \le \tau$ and if E is a unitary map of $\mathbb{C}^2$ to itself with $|E - I| < \tau$, then $E(V_t)$ is transverse to $b\Omega$ and $D_{tE} = E(V_t) \cap \Omega$ is a transversely embedded analytic disc. Moreover, if $\delta > 0$ is sufficiently small, then $\overline{D_{tE}} \setminus \delta\mathbb{B}$ is an arbitrarily small perturbation of $[\overline{D_1} \setminus \delta\mathbb{B}] \cup [\overline{D_2} \setminus \delta\mathbb{B}]$ provided that $t > 0$ and $|E - I|$ are sufficiently small.*

Proof of Lemma 3.3.

Step 1. If $r > 0$ is sufficiently small, then $b(r\mathbb{B})$ is transverse to Λ_1 and to Λ_2, and each Λ_i intersects $b(r\mathbb{B})$ at exactly two points. In addition, $b(r\mathbb{B})$ intersects $b\Omega$ transversely in a slight perturbation of $H \cap b(r\mathbb{B})$ where $H = \{Re(z+w) = 0\}$. Both $b(r\mathbb{B})$ and $b\Omega$ are transverse to L_1 and L_2. For each i, $D_i \setminus r\overline{\mathbb{B}}$ is bounded by the union of an arc $\lambda_i = \Lambda_i \setminus r\mathbb{B}$ and a circular arc γ_i contained in $b(r\mathbb{B})$.

On any compact set missing the origin $E(V_t)$ is an arbitrarily small perturbation of $L_1 \cup L_2$ provided that $t > 0$ and $|E - I|$ are small enough. In this case, $b\Omega$ and $b(r\mathbb{B})$ cut out of V_t two domains $D_i(t, E, r)$, and by the transversality mentioned above, each $D_i(t, E, r)$ is a slight perturbation of $D_i \setminus r\overline{\mathbb{B}}$. It is bounded by two smooth arcs, one lying near λ_i, the other near γ_i.

Suppose for the moment that we have proved that

(**B**) there is $\tau > 0$ such that for each t, $0 < t \le \tau$, $E(V_t)$ is transverse to $b\Omega$ and $E(V_t) \cap b\Omega$ is a simple closed curve Λ_{Et}.

For $0 < t \le \tau$ define D_{tE} to be $\Omega \cap E(V_t)$. By Lemma 3.2 each D_{tE} is a transversely embedded analytic disc.

Step 2. We first describe what we are going to do to prove (B). If $V_0 = \{zw = 0\} = L_1 \cup L_2$, then $V_0 \cap b\Omega$ is a figure eight, a union of two simple closed curves, Λ_1 and Λ_2, in $\mathbb{C}^2$ which meet only at the origin. (Note that they are not tangent to each other there.) Fix a small ball $\mathcal{B}$ centered at the origin. Outside $\mathcal{B}$ the variety V_t is a small perturbation of V_0 when $t > 0$ is small enough. It follows that $b\Omega \cap E(V_t) \setminus \mathcal{B}$ is, for small $t > 0$, a small perturbation of $(\Lambda_1 \setminus \mathcal{B}) \cup (\Lambda_2 \setminus \mathcal{B})$, which is a union to two disjoint arcs. Now we use Lemma 2.5 to analyze what happens in $\mathcal{B}$ as we pass from

$t = 0$ to a small positive t. A careful analysis will show that $\mathcal{B} \cap V_0$, the cross in the figure eight, gets replaced with two arcs that connect the endpoints of the two arcs whose union is $[b\Omega \cap E(V_t)] \setminus \mathcal{B}$ in such a way that the union of all four arcs is a simple closed curve.

Now let us give the details of the proof of (B). Let $\delta > 0$ be smaller than τ in Lemma 2.2. Since $\mathrm{Re}\,(z+w) = 0$ is tangent to $b\Omega$ at 0, it follows that there are $r > 0$ and a smooth function ψ on $\mathbb{R}^3$ such that $\psi(0) = 0$, $|(D\psi)(p)| < \frac{\delta}{2}$ $(p \in \mathbb{R}^3)$ and such that $\mathbb{B}(0,r) \cap Gr(\psi) = \mathbb{B}(0,r) \cap b\Omega$. There is a $\nu > 0$ such that if E is a unitary map with $|E - I| < \nu$, then $E(b\Omega)$ is transverse to both L_1 and L_2, and there is a unique function ψ_E, smooth on $\mathbb{R}^3$ with $\psi_E(0) = 0$, $|(D\psi_E)(p)| < \delta$ $(p \in \mathbb{R}^3)$ and such that $E(Gr(\psi)) = Gr(\psi_E)$.

Provided that δ has been chosen small enough Lemma 2.2 shows that $Gr(\psi_E) \cap L_i = \{(\Phi_{iE}(V), V) : V \in \mathbb{R}\}$ where $\Phi_{iE} = (X_{iE}, Y_{iE}, U_{iE})$ is smooth on $\mathbb{R}$ and $\Phi'_{1E}(V)$ is arbitrarily close to $(0,-1,0)$, and $\Phi'_{2E}(V)$ is arbitrarily close to $(0,1,0)$ uniformly in $V \in \mathbb{R}$ provided that δ and γ are small enough. In particular, we may assume that for $V \in \mathbb{R}$

(3.3)
$$|X'_{iE}(V)| < \frac{1}{2},\ |U'_{iE}(V)| < \frac{1}{2},\ -\frac{3}{2} < Y'_{1E}(V) < -\frac{1}{2},\ \frac{1}{2} < Y'_{2E}(V) < \frac{3}{2}.$$

Thus, Y_{1E} is strictly decreasing, Y_{2E} is strictly increasing, and since $Y_{iE}(0) = 0$ it follows that $Y_{1E}(V) > 0$ and $Y_{2E}(V) < 0$ when $V < 0$ while $Y_{1E}(V) < 0$ and $Y_{2E}(V) > 0$ when $V > 0$.

Step 3. We have to take into account that $Gr(\psi_E)$ coincides with $E(\Omega)$ only within $\mathbb{B}(0,r)$. With M as in Lemma 2.5, choose $\omega > 0$ so small that if

$$\mathcal{A} = \{|X|,\ |Y|,\ |U| < (M+2)\omega \text{ and } |V| < \omega\}$$

then $\mathcal{A} \subset B(0,R)$. The bounds (3.3) imply that

$$|X_{iE}(\pm\omega)| < \frac{1}{2}\omega,\ |U_{iE}(\pm\omega)| < \frac{1}{2}\omega,\ \text{ and } |Y_{iE}(\pm\omega)| < \frac{3}{2}\omega. \tag{3.4}$$

As δ is smaller than the τ of Lemma 2.5, this lemma implies that for each $t > 0$,

$$Gr(\psi_E) \cap V_t = \{(\Phi_{1Et}(V), V) : V \in \mathbb{R}\} \cup \{(\Phi_{2Et}(V), V) : V \in \mathbb{R}\}$$

where $\Phi_{iEt} = (X_{iEt}, Y_{iEt}, U_{iEt})$ is smooth on $\mathbb{R}$ and $Y_{1Et}(V) > 0$, and $Y_{2Et}(V) < 0$ for all $V \in \mathbb{R}$.

If $t_0 > 0$ and $\gamma > 0$ are sufficiently small, then by transversality, the set $\{\Phi_{1Et}(\omega), \Phi_{2Et}\ (\omega)\}$ is a small perturbation of the set $\{\Phi_{1I}(\omega), \Phi_{2I}(\omega)\}$ and the set $\{\Phi_{1Et}(-\omega), \Phi_{2Et}(-\omega)\}$ is a small perturbation of the set $\{\Phi_{1I}(-\omega), \Phi_{2I}(-\omega)\}$ whenever $0 < t \le t_0$ and $|E - I| < \gamma$. In particular by (3.4) we may suppose that

$$|X_{iEt}(\pm\omega)| < \omega,\ |U_{iEt}(\pm\omega)| < \omega \text{ and } |Y_{iEt}(\pm\omega)| < 2\omega. \tag{3.5}$$

Since by Lemma 2.5 $|X'_{iEt}|$, $|Y'_{iEt}|$, $|U'_{iEt}| < M$ on $\mathbb{R}$, (3.5) implies that for $|V| < \omega$,

$$|X_{iEt}(V)| < (M+1)\omega,\ |U_{iEt}(V)| < (M+1)\omega, \text{ and } |Y_{iEt}(V)| < (M+2)\omega$$

so, provided that t_0 and γ are small enough $(X_{iEt}(V), Y_{iEt}, U_{iEt}(V), V) \in \mathcal{A}$ when $|V| < \omega$. Since $\mathcal{A} \subset \mathbb{B}(0,r)$ where $Gr(\psi_E)$ coincides with $E(b\Omega)$, it follows that $\mathbb{B}(0,r) \cap Gr(\psi_E) \cap V_t \cap \{|V| < \omega\} = \mathbb{B}(0,r) \cap E(b\Omega) \cap V_t \cap \{|V| < \omega\}$.

To see that $E(V_t) \cap b\Omega$ is a simple closed curve if $t > 0$ and $|E - I|$ are small is equivalent to seeing that $V_t \cap E(b\Omega)$ is a simple closed curve provided that $t > 0$ and $|E - I|$ are small.

To see that $V_t \cap E(b\Omega)$ is a simple closed curve, observe first that $\Lambda_i = \Lambda_{iI}$ consists of a short arc $\Lambda_i^s = \{(\Phi_{iI}(V), V) : |V| < \omega\}$ and a long arc Λ_i^ℓ, which joins the points $(\Phi_{iI}(\omega), \omega)$ and $(\Phi_{iI}(-\omega), -\omega)$. If $\omega > 0$ is small enough then the long arc Λ_E^ℓ meets $\overline{\mathcal{A}}$ only at its endpoints at which it is transverse to the hyperplanes $V = \omega$ and $V = -\omega$, respectively. This transversality, together with the transversality of L_1 and L_2 to $b\Omega$ implies that when $t = 0$ changes to t, $0 < t \leq t_0$ and I changes to E, $|E - I| < \gamma$, then, provided that t_0 and γ are small enough, the long arc Λ_i^ℓ will change arbitrarily little to an arc Λ_{iEt}^ℓ with endpoints (T_{iEt}^+, ω) close to $(\Phi_{iI}(\omega), \omega)$ and $(T_{iEt}^-, -\omega)$ close to $(\Phi_{iI}(-\omega), -\omega)$, which will still meet $\overline{\mathcal{A}}$ only at its endpoints. In particular,

(**C**) the $Y-$coordinates of T_{1Et}^- and T_{2Et}^+ will be positive, and the Y-coordinates of T_{1Et}^+ and T_{2Et}^- will be negative.

We have $\Lambda_{1Et}^\ell \cup \Lambda_{2Et}^\ell = [V_t \cap E(b\Omega)] \setminus \mathcal{A}$.

On the other hand, provided that $t_0 > 0$ and γ are small enough, for every E, $|E - I| < \gamma$, and for every t, $0 < t \leq t_0$, $\mathcal{A} \cap E(b\Omega) \cap V_t = \mathcal{A} \cap Gr(\psi_E) \cap V_t = \{(\Phi_{1Et}(V), V) : |V| < \omega\} \cup \{(\Phi_{2Et}(V), V) : |V| < \omega\}$ where $Y_{1Et}(V) > 0$, and $Y_{2Et}(V) < 0$ for all $V \in \mathbb{R}$. Now V_t intersects $E(b\Omega) \setminus \mathcal{A}$ transversely provided that γ and t_0 are small enough, and by Lemma 2.5, V_t meets $E(b\Omega) \cap \mathcal{A}$ transversely. Thus $E(V_t)$ intersects $b\Omega$ transversely, which implies that $E(V_t) \cap b\Omega$ is a closed, one-dimensional submanifold of $\mathbb{C}^2$. As $V_t \cap E(b\Omega) \setminus \mathcal{A} = \Lambda_{1Et}^\ell \cup \Lambda_{2Et}^\ell$ and $V_t \cap (b\Omega) \cap \mathcal{A} = \{(\Phi_{1Et}(V), V) : |V| < \omega\} \cup \{(\Phi_{2Et}(V), V) : |V| < \omega\}$, it follows that

$$E(b\Omega) \cap V_t = \Lambda_{1Et}^\ell \cup \Lambda_{2Et}^\ell \cup \{(\Phi_{1Et}(V), V) : |V| < \omega\} \cup \{(\Phi_{2Et}(V), V) : |V| < \omega\}. \tag{3.6}$$

The last two arcs are contained in $\mathcal{A}$, and the first two miss $\mathcal{A}$. As $\Lambda_1^\ell \cap \Lambda_2^\ell = \emptyset$, it follows that $\Lambda_{1Et}^\ell \cap \Lambda_{2Et}^\ell = \emptyset$ if t_0 and γ are small enough. The last two arcs in (3.6) are also disjoint since $Y_{1Et}(V) > 0$ and $Y_{2Et}(V) < 0$ for all $V \in \mathbb{R}$. This, together with **(C)**, implies that the only way for the right side of (3.6) to be a closed submanifold of $\mathbb{C}^2$ is for

$$\Phi_{1Et}(-\omega) = T^-_{1Et} \quad \Phi_{1Et} = T^+_{2Et}$$

and

$$\Phi_{2Et}(-\omega) = T^-_{2Et} \quad \Phi_{2Et}(\omega) = T^+_{1Et}.$$

Thus, $E(b\Omega)\cap V_t$ consists of the long arc Λ^ℓ_{1Et} joining (T_1^+, ω) with $(T_1^-, -\omega)$ followed by the arc $\{(\Phi_{1Et}(V), V) : |V| < \omega\}$ joining $(T_1^-, -\omega)$ with (T_2^+, ω) followed by the long arc Λ^ℓ_{2Et} joining (T_2^+, ω) with $(T_2^-, -\omega)$ followed finally by the arc $\{(\Phi_{Et}(V), V) : |V| < \omega\}$ joining $(T_2^-, -\omega)$ with (T_1^+, ω). This proves that $E(b\Omega)\cap V_t$ is a simple closed curve provided that $t > 0$ and $|E - I|$ are small enough.

This completes the proof of Lemma 3.3.

Discussion 1. Given $\epsilon > 0$, the ω in the proof of Lemma 3.3 can be chosen so small that $4\omega\sqrt{3M^2+1} < \epsilon$. This implies that the length of $E(b\Omega)\cap\mathcal{A}\cap V_t$ does not exceed ϵ. Indeed, length$(E(b\Omega)\cap\mathcal{A}\cap V_t) =$ length$(V_t\cap Gr(\psi_E)\cap\mathcal{A}) = \int_{-\omega}^{\omega}[\sum_{j=1}^2[X'_{jEt}(v)^2 + Y'_{jEt}(v)^2 + V'_{jEt}(v)^2 + 1]]^{\frac{1}{2}}dv$, and since $|X'_{jEt}(v)|$, $|Y'_{jEt}(v)|$, $|V'_{jEt}(v)|$ are all bounded by M, it follows that length$(E(b\Omega)\cap\mathcal{A}\cap V_t) \le 4\omega\sqrt{3M^2+1} < \epsilon$.

Discussion 2. The implicit mapping theorem implies that given an interior point w of the long arc Λ^ℓ_i, there are a neighborhood $W(w)$ of w, a $\nu(w) > 0$, and a $t_0(w) > 0$ such that $|E - I| < \nu(w)$ and $0 < t \le t_0(w)$ imply that $\Lambda^\ell_{iEt}\cap W(w)$ is a smooth graph over $\Lambda^\ell_i\cap W(w)$. These graphs depend smoothly on E and t. The same is true at the endpoints of Λ^ℓ_i. Since we want the endpoints of Λ^ℓ_{iEt} to belong to $|V| = \omega$, we have to write, e.g., in a sufficiently small neighborhood $W(w)$ of $w = (\Phi_{iI}(\omega), \omega)$, $\Lambda_{iI}\cap W(w) = \{(\Phi_{iI}(t), t) : \omega \le t < \omega + \gamma\}$ for some small $\gamma > 0$, and then $\Lambda^\ell_{iEt}\cap W(w) = \{(\Phi_{iEt}(t), t) : \omega \le t < \omega + \gamma\}$. A partition of unity argument together with the compactness of Λ^ℓ_i shows that given a smooth 1-form α on $\mathbb{C}^2$, we have

$$\lim_{E\to I,\ t\to 0}\int_{\Lambda^\ell_{iEt}}\alpha = \int_{\Lambda^\ell_i}\alpha. \tag{3.7}$$

Similar reasoning applies to show that if t_0 and ν are small enough, then the lengths of Λ^ℓ_{iEt}, $|E-I| < \nu$, $0 < t \le t_0$, are uniformly bounded. We already know that the lengths of $E(b\Omega)\cap V_t\cap\mathcal{A}$ are uniformly bounded. Thus, the lengths of $E(b\Omega)\cap V_t$ are uniformly bounded provided that $|E-I| < \nu$, $0 < t \le t_0$.

Discussion 3. Let $\pi_1(z,w) = z, \pi_2(z,w) = w$ be the coordinate projections in $\mathbb{C}^2$. For each t, $0 < t < \tau$, let $D^i_{tE} = \pi_i(E^{-1}(D_{tE}))$, $i = 1, 2$. Since the $E^{-1}(D_{tE})$ are discs in V_t, $0 < t < \tau$, the D^i_{tE} are Jordan domains in $\mathbb{C}$, bounded by simple closed curves $\pi_i(E(V_t)\cap b\Omega)$, which, when $t\to 0$, $E\to I$, by the last statement of Lemma 3.3, converge to D_i in the following sense: Given $\varepsilon>0$ there is a δ, $0<\delta<\tau$, such that if $0<t<\delta$ and $|E-I|<\delta$ then there is a homeomorphism $\psi^i_{tE}\colon bD_i\to bD^i_{tE}$ such that $|\psi^i_{tE}(w)-w|<\varepsilon$ $(w\in bD_i)$, $i=1,2$. This implies [**Po**, p. 26] that given a conformal map $\varphi_1\colon\Delta\to D_1$ and $\varepsilon>0$ there is a δ, $0<\delta<\tau$, such that whenever $0<t<\delta$ and $|E-I|<\delta$, there is a conformal map $\varphi^1_{tE}\colon\Delta\to D^1_{tE}$ such that $|\varphi^1_{tE}-\varphi_1|<\varepsilon$ on Δ which further implies that there is a parametrization $\zeta\mapsto(\varphi^1_{tE}(\zeta),t^2/\varphi^1_{tE}(\zeta))=\Phi_{tE}(\zeta)$ of $E^{-1}(D_{tE})$ such that $|\Phi_{tE}(\zeta)-\Phi_1(\zeta)|<2\varepsilon+\operatorname{diam}(D_2)$ $(\zeta\in\Delta)$ whenever $0<t<\delta$, $|E-I|<\delta$ where $\Phi_1(\zeta)=(\varphi_1(\zeta),0)$ is a parametrization of D_1. In particular, given a neighbourhood $\mathcal{D}$ of D_1 in the space of transversely embedded analytic discs, $D_{tE}\in\mathcal{D}$ $(0<t<\delta,|E-I|<\delta)$ provided that $\operatorname{diam}(D_2)$ and δ are small enough.

Let α be a continuous 1-form on C^2 with compact support. Given $\epsilon>0$, there is a $\delta>0$ such that for every smooth compact curve Σ in $\mathbb{C}^2$, we have

$$\left|\int_{E(\Sigma)}\alpha-\int_\Sigma\alpha\right|=\left|\int_\Sigma E^*\alpha-\alpha\right|<\epsilon.\text{length}(\Sigma)$$

whenever E is a unitary map with $|E-I|<\delta$.

Lemma 3.4. *Let $\Omega,\Lambda_1,\Lambda_2$ and V_t, $t>0$, be as in Lemma* 3.3. *If α is a continuous* 1*-form on $\mathbb{C}^2$, then*

$$\lim_{E\to I,\ t\to 0}\int_{b\Omega\cap E(V_t)}\alpha=\int_{\Lambda_1\cup\Lambda_2}\alpha=\int_{b\Omega\cap V_0}\alpha. \tag{3.8}$$

Proof. With no loss of generality we suppose that α has compact support. Suppose that we have proved that

$$\lim_{E\to I,t\to 0}\int_{E(b\Omega)\cap V_t}\alpha=\int_{\Lambda_1\cup\Lambda_2}\alpha. \tag{3.9}$$

Since the lengths of $E(b\Omega)\cap V_t$, $|E-I|<\delta, 0<t<t_0$, are uniformly bounded, the preceding discussion implies that given $\epsilon>0$ there is a $\delta>0$ such that

$$\left|\int_{\tilde E(E(b\Omega)\cap V_t)}\alpha-\int_{E(b\Omega)\cap V_t}\alpha\right|<\epsilon$$

whenever $|\tilde E-I|<\delta$, $|E-I|<\gamma$, and $0<t<t_0$. In particular, taking $\tilde E=E^{-1}$ yields

$$\left|\int_{b\Omega\cap E^{-1}(V_t)}\alpha-\int_{E(b\Omega)\cap V_t}\alpha\right|<\epsilon$$

whenever $|E^{-1} - I| < \delta, |E - I| < \gamma$, $0 < t < t_0$. Now $E \to I$ is equivalent to $E^{-1} \to I$, so

$$\lim_{E^{-1}\to I, t\to 0} \int_{b\Omega\cap E^{-1}(V_t)} \alpha = \lim_{E\to I, t\to 0} \int_{E(b\Omega)\cap V_t} \alpha$$

provided that the limit on the right exists. Thus (3.9) implies (3.8). It remains to prove (3.9).

Let $\epsilon > 0$, and let L be a uniform bound for the coefficients of α. By the Discussion 1 above, one can choose $\omega > 0$, $t_o > 0$, and $\nu > 0$ such that

$$\text{length}(E(b\Omega) \cap \mathcal{A} \cap (\Lambda_1 \cup \Lambda_2)) < \frac{\epsilon}{16L}$$

and

$$\text{length}(E(b\Omega) \cap \mathcal{A} \cap V_t) < \frac{\epsilon}{16L}$$

whenever $|E - I| < \nu$ and $0 < t < t_0$. It follows that when $0 < t < t_0$ and $|E - I| < \nu$,

$$\left|\int_{E(b\Omega)\cap\mathcal{A}\cap V_t} \alpha\right| < \frac{\epsilon}{4} \quad\text{and}\quad \left|\int_{E(b\Omega)\cap\mathcal{A}\cap(\Lambda_1\cup\Lambda_2)} \alpha\right| < \frac{\epsilon}{4}. \tag{3.10}$$

Further, by (3.7), we can pass to smaller $\nu > 0$ and $t_0 > 0$ if necessarily to get

$$\left|\int_{\Lambda^\ell_{iEt}} \alpha - \int_{\Lambda^\ell_i} \alpha\right| < \frac{\epsilon}{4} \quad (0 < t < t_o,\ |E - I| < \nu,\ i = 1, 2). \tag{3.11}$$

Thus, if $0 < t < t_0$ and $|E - I| < \nu$, then (3.10) and (3.11) imply that

$$\begin{aligned}\left|\int_{E(b\Omega)\cap V_t} \alpha - \int_{\Lambda_1\cup\Lambda_2} \alpha\right| \le \sum_{i=1}^{2} \left|\int_{\Lambda^\ell_{iEt}} \alpha - \int_{\Lambda^\ell_i} \alpha\right| &+ \left|\int_{E(b\Omega)\cap\mathcal{A}\cap V_t} \alpha\right| \\ &+ \left|\int_{E(b\Omega)\cap\mathcal{A}\cap(\Lambda_1\cup\Lambda_2)} \alpha\right| \le \epsilon.\end{aligned}$$

This proves (3.9). The proof of Lemma 3.4 is complete.

Lemma 3.5. *Let $\Omega \subset \mathbb{C}^2$ be a bounded domain with $0 \in b\Omega$. Assume that $b\Omega$ is of class $\mathcal{C}^2$ in a neighborhood U of $(L_1 \cup L_2) \cap b\Omega$, L_1, L_2 the coordinate axes, and that L_1, L_2 intersect $b\Omega$ transversely so that $D_j = L_j \cap \Omega$ are transversely embedded analytic discs, $j = 1, 2$. Let $w_1 \in D_1$. There is then a sequence $\{A_n\}_{n=1,2,\ldots} \subset \Omega$ of transversely embedded analytic discs such that $w_1 \in A_n$, $bA_n \subset U$ $(n \in \mathbb{N})$ and such that for each continuous 1-form α on $\mathbb{C}^2$*

$$\lim_{n\to\infty} \int_{bA_n} \alpha = \int_{bD_1} \alpha + \int_{bD_2} \alpha. \tag{3.12}$$

Proof. Suppose to begin with that $b\Omega$ is of class $\mathcal{C}^2$.

Assume first that $Re(z+w)=0$ is the tangent space to $b\Omega$ at 0. By Lemma 3.3 there is $\tau>0$ such that $b\Omega\cap E(V_t)$ is a simple closed curve contained in U that bounds the analytic disc $E(V_t)\cap\Omega$ transverse to $b\Omega$ provided that $|E-I|$ and t are small enough. Since on a compact set missing 0, V_t is an arbitrarily small perturbation of $L_1\cup L_2$ provided that t is small enough, it follows that given τ_n ($\tau_n<\tau$), there are $t_n,\ 0<t_n<\tau_n$ and $E_n,\ |E_n-I|<\tau_n$, such that $w_1\in E_n(V_{t_n})$. Put $A_n=E_n(V_{t_n})\cap\Omega$. Then the A_n are transversely embedded analytic discs and by Lemma 3.4, (3.12) holds for every continuous 1-form on $\mathbb{C}^2$.

In the general case, let $Re(pz+qw)=0$ be the equation of the tangent space to $b\Omega$ at 0. Since L_1,L_2 are transverse to $b\Omega$ it follows that $p\neq 0, q\neq 0$, so $F(z,w)=(\frac{z}{p},\frac{w}{q})$ is an isomorphism of $\mathbb{C}^2$ with $F(0)=0, F(L_i)=L_i$, $i=1,2$, and $Re(z+w)=0$ is the tangent space to $b\tilde{\Omega}$ if $\tilde{\Omega}=F(\Omega)$. We are now in the situation above with Ω replaced by $\tilde{\Omega}$ and with D_j replaced by $\tilde{D}_j=L_j\cap\tilde{\Omega}$. Thus, there is a sequence $\tilde{A}_n\subset\tilde{\Omega}$ of transversely embedded analytic discs such that $b\tilde{A}_n\subset F(U)$, $F(w_1)\in\tilde{A}_n$ and such that

$$\lim_{n\to\infty}\int_{b\tilde{A}_n}\beta=\int_{b\tilde{D}_1}\beta+\int_{b\tilde{D}_2}\beta$$

for every continuous 1-form β on $\mathbb{C}^2$. In particular given a continuous 1-form α on $\mathbb{C}^2$ it follows that if $\beta=(F^{-1*})\alpha$ then

$$\lim_{n\to\infty}\int_{bA_n}F^*(F^{-1*})\alpha=\int_{bD_1}F^*(F^{-1*})\alpha+\int_{bD_2}F^*(F^{-1*})\alpha$$

where $A_n=F^{-1}(\tilde{A}_n)$, which implies (3.12).

If the boundary $b\Omega$ is of class $\mathcal{C}^2$ only in a neighborhood of the intersection $(L_1\cup L_2)\cap b\Omega$, then a small modification of the argument just given is required. The set U can be taken to lie in the subset of $b\Omega$ that is a manifold of class $\mathcal{C}^2$. Then, in the proof of Lemma 3.3 it is enough to assume that $b\Omega$ is of class $\mathcal{C}^2$ only in a neighborhood of $(L_1\cup L_2)\cap b\Omega$, since we are there intersecting $b\Omega$ with varieties that are small perturbations of $L_1\cup L_2$.

Lemma 3.5 is proved.

Discussion. Again, as in Discussion 3 after Lemma 3.3, given a neighbourhood $\mathcal{D}$ of D_1 in the space of transversely embedded analytic discs, A_n can be chosen to belong to $\mathcal{D}$ provided that $\operatorname{diam}(D_2)$ is small enough.

We now continue the proof of Theorem 1.1, but no longer under the restriction to domains in $\mathbb{C}^2$. We are dealing with the case that Ω is strictly convex at z_0.

Lemma 3.6. *Let Ω be a bounded domain in $\mathbb{C}^N$ with $b\Omega$ of class $\mathcal{C}^2$. Suppose that $D_0\subset\Omega$ is a transversely embedded analytic disc. Let $w_0\in D_0$ and $z_0\in$*

bD_0, and suppose that Ω is strictly convex at z_0. Let $\mathcal{D}$ be a neighbourhood of D_0 in the space of transversely embedded analytic discs. Denote by $\Lambda(z_0)$ a complex line that is tangent to $b\Omega$ at z_0. There are an open ball B centered at z_0 and a neighborhood $\mathcal{L}$ of $\Lambda(z_0)$ in the space of all complex lines in $\mathbb{C}^N$ such that for each $L \in \mathcal{L}$ that meets $\Omega \cap B$

(a) *$L \cap B \cap b\Omega$ is a compact convex curve, and*
(b) *there are $A \in \mathcal{D}$ and a sequence $\{A_n\}_{n=1,2,\ldots} \subset \mathcal{D}$ such that $w_0 \in A, w_0 \in A_n$ $(n \in \mathbb{N})$ and such that for each smooth 1-form α on $\mathbb{C}^2$*

$$\lim_{n\to\infty} \int_{bA_n} \alpha = \int_{bA} \alpha + \int_{L\cap Q\cap b\Omega} \alpha. \tag{3.13}$$

As in the two-dimensional analysis, this implies that $\int_{L\cap Q\cap b\Omega} f\omega = 0$. Granted this lemma, Theorem 1.1, in the case of convexity, now follows from the following lemma.

Lemma 3.7. *Let D be a bounded convex domain with bD of class $\mathcal{C}^2$ and strictly convex at $z_0 \in bD$. Let $\mathcal{L}$ be an open set of complex lines in $\mathbb{C}^N$ that contains a line tangent to bD at z_0. If f is a continuous function on bD that with the property that $\int_{L\cap bD} f\alpha = 0$ whenever $L \in \mathcal{L}$ meets D and whenever $\alpha \in \mathbb{C}^{1;0}[dz]$, then f is a CR-function on a neighborhood of z_0 in bD.*

Proof. As $\mathcal{L}$ contains L_0, a complex line tangent to bD at z_0 and is open, the result follows, in the case that $N = 2$, from a result in [**Gl2**]. In the case of arbitrary N, a different analysis is necessary.

Thus, consider the case of general N. Let H_0 be the real hyperplane tangent to bD at z_0, and let $T_0 \subset H_0$ be the complex hyperplane in $\mathbb{C}^N$ that goes through z_0. We shall show that if T is a complex hyperplane in $\mathbb{C}^N$ that is near T_0 and that meets D, then $\int_{T\cap bD} f\vartheta = 0$ for all $(N, N-2)$-forms ϑ on $\mathbb{C}^N$ with constant coefficients. Granted this, the result we want is a consequence of Theorem 3.2.1 of [**GlS**].

The complex hyperplane T_0 is a disjoint union of complex lines parallel to L_0. Continuity and the openness of $\mathcal{L}$ imply the existence of an open set $\mathcal{T}$ in the space of complex hyperplanes in $\mathbb{C}^N$ such that $T_0 \in \mathcal{T}$ and such that each $T \in \mathcal{T}$ is a union of complex lines L each of which is parallel to an element of $\mathcal{L}$ and each of which is either disjoint from $\overline{D}$, meets $\overline{D}$ in a single point, or else meets bD in a small convex curve lying near z_0. If $\mathcal{T}$ is small enough, then as $\mathcal{L}$ is open, each $T \in \mathcal{T}$ is a union of lines $\{L_\lambda\}_{\lambda\in\Lambda(T)}$ parallel to an element of $\mathcal{L}$ with the additional property that if L_λ meets bD, then $L_\lambda \in \mathcal{L}$.

By hypothesis, for a given $T \in \mathcal{T}$, $\int_{L_\lambda\cap bD} f\omega = 0$ for every $\lambda \in \Lambda(T)$. Lemma 2.2.1 of [**GlS**] implies that for each $T \in \mathcal{T}$, $\int_{T\cap bD} f\vartheta = 0$ for each ϑ, an $(N, N-2)-$form on $\mathbb{C}^N$ with constant coefficients. The lemma is proved.

The proof of Lemma 3.6 depends on the following simple observation.

Lemma 3.8. *Let $\delta > 0$, and let V be a closed one-dimensional complex submanifold of an open set in $\mathbb{C}^N$ such that for each ζ, $|\zeta| < \delta$, V intersects the hyperplane $H_\zeta = \{z : z_1 = \zeta\}$ at one point and transversely. Assume that V meets H_0 at the origin. There is a biholomorphic map Φ of $\{z \in \mathbb{C}^N : |z_1| < \delta\}$ onto itself that fixes H_0 and has the property that $\Phi(V \cap \{|z_1| < \delta\}) = \{(\zeta, 0, \dots, 0) : |\zeta| < \delta\}$.*

Proof. Let $V \cap H_\zeta = \varphi(\zeta) = (\zeta, \varphi_2(\zeta), \dots, \varphi_N(\zeta))$, $|\zeta| < \delta$. The map $\varphi : \delta\Delta \to \mathbb{C}^N$ is holomorphic by the transversality. As it satisfies $\varphi(0) = 0$ the map Φ given by

$$\Phi(z_1, \dots, z_N) = (z_1, z_2 - \varphi_2(z_1), \dots, z_N - \varphi_N(z_1))$$

has the desired properties.

Proof of Lemma 3.6. The idea is very simple; we describe it in a special case. Since Ω is strictly convex at z_0 there are an open ball B centered at z_0 and a neighbourhood $\mathcal{L}$ of $\Lambda(z_0)$ in the space of complex lines in $\mathbb{C}^N$ such that if $L \in \mathcal{L}$ and $L \cap B$ meets Ω then $L \cap B$ meets $b\Omega$ transversely in a compact convex curve. Moreover, given a neighbourhood $\mathcal{E}$ of z_0 in $\mathbb{C}^N$, $L \cap B \cap \overline{\Omega} \subset \mathcal{E}$ provided that $\mathcal{L}$ is small enough. Let V be a one dimensional submanifold of an open neighbourhood of $\overline{\Omega}$ such that $V \cap \Omega = D_0$. We find a biholomorphic map G from a neighbourhood Q of $\overline{D_0}$ in $\mathbb{C}^N$ to a domain $G(Q)$ in such a way that $G(z_0) = 0$ and that G maps $Q \cap V$ into the z_1-axis. Shrink $\mathcal{L}$ if necessary so that $L \cap B \cap \overline{\Omega} \subset Q$ whenever $L \in \mathcal{L}$ and assume that $L \in \mathcal{L}$ and that $z_0 \in L$. We modify G so that, in addition, it maps $L \cap B \cap Q$ into the z_2-axis. Then we intersect $G(Q)$ with the two-dimensional subspace M spanned by z_1- and z_2-axes. This gives a domain in the copy M of $\mathbb{C}^2$ to which Lemma 3.5 applies to yield a sequence $\{\tilde{A}_n\}$ of transversely embedded analytic discs whose boundaries $b\tilde{A}_n$, in the sense of that lemma, converge to $G(bD_0) \cup G(L \cap B \cap \Omega)$. Then $A_n = G^{-1}(\tilde{A}_n)$, and $A_0 = D_0$ will do the job. Of course, we must be more careful when L does not pass through z_0.

Step 1. Let V be a closed one-dimensional submanifold of an open neighborhood of $\overline{\Omega}$ that intersects $b\Omega$ transversely so that $D_0 = V \cap \Omega$, $bD_0 = V \cap b\Omega$. By Lemma 3.1 above, $\overline{D_0}$ is polynomially convex, so it has a Stein neighborhood basis. Thus, there are arbitrarily small neighborhoods Q of $\overline{D}_0$ in $\mathbb{C}^N$ that are biholomorphically equivalent, say under F, to a domain P in $\mathbb{C}^N$ and in such a way that $F(V \cap Q) = \{(z_1, \dots, z_N) \in P : z_2 = \cdots = z_N = 0\}$, i.e., that $F(V \cap Q)$ is the intersection of P with the z_1-axis. This follows from a result of Docquier and Grauert on the existence of holomorphic tubular neighborhoods -see [**GR**, pp. 256-257]- and the holomorphic triviality of holomorphic vector bundles over discs. The strict convexity of Ω at z_0

implies that there are an open ball B centered at z_0 and a neighbourhood $\mathcal{L}$ of $\Lambda(z_0)$ in the space of complex lines such that

(D) if $L \in \mathcal{L}$ and $L \cap B$ meets Ω then $L \cap B$ meets $b\Omega$ transversely in a small compact convex curve. Moreover, given a neighbourhood $\mathcal{E}$ of z_0, $\mathcal{L}$ can be chosen so small that $B \cap L \cap \overline{\Omega} \subset \mathcal{E}$ for every $L \in \mathcal{L}$.

Passing to smaller P, Q if necessary we may assume that there are a neighbourhood $\mathcal{T}$ of the z_1-axis in the space of all complex lines passing through $\tilde{w}_0 = F(w_0)$ and a $\delta > 0$ such that whenever a complex line T' is parallel to a line $T \in \mathcal{T}$, $\operatorname{dist}(T, T') < \delta$ then T' intersects $F(Q \cap b\Omega)$ transversely in a simple closed curve that bounds the domain $\mathcal{D}(T') = T' \cap F(Q \cap \Omega)$.

Step 2. Since V is transverse to $b\Omega$ at z_0 it follows that $\Lambda(z_0)$ is not tangent to V at z_0 so the complex tangent line to $F(\Lambda(z_0) \cap B \cap Q)$ at $\tilde{z}_0 = F(z_0)$ does not coincide with the z_1-axis. Thus, after composing F with a unitary map that fixes the z_1-axis we may, after passing to a smaller $\mathcal{L}$ and δ, assume that there are a small open ball $\mathcal{E} \subset B \cap Q$ such that (D) holds, and a neighbourhood $\mathcal{H}$ of $\{z_2 = 0\}$ in the space of complex hyperplanes passing through $\tilde{w}_0$ such that

(E) for each $L \in \mathcal{L}$, $F(\mathcal{E} \cap L)$ intersects each H', a complex hyperplane parallel to an $H \in \mathcal{H}$, $\operatorname{dist}(H, H') < 2\delta$, at precisely one point and transversely.

By passing to a smaller $\mathcal{T}$ we can suppose that for each $T \in \mathcal{T}$ there is an $H \in \mathcal{H}$ such that $T \subset H$.

For each $T \in \mathcal{T}$ let

$$\mathcal{D}(T, \delta) = \cup\{\mathcal{D}(T') : T' \text{ parallel to } T, \operatorname{dist}(T, T') < \delta\}$$

and

$$P(T, \delta) = \cup\{T' \cap P : T' \text{parallel to } T, \operatorname{dist}(T, T') < \delta\}.$$

Note that $\mathcal{D}(T, \delta) = P(T, \delta) \cap F(\Omega \cap Q)$ is a connected component of $P(T, \delta) \setminus S$ where $S = F(b\Omega \cap Q)$. Choose a neighbourhood $\mathcal{P}$ of $\tilde{z}_0$ in Q so small that

(F) if $z \in \mathcal{P}$ and if T is the complex line passing through z and $\tilde{w}_0$ then $T \in \mathcal{T}$ and $\mathcal{P} \subset\subset P(T, \delta)$.

By (D) we can pass to a smaller $\mathcal{L}$ such that

$$L \cap B \cap \overline{\Omega} \subset Q, \quad F(L \cap B \cap \overline{\Omega}) \subset \mathcal{P} \quad (L \in \mathcal{L}). \tag{3.14}$$

Fix $L \in \mathcal{L}$ that meets $\Omega \cap B$. By (D) and by (3.14), L meets $B \cap b\Omega$ transversely in a compact convex curve that bounds the convex domain $L \cap B \cap \Omega \subset\subset Q$. Moreover, by (3.14) we have

(G) $F(L\cap B\cap \Omega) = F(L\cap B)\cap F(\Omega\cap Q)\subset \mathcal{P}$ is a domain in $F(L\cap B\cap Q)$ bounded by the simple closed curve $F(L\cap B\cap b\Omega) = F(L\cap B\cap Q)\cap F(b\Omega\cap Q)\subset \mathcal{P}$.

Recall that

$$P(T,\delta)\cap b\mathcal{D}(T,\delta) = P(T,\delta)\cap F(b\Omega\cap Q)$$

and

$$P(T,\delta)\cap F(Q\cap\Omega) = \mathcal{D}(T,\delta).$$

By (G) and by the fact that $\mathcal{P}\subset\subset P(T,\delta)$ it follows that $F(L\cap B\cap Q)\cap F(\Omega\cap Q)\subset \mathcal{D}(T,\delta)\cap\mathcal{P}$ and $F(L\cap B\cap Q)\cap F(b\Omega\cap Q)\subset b\mathcal{D}(T,\delta)\cap\mathcal{P}\subset S\cap P(T,\delta)$ which implies that

$$F(B\cap L\cap Q)\cap \mathcal{D}(T,\delta) = F(B\cap L\cap Q)\cap F(\Omega\cap Q) \tag{3.15}$$

and

$$F(B\cap L\cap Q)\cap b\mathcal{D}(T,\delta) = F(B\cap L\cap Q)\cap F(b\Omega\cap Q)\subset S\cap P(T,\delta). \tag{3.16}$$

Clearly, $F(B\cap L\cap Q)$ intersects S transversely.

Step 3. Choose $p\in F(L\cap B\cap Q)\cap F(b\Omega\cap Q)$. By (3.14) and by (F) the complex line T passing through p and $\tilde w_0$ belongs to $\mathcal{T}$ and is contained in some $H\in\mathcal{H}$ so by (E), $F(L\cap B\cap Q)$ intersects each complex hyperplane H' parallel to H with $\mathrm{dist}(H,H')<2\delta$ at precisely one point and transversely. Lemma 3.8 now implies that there is a biholomorphic map Φ from $H+2\delta\mathbb{B}$ onto itself that fixes H and maps $F(B\cap L\cap Q)\cap(H+2\delta\mathbb{B})$ onto $p+2\delta\Delta q$ where q is a unit vector orthogonal to H. Note that $P(T,\delta)\subset\subset H+2\delta\mathbb{B}$.

Let $U=\Phi(\mathcal{D}(T,\delta))$ and $\Sigma=\Phi(P(T,\delta)\cap S)=\Phi(b\mathcal{D}(T,\delta)\cap P(T,\delta))$. Now (3.15) and (3.16) imply that $\Phi(F(B\cap L\cap Q))\cap(H+2\delta\mathbb{B}) = p+2\delta\Delta q$ meets $\overline{U}$ in a domain bounded by a simple closed curve contained in $\Sigma\subset bU$, and the intersection is transverse. Moreover, since Φ fixes $H\supset T$ we have $p\in\Sigma\cap T$ and T meets $\overline{U}$ in a domain bounded by a simple closed curve $\Sigma\cap T = S\cap T$ obtained as a transverse intersection of T with $\Sigma\subset bU$.

With no loss of generality, assume that $p=0$. Let M be the two-dimensional complex subspace spanned by T and $\mathbb{C}q$. Recall that $\mathbb{C}q$ meets $\overline{U}$ in a domain bounded by a simple closed curve which is a transverse intersection of $\mathbb{C}q$ with $\Sigma\subset bU$ and that T intersects $\overline{U}$ in a domain bounded by a simple closed curve which is the transverse intersection of T with $\Sigma\subset bU$. Since both $T\cap bU$ and $\mathbb{C}q\cap bU$ are contained in Σ, an open subset of bU which is smooth, it follows that near $(T\cup bU)\cup(\mathbb{C}q\cap bU)$, bU is smooth and transverse to M. Thus $\tilde U = U\cap M$ is a bounded open set in M which has smooth boundary near $(T\cap bU)\cap(\mathbb{C}q\cap bU)$. The component $\tilde U_0$ of U containing $(T\cap U)\cup(\mathbb{C}q\cap U)$ is a bounded domain in M which has smooth boundary near $(T\cap b\tilde U_0)\cup(\mathbb{C}q\cap b\tilde U_0)$ and which T and $\mathbb{C}q$ meet in domains bounded by simple closed curves, which are transverse intersections of T and $\mathbb{C}q$ with $b\tilde U_0$. Now apply Lemma 3.5 to get a sequence

$\tilde{A}_n \subset \tilde{U}_0$ of transversely embedded analytic discs whose boundaries $b\tilde{A}_n$, in the sense of that lemma, converge to $(T \cap b\tilde{U}_0) \cup (\mathbb{C}q \cap b\tilde{U}_0)$. It is now clear that by pulling back to $b\Omega \cap Q$ with $(\Phi \circ F)^{-1}$ that $A_n = (\Phi \circ F)^{-1}(\tilde{A}_n)$, $A = (\Phi \circ F)^{-1}(T \cap U_0)$, and $L \cap B \cap Q \cap b\Omega = (\Phi \circ F)^{-1}(\mathbb{C}q \cap b\tilde{U})$ satisfy (3.13) and $w_0 \in A,\ w_0 \in A_n (n \in \mathbb{N})$.

It remains to show that everything can be done in such a way that A_n $(n \in \mathbb{N})$ and A belong to $\mathcal{D}$. By transversality, for each $T \in \mathcal{T}$ the disc $T \cap F(Q \cap \Omega)$ is arbitrarily small perturbation of the intersection of $F(Q \cap \Omega)$ with z_1-axis provided that $\mathcal{T}$ is small enough. Further, our construction implies that the maps Φ are uniformly close to the identity provided that $\mathcal{H}$ and δ are small enough. Since $\mathcal{P}$ can be chosen arbitrarily small the reasoning from Discussion 3 following Lemma 3.3 applies to show that the discs $\tilde{A}_n$ belong to an arbitrarily small neighbourhood of $T \cap F(Q \cap \Omega)$ provided that $\mathcal{P}$ is small enough. It follows that $A_n,\ n \in \mathbb{N}$, and A can be chosen to belong to $\mathcal{D}$.

Theorem 1.1 is thus proved in the case of convexity. It remains to prove it in the strictly pseudoconvex case.

To this end, observe first that in the proof of Lemma 3.6 we never used the fact that the elements of $\mathcal{L}$ are complex lines. What we needed was the following:
(a) There is an open neighborhood $\mathcal{W}$ of z_0 such that all $L \in \mathcal{L}_0$ are one-dimensional complex submanifolds of $\mathcal{W}$,
(b) the initial $\Lambda(z_0) \in \mathcal{L}_0$ is tangent to $\mathcal{W} \cap b\Omega$,
(c) if $L \in \mathcal{L}_0$ is sufficiently close to $\Lambda(z_0)$ in the $\mathcal{C}^1-$sense and if L meets $\mathcal{W} \cap \Omega$, then L meets $b\Omega$ transversely in a simple closed curve bounding the domain $\Omega \cap L$, and
(d) given a neighborhood $\mathcal{V}$ of z_0, we have $L \cap \overline{\Omega} \subset \mathcal{V}$ provided that $L \in \mathcal{L}_0$ is sufficiently close to $\Lambda(z_0)$ in the $\mathcal{C}^1$-sense.

Now the following lemma is proved in exactly the same way as Lemma 3.6.

Lemma 3.9. *Let Ω be a bounded domain in $\mathbb{C}^N$ with $b\Omega$ of class $\mathcal{C}^2$. Suppose that $D_0 \subset \Omega$ is a transversely embedded analytic disc. Let $w_0 \in D_0$ and $z_0 \in bD_0$. Let $\mathcal{W}, \Lambda(z_0)$ and $\mathcal{L}_0$ be as above. Let $\mathcal{D}$ be a neighbourhood of D_0 in the space of transversely embedded analytic discs. There is an $\mathcal{L} \subset \mathcal{L}_0$, a $\mathcal{C}^1$-neighbourhood of $\Lambda(z_0)$ in $\mathcal{L}_0$, such that for each $L \in \mathcal{L}$ that meets $\Omega \cap \mathcal{W}$*

(a) *$L \cap b\Omega$ is a compact convex curve, and*
(b) *there are $A \in \mathcal{D}$ and a sequence $\{A_n\}_{n=1,2,\dots} \subset \mathcal{D}$ such that $w_0 \in A, w_0 \in A_n$ $(n \in \mathbb{N})$ and such that for each smooth 1-form α on $\mathbb{C}^2$*

$$\lim_{n\to\infty} \int_{bA_n} \alpha = \int_{bA} \alpha + \int_{L \cap b\Omega} \alpha. \tag{3.17}$$

To conclude the proof, we suppose that $b\Omega$ is strictly pseudoconvex at z_0. There exist a neighborhood $\mathcal{W}$ of z_0 and a biholomorphic map $\Psi : \mathcal{W} \to \tilde{\mathcal{W}}$, $\mathcal{W}$ a neighborhood of $0 = \Psi(z_0)$ such that Ψ is a polynomial map of degree two and such that $\Psi(\Omega \cap \mathcal{W})$ is strictly convex. Let $\tilde{L}_0$ be the complex tangent line to $b\Psi(\Omega \cap \mathcal{W})$ at 0. Let $\tilde{\mathcal{L}}_0$ be a neighborhood of $\tilde{L}_0$ in the space of complex lines, and for each $\tilde{L} \in \tilde{\mathcal{L}}_0$ let $L = \Phi^{-1}(\tilde{L} \cap \tilde{\mathcal{W}})$. If $\mathcal{W}$ and $\tilde{\mathcal{L}}_0$ are sufficiently small then $\mathcal{L}_0 = \{L : \tilde{L} \in \tilde{\mathcal{L}}_0\}$ has the properties (a)-(d). Recall that by our assumptions, $\int_{bD} f\omega = 0$ for every form $\omega \in \mathbb{C}^{1;1}[dz]$ and for each transversely embedded analytic disc $D \in \mathcal{D}$. By Lemma 3.9 it follows that there is a neighborhood $\tilde{\mathcal{L}} \subset \tilde{\mathcal{L}}_0$ of $\tilde{L}_0$ in the space of complex lines and a neighborhood Q of z_0 such that for each $L \in \mathcal{L} = \{L : \tilde{L} \in \tilde{\mathcal{L}}\}$ which meets $\Omega \cap \mathcal{W}$,

$$\int_{L\cap b\Omega} f\omega = 0 \tag{3.18}$$

for each $\omega \in \mathbb{C}^{1;1}[dz]$. Since Ψ is a polynomial map of degree two, it follows that for all $\alpha \in \mathbb{C}^{1;0}[dz]$ the form $\Psi^*\alpha$ lies in $\mathbb{C}^{1;1}[dz]$.

If $\tilde{L} \in \tilde{\mathcal{L}}$, then given a form $\alpha \in \mathbb{C}^{1,0}[dz]$ (3.18) implies that

$$\int_{\tilde{L}\cap\Psi(b\Omega\cap\mathcal{W})} (f\circ\Psi^{-1})\alpha = \int_{\Psi(L\cap b\Omega)} (f\circ\Psi^{-1})\alpha = \int_{L\cap b\Omega} (f\circ\Psi^{-1}\circ\Psi)\Psi^*\alpha = 0.$$

By [**GlS**, Th. 3.2.1] it follows that $f\circ\Psi^{-1}$ is a CR-function in a neighborhood of 0 in $\Psi(b\Omega\cap\mathcal{W})$, which implies that f is a CR-function in a neighborhood of z_0.

Theorem 1.1 is finally proved.

4. Concluding Remarks.

In this final section we note a consequence of the main theorem.

Let Ω be as in Theorem 1.1 and let $D \subset \Omega$ be a transversely embedded analytic disc. If a continuous function f on bD has a continuous extension to $\overline{D}$ which is holomorphic on D then $\int_{bD} f\omega = 0$ for all $(1,0)$-forms with linear coefficients. The holomorphic extendibility is invariant with respect to biholomorphic maps. Accordingly the following theorem holds in Stein manifolds, which we view as closed complex submanifolds of $\mathbb{C}^M$.

Theorem 4.1. *Let Ω be a relatively compact domain in a closed complex submanifold $\mathcal{M}$, $\dim \mathcal{M} \geq 2$ of $\mathbb{C}^M$ that has $\mathcal{C}^2$ boundary. Let $D_0 \subset \Omega$ be a transversely embedded analytic disc, and let Ω be strictly pseudoconvex at $z_0 \in bD_0$. Let $\mathcal{D}$ be a neighbourhood of D_0 in the space of transversely embedded analytic discs $D \subset \Omega$. Suppose that f is a continuous function on a neighbourhood of bD_0 in $b\Omega$ such that for each $D \in \mathcal{D}$ satisfying $w_0 \in D$,*

(4.1) *$f|bD$ has a continous extension to D which is holomorphic on D.*

Then f is a CR function in a neighbourhood of z_0. If (4.1) holds for all $D \in \mathcal{D}$ then there are a neighbourhood W of D_0 in $\mathcal{M}$ and a continuous function $\tilde{f}$ on $W \cap \overline{\Omega}$ such that $\tilde{f} = f$ on $W \cap b\Omega$.

Proof. Let $\dim\mathcal{M} = N$. The set $\overline{D} \subset \mathcal{M}$ is a compact, polynomially convex set in C^M and thus has a Stein neighborhood basis in $\mathbb{C}^M$. The intersections of the elements of this basis with $\mathcal{M}$ form a Stein neighborhood basis of $\overline{D}$ in $\mathcal{M}$. Thus, by the result of Docquier and Grauert and the holomorphic triviality of holomorphic vector bundles on discs as used above and by the fact that the extendibility assumptions we make are invariant under biholomorhic maps, we can assume that Ω is a domain in $\mathbb{C}^N$ and that D is the intersection of Ω with the z_1-axis. The preceding theorem now implies that f is a CR-function in a neighborhood of z_0 in $b\Omega$. This proves the first part of the theorem. As $b\Omega$ is strictly pseudoconvex at z_0, there are a neighborhood P of z_0 and a continuous function $\tilde{\tilde{f}}$ on $P \cap \overline{\Omega}$ that is holomorphic on $P \cap \Omega$ and that satisfies $\tilde{\tilde{f}} = f$ on $P \cap b\Omega$. Our assumptions imply that there is $\delta > 0$ such that if $\mathcal{T}$ denotes the set of all complex lines parallel to the z_1-axis and at distance not exceeding δ from it, then each $T \in \mathcal{T}$ meets $P \cap b\Omega$ transversely (and thus also meets $P \cap \Omega$), T meets $b\Omega$ in a simple closed curve bounding $T \cap \Omega$, and $f|(T \cap b\Omega)$ has a continuous extension f_T to $T \cap \overline{\Omega}$ that is holomorphic in $T \cap \Omega$. Since $T \cap b\Omega \cap P$ contains an arc, f_T coincides with $\tilde{\tilde{f}}|T$ near $b\Omega \cap P$. Thus, by Hartogs's lemma, the function $\tilde{f}$ defined by $\tilde{f}|(T \cap \overline{\Omega}) = f_T$ has all the required properties on $\mathcal{W} = \cup\{T : T \in \mathcal{T}\}$.

This completes the proof.

References

[Al1] H. Alexander, *Polynomial approximation and hulls in sets of finite linear measure in C^n*, American J. Math., **93** (1971), 65-74.

[Al2] ———, *The polynomial hull of a rectifiable curve in C^n*, American J. Math, **110** (1988), 629-640.

[Gl1] J. Globevnik, *Testing analyticity on rotation invariant families of curves*, Trans. Amer. Math. Soc., **306** (1988), 401-410.

[Gl2] ———, *A boundary Morera theorem in C^2*, J. Geom. Anal., **3** (1993), 269-277.

[GlS] J. Globevnik and E.L. Stout, *Boundary Morera theorems for holomorphic functions of several complex variables*, Duke Math. J., **64** (1991), 571-615.

[GR] R.C. Gunning and H. Rossi, *Analytic Functions of Several Complex Variables*, Prentice-Hall, Englewood Cliffs, 1965.

[Po] Ch. Pommerenke, *Boundary Behaviour of Conformal Maps*, Springer-Verlag, Berlin-Heidelberg-New York, 1992.

[Tu1] A.E. Tumanov, *Extension of CR-functions into a wedge from a manifold of finite type* (Russian), Mat. Sb. (N.S.), **136(178)** (1988), 128-139.

[Tu2] ———, *Thin discs and a Morera theorem for CR functions*, Math. Z., (1997), 327-334.

Received April 20, 1998. Research of the first author was supported in part by a grant from the Ministry of Science and Technology of the Republic of Slovenia. Research of the second author was supported in part by grant 9322326 from the National Science Foundation.

University of Ljubljana
Ljubljana
Slovenia
E-mail address: josip.globevnik@fmf.uni-lj.si

University of Washington
Seattle, WA 98195
E-mail address: stout@math.washington.edu

PACIFIC JOURNAL OF MATHEMATICS
Vol. 192, No. 1, 2000

THE CLASSIFICATION OF COHOMOLOGY ENDOMORPHISMS OF CERTAIN FLAG MANIFOLDS

Duan Haibao and Zhao Xuan

This note describe a method for dertermining cohomology endomorphisms of a space. Examples are taken for some flag manifolds.

1.

Let $U(n)$ be the unitary group of rank n, $SO(m)$ the special orthogonal group of rank m, and $Sp(n)$, the symplectic group of rank n. Fix, once and for all, a maximal torus $T^n \subset U(n)$, $T^{[\frac{m}{2}]} \subset SO(m)$, $T^n \subset Sp(n)$ in each of the groups.

For a topological space X the problem of finding all integral cohomology endomorphisms $H^*(X;Z) \to H^*(X;Z)$ is a step toward classifying all homotopy classes of self maps $X \to X$. In [**3**] M. Hoffman solved this problem for the flag manifold $F(n) = U(n)/T^n$. In this note we settle the problem for the spaces $D(m) = SO(m)/T^{[\frac{m}{2}]}$ and $S(n) = Sp(n)/T^n$.

It is worth to point out that S. Papadima determined all cohomology automorphisms of G/T with coefficients in R or Q, where G is a compact connected Lie Group and T its maximal torus [**5**].

2.

We describe the integral cohomologies (in terms of generators-relations) of these spaces. Let $e_i(z_1, \cdots, z_n)$ be the i^{th} elementary symmetric function in the variables $z_1, \cdots, z_n$. The following results is well known from A. Borel [**1**].

Lemma 1.

$$H^*(F(n);Z) = Z[t_1, \dots, t_n]/e_i(t_1, \dots, t_n),\ 1 \le i \le n,\ |t_i| = 2;$$

$$H^*(S(n);Z) = Z[y_1, \dots, y_n]/e_i(y_1^2, \dots, y_n^2),\ 1 \le i \le n,\ |y_i| = 2,$$

and there exist integral classes $x_1, \dots, x_n \in H^2(D(m);Z)$, $n = [\frac{m}{2}]$, so that

$$H^*(D(2n);Q) = Q[x_1, \dots, x_n]/e_i(x_1^2, \dots, x_n^2),\quad 1 \le i \le n-1;\ e_n(x_1, \dots, x_n);$$

$$H^*(D(2n+1);Q) = Q[x_1, \dots, x_n]/e_i(x_1^2, \dots, x_n^2),\ 1 \le i \le n,$$

where $|u|$ stands for the grade of the enclosed u and Q, the field of rationals.

The integral cohomology for the space $D(m)$ is not available in [**1**] due to the presence of 2-torsion in $H^*(SO(m);Z)$. However, with slightly more geometric investigation, the algebra $H^*(D(m);Z)$ can be shown to be as follows.

Let $G(2n,2)$ be the Grassmannian of oriented 2-planes through the origin in the Euclidean $2n$-space R^{2n}. The canonical 2-plane bundle γ over $G(2n,2)$ will have the preferred orientation. Let $x \in H^2(G(2n,2);Z)$ be the Euler classes for the oriented bundle γ.

The space $D(2n)$ can be identified with the subspace of $G(2n,2)\times\cdots\times G(2n,2)$ (n-copies) consisting of n mutually orthogonal oriented 2-planes with their induced orientation on R^{2n} agrees with the standard one. Evidently we have n projections $\pi_i : D(2n) \to G(2n,2)$, $1 \le i \le n$, given by taking the i^{th} plane. Set $x_i = \pi_i^*(x) \in H^2(D(2n);Z)$.

Lemma 2. *The integral classes $e_i(x_1,\dots,x_n) \in H^*(D(2n);Z)$, $1 \le i \le n$, are all divisible by 2. Further the algebra $H^*(D(2n);Z)$ is generated multiplicatively by $x_1,\dots,x_{n-1}$ and $v = \frac{1}{2}e_1(x_1,\dots,x_n)$ subject to the relations*

$$\frac{1}{4}e_i(x_1^2,\dots,x_n^2) = 0,\ 1 \le i \le n-1, \quad \textit{and} \quad \frac{1}{2}e_n(x_1,\dots,x_n) = 0,$$

where the 4-divisibility of $e_i(x_1^2,\dots,x_n^2)$ follows from the 2-divisibility of $e_i(x_1,\dots,x_n)$.

From the Wang sequence of the standard fibration

$$D(2n) \subset D(2n+1) \xrightarrow{p} S^{2n}$$

we find classes $x_1',\dots,x_n' \in H^2(D(2n+1);Z)$ so that $x_i' \mid D(2n) = x_i \in H^2(D(2n);Z)$, i.e., the bundle p has Leray-Hirsch property [**6**, p. 365]. Moreover it can be shown that $p^*(\iota) = \pm\frac{1}{2}x_1'\cdots x_n'$, where $\iota \in H^{2n}(S^{2n};Z) = Z$ is a generator and p^*, the induced homomorphism. Therefore, Lemma 2 implies:

Lemma 3. *The integral classes $e_i(x_1',\dots,x_n') \in H^*(D(2n+1);Z)$, $1 \le i \le n$, are all divisible by 2. Further the algebra $H^*(D(2n+1);Z)$ is generated multiplicatively by $x_1',\dots,x_{n-1}'$ and $v' = \frac{1}{2}e_1(x_1',\dots,x_n')$ subject to the relations*

$$\frac{1}{4}e_i(x_1'^2,\dots,x_n'^2) = 0, \quad 1 \le i \le n.$$

3.

Let X denote one of the spaces $F(n)$, $D(m)$, or $S(n)$. From the descriptions for $H^*(X;Z)$ two types of endomorphisms become apparent:

Example 1. For an integer $k \in Z$, the Adams map $l_k : H^*(X;Z) \to H^*(X;Z)$ is defined by $l_k(u) = ku$ for all $u \in H^2(X;Z)$.

Example 2. Let Σ_n be the permutation group on $\{1, 2, \dots, n\}$, and let S_n be the full permutation group on $\{\pm 1, \pm 2, \dots, \pm n\}$. We set

$$\Phi(X) = \begin{cases} \Sigma_n & \text{if } \; X = F(n) \quad \text{and} \\ S_n & \text{if } \; X = D(2n),\; D(2n+1) \text{ or } S(n). \end{cases}$$

For a $\alpha \in \Phi(X)$ the standard action of α on $H^2(X; Z)$ extends uniquely to an algebra map $H^*(X; Z) \to H^*(X; Z)$, for which we denote still by α.

For an $k \in Z$ and a $\alpha \in \Phi(X)$ write h_k^α for the composition $\alpha \circ l_k$. In [**3**] M. Hoffman proved that any endomorphism of $H^*(F(n); Z)$ has the form h_k^α for some $k \in Z$ and $\alpha \in \Phi(F(n))$. We extends his result in:

Theorem 1. *For $X = D(m)$, $m > 5$, $m \neq 8$ or $S(n)$, $n > 2$, any endomorphism of $H^*(X; Z)$ has the form h_k^α for some $k \in Z$ and $\alpha \in \Phi(X)$.*

Exceptions do occur for the spaces $D(4)$, $D(5)$, $D(8)$ and $S(2)$. Indeed, as only small values of n are involved, a direct computation based on Lemma 1-3 yields the following. Let h be an endomorphism of $H^*(X; Z)$ with $X = D(4)$, $D(5)$, $D(8)$ or $S(2)$. With respect to the Q-basis of $H^2(X; Z)$ given in Lemma 1 the action of h on $H^2(X; Z)$ has the representations:

1) h_k^α, or $\begin{pmatrix} -1 & 1 \\ 1 & 1 \end{pmatrix} \circ h_k^\alpha$, with $k \in Z$, $\alpha \in S_2$ for $X = S(2)$;

2) $\frac{1}{2}\begin{pmatrix} a & \epsilon a \\ b & \epsilon b \end{pmatrix}$, or $\begin{pmatrix} a & \epsilon b \\ b & \epsilon a \end{pmatrix}$, with $a, b \in Z$, $\epsilon = \pm 1$ for $X = D(4)$;

3) h_k^α, or $\frac{1}{2}\begin{pmatrix} -1 & 1 \\ 1 & 1 \end{pmatrix} \circ h_k^\alpha$, with $k \in Z$, $\alpha \in S_2$ for $X = D(5)$ and

4) h_k^α, or $\frac{1}{2}\begin{pmatrix} -1 & 1 & 1 & 1 \\ 1 & -1 & 1 & 1 \\ 1 & 1 & -1 & 1 \\ 1 & 1 & 1 & -1 \end{pmatrix} \circ h_k^\alpha$, with $k \in Z$, $\alpha \in S_4$ for $X = D(8)$.

4.

In studying endomorphisms of an algebra, a technique of handling relations appears to be crucial. The proof of Theorem 1 serves the purpose of introducing the so called "quantification" method. By this we mean polynomial relations might be quantified, hence simplified, by evaluating them at appropriate quantities. More precisely, the proof of Theorem 1 will be based on the following numerical results.

For a complex number $y \in C$ denote respectively by $\langle y \rangle$, and $[y]$ for the sets of sequences

$$\left\{(\epsilon_1, \epsilon_2 y, \dots, \epsilon_n y^{n-1})_\alpha \mid \alpha \in \Sigma_n,\; \epsilon_i = \pm 1\right\};$$
$$\left\{(0, \epsilon_1, \epsilon_2 y, \dots, \epsilon_{n-1} y^{n-2})_\alpha \mid \alpha \in \Sigma_n,\; \epsilon_i = \pm 1\right\},$$

where, for instance, $(z_1, \dots, z_n)_\alpha = (z_{\alpha(1)}, \dots, z_{\alpha(n)})$.

Theorem 2. *Let* $a_1, \dots, a_n$; r *be some reals, and let* $\xi = e^{\frac{\pi i}{n}}$, $\eta = e^{\frac{\pi i}{n-1}}$.

(1) *If* $n > 2$ *and if the equality* $(a_1x_1 + \cdots + a_nx_n)^{2n} = r$ *holds for all* $(x_1, \dots, x_n) \in \langle\xi\rangle$, *then at most one of* $a_1, \dots, a_n$ *is non-zero.*

(2) *If* $n > 4$ *and if the equality* $(a_1x_1+\cdots+a_nx_n)^{2n} = r(a_1x_1+\cdots+a_nx_n)^2$ *holds for all* $(x_1, \dots, x_n) \in [\eta]$, *then at most one of* $a_1, \dots, a_n$ *is non-zero.*

5.

The proof of Theorem 2 will be postponed until the next section and at this moment, we show how it leads to a proof of Theorem 1. Let h be an endomorphism of $H^*(X; Z)$. Define a matrix $A = (a_{ij})_{n\times n}$ by

$$h(y_i) = \Sigma_j a_{ij} y_j,\ a_{ij} \in Z, \quad \text{for}\ \ X = S(n);$$

(resp. $h(x_i) = \Sigma_j a_{ij}x_j$, $a_{ij} \in Q$, for $X = D(m)$ with $m = 2n$ or $2n + 1$).

Lemma 4. *If* $n > 2$ (*resp.* $m > 5$ *and* $m \neq 8$), *at most one entry in each row of* A *is nonzero.*

Proof. Consider first the case $X = S(n)$, $n > 2$ (resp. $D(2n + 1)$, $n > 2$; for rationally the cohomology of $S(n)$ is isomorphic to that of $D(2n+1)$ by Lemma 1). The obvious formula in $Z[y_1, \dots, y_n]$

$$\begin{aligned} y_i^{2n} &= y_i^{2(n-1)} e_1(y_1^2, \dots, y_n^2) - y_i^{2(n-2)} e_2(y_1^2, \dots, y_n^2) \\ &\quad + \cdots + (-1)^{n-1} e_n(y_1^2, \dots, y_n^2) \end{aligned}$$

implies, in $H^*(S(n); Z)$, that $y_i^{2n} = 0$, and consequently $h(y_i)^{2n} = 0$. Thus in $Z[y_1, \dots, y_n]$ we have

$$h(y_i)^{2n} = (a_{i1}y_1 + \cdots + a_{in}y_n)^{2n} = g_1e_1(y_1^2, \dots, y_n^2) + \cdots + g_ne_n(y_1^2, \dots, y_n^2),$$

for some $g_i \in Z[y_1, \dots, y_n]$, $1 \le i \le n$, with $|g_i| = 4(n-i)$. Note that $g_n = r \in Z$ for degree reason. Since the systems

$$e_i(y_1^2, \dots, y_n^2) = 0,\ 1 \le i \le n-1; \quad e_n(y_1^2, \dots, y_n^2) = 1$$

is satisfied by all $(y_1, \dots, y_n) \in \langle\xi\rangle$, $\xi = e^{\frac{\pi i}{n}}$, we get

$$(a_{i1}y_1 + \cdots + a_{in}y_n)^{2n} = r \quad \text{for all}\ \ (y_1, \dots, y_n) \in \langle\xi\rangle.$$

So $h(y_i) = ay_j$ for some $a \in Z$, $1 \le j \le n$, by 1) of Theorem 2.

Consider next $X = D(2n)$, $n > 4$. Applying h to the standard equality

$$\begin{aligned} x_i^{2n} &= x_i^{2(n-1)} e_1(x_1^2, \dots, x_n^2) - x_i^{2(n-2)} e_2(x_1^2, \dots, x_n^2) \\ &\quad + \cdots + (-1)^{n-1} e_n(x_1, \dots, x_n)^2 \end{aligned}$$

in $Q[x_1, \dots, x_n]$ and using (since h preserves the ideal) the facts

$$h(e_i(x_1^2, \dots, x_n^2)) = \Sigma_{j \le i} g_{ij} e_j(x_1^2, \dots, x_n^2) + l_i e_n(x_1, \dots, x_n);$$
$$h(e_n(x_1, \dots, x_n)) = \Sigma_{2i<n} k_i e_i(x_1^2, \dots, x_n^2) + l e_n(x_1, \dots, x_n)$$

we find

$$h(x_i)^{2n} = f \in \{\text{the ideal generated by } \ e_i(x_1^2, \dots, x_n^2),\ 1 \le i \le n-2; \\ e_n(x_1, \dots, x_n)\} + g_{n-1,n-1} h(x_i)^2 e_{n-1}(x_1^2, \dots, x_n^2),$$

where $g_{n-1,n-1} = r \in Q$ for degree reasons. Since the systems

$$e_n(x_1, \dots, x_n) = e_i(x_1^2, \dots, x_n^2) = 0,\ 1 \le i \le n-2;\ e_{n-1}(x_1^2, \dots, x_n^2) = 1$$

is satisfied by all $(x_1, \dots, x_n) \in [\eta]$ we get

$$h(x_i)^{2n} = r\, h(x_i)^2 \ \text{ for all } \ (x_1, \dots, x_n) \in [\eta].$$

Thus $h(x_i) = a x_j$ for some $a \in Q$, $1 \le j \le n$, by 2) of Theorem 2.

The case $X = D(6)$ requires additional treatment as assertion 2) of Theorem 2 does not apply to it. Assume, with respect to the Q-basis $x_1, x_2, x_3 \in H^2(D(6); Z)$, that h has the matrix representation

$$A = \begin{pmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{pmatrix},\ a_i, b_j, c_k \in Q.$$

Since h preserves the ideal, we have

$$h(x_1^2 + x_2^2 + x_3^2) = l(x_1^2 + x_2^2 + x_3^2); \qquad \text{and} \tag{5.1}$$

$$h(x_1 x_2 x_3) = g(x_1^2 + x_2^2 + x_3^2) + k x_1 x_2 x_3 \tag{5.2}$$

for some $l, k \in Q$, $g \in Q[x_1, x_2, x_3]$. From (5.1) we get $\Sigma a_i^2 = \Sigma b_i^2 = \Sigma c_i^2 = l$. If $l = 0$ we are done, so we may assume that $l \ne 0$ (thus A has no zero row). Putting $(x_1, x_2, x_3) = (0, it, t)$, $i = \sqrt{-1}$, $t \in R$, in (5.2) gives

$$(a_2 i + a_3)(b_2 i + b_3)(c_2 i + c_3) t^3 = 0 \quad \text{for all } \ t \in R.$$

So one of the three complex numbers $(a_2 i + a_3)$, $(b_2 i + b_3)$, $(c_2 i + c_3)$ must be zero. Assume, without losing the generality, that $a_2 = a_3 = 0$ (hence $a_1 \ne 0$). Then (5.2) becomes

$$a_1 x_1 (\Sigma b_i x_i)(\Sigma c_i x_i) = g(x_1^2 + x_2^2 + x_3^2) + k x_1 x_2 x_3.$$

Setting $(x_1, x_2, x_3) = (t, it, 0)$ we get $a_1(b_1 + b_2 i)(c_1 + c_2 i) t^3 = 0$ for all $t \in R$. So we may assume further that $b_1 = b_2 = 0$, $b_3 \ne 0$. Finally from

$$a_1 b_3 x_1 x_3 (\Sigma c_i x_i) = g(x_1^2 + x_2^2 + x_3^2) + k x_1 x_2 x_3$$

we get $c_1 = c_3 = 0$. This completes the proof of Lemma 4. $\square$

Proof of Theorem 1.

Case 1. $X = S(n)$, $n \geq 3$. Since h must preserve the ideal we have, in particular, that $h(y_1^2 + \cdots + y_n^2) = l(y_1^2 + \cdots + y_n^2)$ for some $l \in Z$. Now $h(y_1^2 + \cdots + y_n^2) = \Sigma_j(\Sigma_i a_{ij}^2)y_j^2$ (by Lemma 4) tells that

(5.3) all column sum of $(a_{ij}^2)_{n\times n}$ are equal (to l).

If $l = 0$, then $A = 0_{n\times n}$, we are clearly done. If $l \neq 0$ then (5.3) indicates that A has no zero column. A combination of Lemma 4 and (5.3) now gives that $A = kP$, where $k \in Z$ with $k^2 = l$, and where P is the matrix representation of some $\alpha \in S_n$. This completes the proof for $X = S(n)$.

Case 2. $X = D(m)$, $m \geq 6$ and $m \neq 8$. A discussion similar to Case 1 shows that $h = h_k^\alpha$ for some $k \in Q$, $\alpha \in S_n$. Further, as h must send integral classes to integral ones we have $k \in Z$ by Lemma 2 and 3. This finishes the proof of Theorem 1. □

6. Proof of Theorem 2.

Assume throughout this section that $a_1, \ldots, a_n; r$ are some real numbers satisfying

$$\begin{aligned}&(a_1x_1 + \cdots + a_nx_n)^{2n}\\ &= r \quad \text{for all } \ (x_1, \ldots, x_n) \in \langle\xi\rangle\\ &\Big(\text{resp. } \ (a_1x_1 + \cdots + a_nx_n)^{2n}\\ &= r(a_1x_1 + \cdots + a_nx_n)^2 \ \text{ for all } \ (x_1, \ldots, x_n) \in [\eta]\Big)\end{aligned}$$

where $\xi = e^{\frac{\pi}{n}}$(resp. $\eta = e^{\frac{\pi}{n-1}}$). Clearly we can also assume that $a_i \geq 0$ for all $1 \leq i \leq n$ (by replacing a_i with $sign(a_i)a_i$ when $a_i < 0$). If $r = 0$, Theorem 2 is obviously true. So we may adopt the convention that $r \neq 0$.

Denote by C^+ the upper half complex plane $\{a + bi \mid a, b \in R, \ b \geq 0\}$, and write

$$\begin{aligned}&H_1 = \{\omega_\alpha = a_{\alpha(1)} + a_{\alpha(2)}\xi + \cdots + a_{\alpha(n)}\xi^{n-1} \mid \alpha \in \Sigma_n\};\\ &\Big(\text{resp. } \ H_2 = \big\{\omega'_\alpha = a_{\alpha(1)} + a_{\alpha(2)}\eta\\ &\qquad + \cdots + a_{\alpha(n-1)}\eta^{n-2} + a_{\alpha(n)}\cdot 0 \mid \alpha \in \Sigma_n\big\}\Big).\end{aligned}$$

Since $\xi^k \in C^+$ for all $0 \leq k \leq n-1$ (resp. $\eta^l \in C^+$ for $0 \leq l \leq n-2$), and since $a_i \geq 0$, the set H_1 (resp. H_2) of complex numbers can be considered as a subset of C^+. We observe first of all that:

Lemma 5. *If the sequence $a_1, \ldots, a_n$, $n \geq 3$, contains more than one non-zero entry then $\#H_i \leq n-1$, $i = 1, 2$.*

Proof. Let φ be the polar angle function (with respect to the oriented real axis) on nonzero complexes. Take a base point $\omega_0 \in H_1$(resp. $\omega_0' \in H_2$) with

$$\varphi(\omega_0) = \min\{\varphi(\omega_\alpha) \mid \omega_\alpha \in H_1 \setminus 0\}$$
$$(\text{resp. } \varphi(\omega_0') = \min\{\varphi(\omega_\alpha') \mid \omega_\alpha' \in H_2 \setminus 0\}).$$

Clearly if the sequence $a_1, \cdots, a_n$ contains more than one nonzero entry then we have $0 \notin H_i$, and for all $\omega_\alpha \in H_1$ (resp. $\omega_\alpha' \in H_2$)

$$0 < \varphi(\omega_0) \le \varphi(\omega_\alpha) < \frac{n-1}{n}\pi$$
$$\left(\text{resp. } 0 \le \varphi(\omega_0') \le \varphi(\omega_\alpha') \le \frac{n-2}{n-1}\pi\right).$$

This indicates that, since $\omega_\alpha^{2n} = r$ (resp. $(\omega_\alpha')^{2n} = r(\omega_\alpha')^2$), for each $\alpha \in \Sigma_n$ there exists an integer $0 \le k_\alpha \le n-2$ (resp. $0 \le l_\alpha \le n-2$) so that $\omega_\alpha = \xi^{k_\alpha}\omega_0$ (resp. $\omega_\alpha' = \eta^{l_\alpha}\omega_0'$). This verifies $\#H_1$ (resp. $\#H_2$)$\le n-1$. □

On the other hand a common lower bound for $\#H_i$ can be obtained as follows. Since both sets H_i are invariant under the given action of Σ_n we can assume, without loss the generality, that $a_1 \ge a_2 \ge \cdots \ge a_n$. Thus a partition $r_1, \cdots, r_k$ for n (i.e., $r_i \ge 1$, $\Sigma r_i = n$) can be found so that, if $s_h = r_1 + \cdots + r_h$ then

$$a_{s_h+1} = \cdots = a_{s_{h+1}}, \quad a_{s_h+1} > a_{s_h} \quad \text{for all } h = 1, \dots, k.$$

For a pair of integers $1 \le i < j \le n$ let $(i,j) \in \Sigma_n$ be the transposition of i and j. It is easy to see that the cardinality of the set

$$\overline{\Sigma}_n = \{\alpha_0,\ (i,j) \mid i \in [s_h+1, s_{h+1}], j \in [s_{h'}+1, s_{h'+1}],\ 1 \le h < h' \le k\}$$

is $1 + \Sigma_{1 \le i < j \le k} r_i r_j$. However for a $\alpha = (i,j)$ we have

$$\omega_\alpha = \omega_{id} + (a_i - a_j)(\xi^{j-1} - \xi^{i-1}) \in H_1 \tag{6.1}$$
$$\left(\text{resp. } \omega_\alpha' = \omega_{id}' + \begin{cases} (a_i - a_j)(\eta^{j-1} - \eta^{i-1}) & \text{if } j < n \\ (a_n - a_i)\eta^{i-1} & \text{if } j = n \end{cases} \in H_2\right),$$

where $id \in S_n$ is the identity. We use this to show:

Lemma 6. *Assume as above. Then $\#H_i \ge 1 + \Sigma_{1 \le i < j \le k} r_i r_j$.*

Proof. It suffices to show that if $\alpha, \beta \in \overline{\Sigma}_n$ with $\alpha \ne \beta$, then $\omega_\alpha \ne \omega_\beta$ (resp. $\omega_\alpha' \ne \omega_\beta'$). We verify the first assertion for instance. If one of α, β is the identity, it is clearly true. Assume $\alpha = (i,j), \beta = (i\prime, j\prime) \in \overline{\Sigma}_n$ are such that $\omega_\alpha = \omega_\beta$. Then $(a_i - a_j)(\xi^{j-1} - \xi^{i-1}) = (a_{i'} - a_{j'})(\xi^{j'-1} - \xi^{i'-1})$ by (6.1).

Comparing the real and imaginary parts gives

(6.2)

$$(a_i - a_j)\sin\frac{j-i}{2n}\pi\sin\frac{i+j-2}{2n}\pi = (a_{i'} - a_{j'})\sin\frac{j'-i'}{2n}\pi\sin\frac{i'+j'-2}{2n}\pi;$$

(6.3)

$$(a_i - a_j)\sin\frac{j-i}{2n}\pi\cos\frac{i+j-2}{2n}\pi = (a_{i'} - a_{j'})\sin\frac{j'-i'}{2n}\pi\cos\frac{i'+j'-2}{2n}\pi.$$

Taking the quotient of (6.2) by (6.3) yields $tg\frac{i+j-2}{2n}\pi = tg\frac{i'+j'-2}{2n}\pi$. Since $1 \le i < j \le n, 1 \le i' < j' \le n$, the monodromy of tg on $[0,\pi]$ tells

$$i + j = i\prime + j\prime. \tag{6.4}$$

Now (6.2) gives

$$(a_i - a_j)\sin\frac{j-i}{2n}\pi = (a_{i'} - a_{j'})\sin\frac{j'-i'}{2n}\pi. \tag{6.5}$$

We may assume that $i \le i'$. (6.4) then implies $i \le i' < j' \le j$. So we have $(a_i - a_j) \ge (a_{i'} - a_{j'})(> 0)$ and that

$$\sin\frac{j-i}{2n} \ge \sin\frac{j'-i'}{2n}, \tag{6.6}$$

while the equality holds if and only if $j - i = \ j' - i'$
(by the monotonicity of sin on $[0,\frac{\pi}{2}]$). Now (6.5) says that

$$j - i = \ j' - i'. \tag{6.7}$$

Summarizing $\omega_\alpha = \omega_\beta$ (resp. $\omega'_\alpha = \omega'_\beta$), $\alpha,\beta \in \overline{\Sigma}_n$, will imply $\alpha = \beta$ by (6.4) and (6.7). This completes the proof of Lemma 6. □

Since

$$\begin{aligned}
&1 + \Sigma_{1\le i<j\le k} r_i r_j \\
&= 1 + (n - r_k)r_k + \Sigma_{1\le i<j\le k-1} r_i r_j \text{ (since } \ r_1 + \cdots + r_k = n) \\
&= n + (n - r_k - 1)(r_k - 1) + \Sigma_{1\le i<j\le k-1} r_i r_j \\
&\begin{cases} > n & \text{if } \ k \ge 3 \\ = n + (r_1 - 1)(r_2 - 1) & \text{if } \ k = 2 \\ = 1 & \text{if } \ k = 1 \end{cases}
\end{aligned}$$

a combination of Lemma 5 and 6 gives:

Lemma 7. *If the sequence $a_1, \dots, a_n$, $n \ge 3$, contains more than one non-zero entry, then $a_1 = \cdots = a_n = b$ for some $b \ne 0$.*

We complete the proof of Theorem 2 by showing:

Lemma 8. *If $a_1 = \cdots = a_n = b$ for some $b \ne 0$, then $n \le 2$ (resp. $n \le 4$).*

Proof. Here we have $H_1 = \{\omega_0 = \frac{2b}{1-\xi}\}$ (resp. $H_2 = \{\omega_0' = \frac{2b}{1-\eta}\}$). We put

$$J_1 = \{\omega_k = \omega_0 - 2b\xi^{k-1} \mid k = 1, \dots, n\}$$
$$(\text{resp. } \; J_2 = \{\omega_k' = \omega_0' - 2b\eta^{k-1} \mid k = 1, \dots, n-1\}).$$

It suffices to show that

$$\#J_1 = n \leq 2 \text{ (resp. } \; \#J_2 = n - 1 \leq 3). \tag{6.8}$$

Let S_1 be the circle in the complex plane centered at ω_0 (resp. ω_0') with radius $2|b|$. Then $J_i \subset S_1$, $i = 1, 2$. On the other hand, if we let S_2 be the circle centered at 0 with radius $\sqrt{|r|^{\frac{1}{2n}}}$ $\left(\text{resp. } \sqrt{|r|^{\frac{1}{2n-2}}}\right)$, then we have $J_1 \subset S_2$ since $(\omega_k)^{2n} = r$ for all $1 \leq k \leq n$ (resp. $J_2 \subset S_2 \cup 0$ since $(\omega_k')^{2n} = r(\omega_k')^2$ for all $1 \leq k \leq n-1$). (6.8) is now verified by

$$J_1 \subseteq S_1 \cap S_2 \text{ (resp. } \; J_2 \subseteq S_1 \cap (S_2 \cup 0)).$$

□

Remark 1. The following result, on which the main results of [**2**] and [**3**] are based, was first proved in [**4**] and a different proof was given in [**2**].

"If $u \in H^2(F(n); Z)$ and $u^n = 0$, then $u = at_i$ for some $a \in Z$, $1 \leq i \leq n$." Our proof of Lemma 4 indicates that its analogue holds also for the spaces $S(n)$ and $D(2n+1)$.

Remark 2. One may find the following corollary of Lemma 7 useful in verifying the assertion made in Section 3, exceptional Case 4.

"Let $a_1, \cdots, a_4$; r be some reals, and let $\eta = e^{\frac{\pi i}{3}}$. If the equality $(a_1x_1 + \cdots + a_4x_4)^8 = r(a_1x_1 + \cdots + a_4x_4)^2$ holds for all $(x_1, \dots, x_4) \in [\eta]$, then we have either 1) at most one of $a_1, \dots, a_n$ is non-zero or, 2) $|a_1| = \cdots = |a_4|$." Instead, a proof based on Lemma 2 will cause tedious computation.

Acknowledgement. The authors feel grateful to Boju Jiang. The present proof of Lemma 8, more intuitive and shorter than the previous one, is due to him.

References

[1] A. Borel, *Sur la cohomologie des espaces fibres principaux et des espaces homogenes de groupes de Lie compacts*, Ann. of Math., **57** (1953), 115-207.

[2] J. Ewing and A. Liulevicius, *Homotopy rigidity of linear actions on homogeneous spaces*, J. Pure and Applied Algebra, **18** (1990), 259-267.

[3] M. Hoffman, *On fixed point free maps of the complex flag manifold*, Indiana U. Math. J., **33** (1984), 249-255.

[4] J.D. Monk, *The geometry of flag manifolds*, Proc. London Math. Soc., **9**(**3**) (1959), 253-286.

[5] S. Papadima, *Rigidity properties of compact Lie group module maximal tori*, Math. Ann., **275** (1986) 637-652.

[6] R.M. Switzer, *Algebraic topology-homotopy and homology*, Springer-Verlag Berlin-Heidelberg, 1975.

Received October 22, 1997 and revised October 20, 1998. Project 19771001 supported by NSFC.

PEKING UNIVERSITY
BEIJING 100871
P. R. CHINA
E-mail address: dhb@sxx0.math.pku.edu.cn

PEKING UNIVERSITY
BEIJING 100871
P. R. CHINA

PACIFIC JOURNAL OF MATHEMATICS
Vol. 192, No. 1, 2000

THE STATISTICS OF THE CONTINUED FRACTION DIGIT SUM

Doug Hensley

The statistics of the digits of a continued fraction, also known as partial quotients, have been studied at least since the time of Gauss. The usual measure m on the open interval $(0,1)$ gives a probability space $\mathcal{U}$. Let a_k, $k \geq 1$ be integer-valued random variables which take $\alpha \in (0,1)$ to the k^{th} partial quotient or digit in the continued fraction expansion $\alpha = 1/(a_1 + 1/(a_2 + \cdots))$. Let $S_r = S_r(\alpha) = \sum_{k=1}^{r} a_k$. It is well known that although there is an average value for $\log a_k$, each a_k, let alone each S_r, has infinite expected value or first moment.

The main result of this work is that there exists a stable probability density function ϕ on $\mathbf{R}$ so that

$$\lim_{r\to\infty} \sup_{z\in\mathbf{R}} \left| m(\{x \in (0,1) : S_r(x) \leq z\}) - \int_{-\infty}^{z\log 2/r+\gamma-\log(r/\log 2)} \phi(x)\,dx \right|$$
$$= 0.$$

Explicit error bounds, and some interesting properties of ϕ, are given. This ϕ occupies the boundary zone between distributions with support R and those with support of the form $[a,\infty)$, and was enough of an anomaly that Lévy classified it as only ‘quasi-stable’. Such distributions arise also in connection with the behavior of the sum of independent, identically distributed random variables of infinite expected value, and ϕ in particular is associated with the sum $X_1 + X_2 + \cdots + X_n$ where X_j is the reciprocal of a random variable uniformly distributed on $[0,1]$. Lévy considered this sum, and conjectured that $X_1 + X_2 + \cdots + X_n < n(\log n + \log c - \log\log\log n)$ infinitely often if and only if $c < 1$. From the results obtained here, the triple logarithm should be a double, and the cutoff is at $c = e^{-\gamma}$. We give a quick proof that if $c < e^{-\gamma}$ then almost surely $X_1 + X_2 + \cdots + X_n < n(\log n + \log c - \log\log n)$occurs but finitely often.

A stable probability distribution function F is infinitely divisible and has the additional property that for all $a_1 > 0, a_2 > 0$ and for all real b_1, b_2 there exist $a > 0$ and b so that $F(a_1x + b_1) * F(a_2x + b_2) = F(ax + b)$. The history of our understanding of such F is convoluted. In the 1950’s, Lapin [**10**] claimed to have proved that the convolution of unimodal functions is unimodal, but there were counterexamples, as noted by Chung in the course of translating the book 'Limit distributions for sums of independent random variables' [**4**]. In 1971, Wolfe published a paper [**14**, **1**] on the unimodality of distributions in class L, a class which includes the stable distributions. His result was that the class of 'one-sided' distributions in L is unimodal. Later, Yamazato [**15**, **1**] showed that from Wolfe's theorem it follows that all distributions in class L are unimodal. Stable F, then, have unimodal density when the dust settles. Their characteristic functions (Fourier transforms) have the form

$$f(t) = \exp\left(i\gamma t - c|t|^{\alpha}\left\{1 + i\beta(t/|t|)\omega(t,\alpha)\right\}\right)$$

where in general γ is an arbitrary real number, c is an arbitrary positive number, $0 < \alpha \le 2$, $-1 \le \beta \le 1$, and $\omega(t,\alpha) = \tan(\pi\alpha/2)$ if $\alpha \ne 1$, while if $\alpha = 1$ (the case at hand!), $\omega(t,\alpha) = (2/\pi)\log|t|$ [**4**]. For the particular stable density function ϕ, we have moreover:

(1) ϕ is positive on $\mathbf{R}$.

(2) $\phi(x) \approx \dfrac{1}{\sqrt{2\pi}} \exp\left(|x| - 1 - e^{|x|-1}\right)$ as $x \to -\infty$.

(3) $\phi(x) = x^{-2} + (2\gamma - 3 + 2\log x)x^{-3} + O(x^{-4}\log^2 x)$ as $x \to \infty$.

(4) ϕ extends to an entire function as the Laplace transform of $t^{-t}\sin \pi t$.

(6) $\hat{\phi}(t) = \exp(-(\pi/2)|t| + it\log|t|)$.

A computer-generated plot of ϕ, and a couple of computer-generated histograms (10000 and 80000 values of S_{100} respectively), for the distribution of $S_r(X)$, are presented below. Some tricks are needed, and some corners have to be cut, to get this many data points in reasonable time. See the author's home page on the Internet for details.

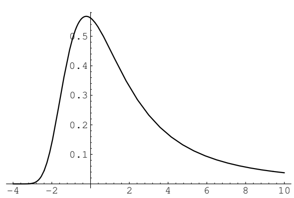

Fig. 1.

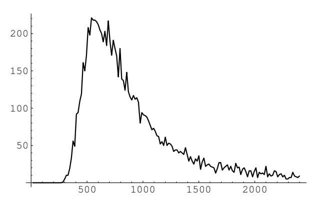

Fig. 2.

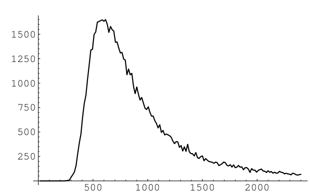

Fig. 3.

The proof of the main result breaks into two main parts. The typical digit sequence (a_i) consists mostly of small digits, punctuated by occasional larger digits. We choose the cutoff at R with $R := [r^{1/2+\epsilon}]$, reserving the choice of ϵ for later. (Eventually, ϵ is taken to be $1/12$.) There will almost always be on the close order of $(\log 2)^{-1}r^{1/2-\epsilon}$ of these large digits, and they are effectively statistically independent. The conditional expected value M of a small digit, (one bounded by R), is finite. The sum of the small digits normally turns out to contribute rather more than half the grand total, and for most x is close to rM. That is, while the shape of the distribution function is determined by the large digits, the small digits move that distribution (of the sum of the large digits) to the right by a nearly-deterministic rM. Proving this involves some probabilistic approximation of the original 'game' with other, more suitable games that assign very nearly the same probability to 'most' of the possible digit sequences, followed by an appeal to known facts concerning the functional analysis of continued fractions. The idea of approximating a measure G with a similar but more tractable G^* goes back to P. Lévy, who used it in [**11**] to study continued fraction questions similar to ours.

The distribution of the sum of the large digits is effectively modeled by the sum of $N = [(\log 2)^{-1} r^{1/2-\epsilon}]$ independent, identically distributed random variables, each with probability density function $f_R := R/x^2$ on the interval (R, ∞). The second main part of the proof is an analysis of this sum. For all stable laws F with *characteristic exponent* $\alpha > 1$, F extends from the real line to an entire function, [**4**, **10**]. In our case, $\alpha = 1$ and so a particular argument is needed. From the Fourier transform of ϕ we are able to work back to the asymptotic behavior of $\phi(x)$ as $x \to \infty$ and as $x \to -\infty$, by way of classical complex analysis starting with the formula for the inverse Fourier transform. The analysis of the sum of the large digits yields, en passant, the asymptotic distribution of the sum of n independent identically distributed random variables, each with the same probability density function ($1/x^2$ for $x > 1$ else 0) as $1/X$: It too is a scaled and shifted copy of $\phi(x)$: As $n \to \infty$,

$$\text{prob}\left[\sum_1^n 1/X_k \le n(\log n + 1 - \gamma + Z)\right] \to \int_{-\infty}^{Z} \phi(x)\,dx$$

uniformly over $Z \in R$.

2. Games with essentially equal probability measures. Here we introduce a variety of countable probability spaces, described as games of chance. Apart from the first 'game', in which a single number is chosen from the open interval (0,1) with the usual measure, our games are Markov chains. The states of such a chain can involve a positive integer, a real number in $(0, 1)$, or both, but there is no dependence on the past, as there is in the evolution of the digits of the continued fraction expansion of a random real number. Two probability measures μ_j and μ_k say, both on the same countable event set, will be termed equivalent to within a given margin of error δ, if the set of elements a for which $\mu_j(a) \notin (1 \pm \delta)\, \mu_k(a)$, or vice-versa, has measure less than δ in both spaces.

The original game is to pick a random number $X \in (0, 1)$, extract the first r partial quotients or digits in its continued fraction expansion, and form their sum. The issue is to characterize the asymptotic behavior of the distribution function $p_r(z) := \text{prob}(S_r(X)) \le z$. This game, call it G_1, is in some respects an unsatisfactory representation of random continued fraction digits. There is some short-range correlation between one digit and the previous ones in the continued fraction expansion of X; this game does not amount to a Markov chain on the digits. The game G_2 described below provides a Markov chain (game) for which the probability of any digits string is the same as in game 1.

Play proceeds through r stages, with one digit being chosen at each stage. At the outset, a state parameter θ is set to zero. At stage $k+1$, the probability of choosing $a_{k+1} = n$ is $(1+\theta)/((n+\theta)(n+\theta+1))$. Once a_{k+1} is chosen, set $\theta \leftarrow 1/(\theta + a_{k+1})$. The probability of a given sequence $(a_1, a_2, \dots a_r)$ of continued fraction partial quotients is exactly the same in both games: $\langle a_1, a_2, \dots a_r\rangle^{-2}(1+\{a_1, a_2, \dots a_r\})^{-1}$, where $\{a_1, a_2, \dots a_r\}$ denotes the continued fraction $1/(a_r + 1/(a_{r-1} + \cdots + 1/a_1)\dots)$, and $\langle a_1, a_2, \dots a_r\rangle$, its denominator. This fraction is the value of θ after those r digits have been chosen, incidentally. An equivalent characterisation of this game is that after digits $a_1 \cdots a_k$ have been chosen, a real number Y in $[0,1]$ is chosen according to the probability density function $(1 + \{a\})/(1 + \{a\}t)^2$ where $\{a\}$ denotes the continued fraction $[a_k, a_{k-1}, \dots a_1]$. The next digit is then chosen to be the integer part of $1/Y$. The other variants introduced below, of the 'game' of choosing digits, can likewise be given this kind of formulation. In every case, the conditional probability density function for Y is a positive linear combination of functions of the form $(1 + \theta)/(1 + \theta t)^2$, with $0 \le \theta \le 1$.

Equivalent$_{r^{-\epsilon}}$ variants of this game change the probabilities, but rarely by much. That is, the set of sequences a of r positive integers falls into two subsets, 'typical' and 'atypical' say. The probabilities the two games assign to elements of 'typical' differ by a factor of $1 \pm O(r^{-\epsilon})$. The probabilities assigned to elements of 'atypical' may be quite different, but the whole of 'atypical' has probability $O(r^{-\epsilon})$ in both games. One such game is G_3. In this game, we round θ down to zero following any occurence of a large digit: $a_k \ge R \Rightarrow \theta \leftarrow 0$. Again, the 'state' of this Markov chain is the current value of θ. This game has the property that for all $1 \le j \le r$ and all sequences $a = (a_1, a_2, \dots a_r)$ with $a_j \ge R$,

$$\text{prob}_{G_3}(a) = \text{prob}_{G_3}(a_1 \cdots a_j)\text{prob}_{G_3}(a_{j+1} \cdots a_r).$$

To see that games 2 and 3 are nearly equivalent, and find the extent δ to which they are not identical (in the probabilities assigned to digit sequences), we begin with the observation that in both games, the conditional probability, given an arbitrary initial sequence $(a_1, a_2, \dots a_k)$ of digits, that the next digit a_{k+1} will be large, is never more than $2/R$. Therefore, in both games the probability of having several large digits is small: If N denotes the number of large digits, then both $\text{prob}_{G_1}(N \ge z) = \text{prob}_{G_2}(N \ge z)$

and $\mathrm{prob}_{G_3}(N \geq z)$, are less than the probability of z or more heads in r independent tosses of a coin that shows heads with probability $2/R$ and tails with probability $1 - 2/R$. Taking $z := [3r^{1/2+\epsilon}]$ makes even this latter probability $\ll r^{\epsilon-1/2}$ on grounds of Cauchy's inequality. Apart from sequences $(a_1, \dots, a_r)$ with $N(a) \geq 3r^{1/2-\epsilon}$, though, all digit sequences have essentially the same probability in either game. First of all, the two games assign exactly the same probability to any sequence a with no large digits. Consider next a sequence with exactly one large digit, say a_k.

The probability of a digit sequence $(a_1 \cdots a_k, a_{k+1} = b_1 \cdots a_r = b_{r-k}) = ab$, say, is given in games 1 and 2 by

$$\langle a\rangle^{-2}\langle b\rangle^{-2}(1 + \{a\}[b])^{-2}(1 + \{ab\})^{-1}$$

while in game 3, and assuming $a_k \geq R$, it is

$$\langle a\rangle^{-2}\langle b\rangle^{-2}(1 + \{a\})^{-1}(1 + \{b\})^{-1}.$$

Now $\{a\} \leq 1/R$, and $\{b\} = (1+O(1/R))\{ab\}$. The ratio here is $(1+O(1/R))$. For a sequence a with N large digits, the same calculation shows that the ratio of probabilities assigned to a by games 2 and 3 is $(1 + O(1/R))^N$, and with $N \ll r^{1/2-\epsilon}$ that's $1 + O(r^{-2\epsilon})$. So we have equivalence $_\delta$ with $\delta = O(r^{-2\epsilon})$.

For some purposes we need yet more statistical independence in our model of digit-choosing. In game 4, large and small digits are chosen independently. This game is played exactly like game 3 except on those occasions when a large digit happens to be chosen. In this case, instead of simply using it and resetting the auxiliary θ to zero, we discard the large digit chosen, and draw again from the "urn of large digits", this time with $\mathrm{prob}_4(n) := R/(n(n+1))$ for $n \geq R$, and zero otherwise. We then set θ to zero as before. The new probability differs by a factor of $(1 + O(1/R))$ from the conditional probability in game 3 that digit $a_k = n$ given the history $a_1 \cdots a_{k-1}$ (i.e. given θ), and given that $a_k \geq R$. Since the probabilities in games 3 and 4 are changed only by reshuffling the mass allotted to the individual large digits, the probabilities for N are unchanged and this game, too, is equivalent$_\delta$ to the others, still with $\delta = O(r^{-2\epsilon})$. Consequently, insofar as Theorem 1 is concerned, all these games are equivalent.

The point of game 4 is that we can partition the event space according to the subset $T \subset \{1, 2, \dots, r\}$ of places at which a_k is large ($\geq R$). For fixed T, we consider the conditional games $G_4(T)$ described below. Given a set $T \subset \{1, 2, \dots, R\}$, and a sequence $a = (b_0 n_1 b_1 \cdots n_N b_N)$ in which the b_j represent strings or sequences of $k_j := (t_{j+1} - t_j - 1)$ integers all between 1 and $R - 1$, and the n_j are integers $\geq R$, (that is, a sequence of r positive integers for which T is the set of positions at which 'large' integers are found), game $G_4(T)$ assigns probability $\mathrm{prob}_4(a)/\mathrm{prob}_4(T)$ to a. Equivalently, with $N = N(T) =$ the cardinality of T, and with b's, k's and n's defined in terms of a as above, we have

$$\operatorname{prob}_{4,T}(a) = \frac{\prod_{j=0}^{N-1}\langle b_j\rangle^{-2}\left(\frac{R}{R+\{b_j\}}\right)\prod_{j=1}^{N} n_j^{-1}(n_j+1)^{-1}\cdot\langle b_N\rangle^{-2}(1+\{b_N\})^{-1}}{\prod_{j=0}^{N-1}\sum_{b\in V_R(k_j)}\langle b\rangle^{-2}(R+\{b\})^{-1}\cdot\sum_{b\in V_R(k_N)}\langle b\rangle^{-2}(1+\{b\})^{-1}}.$$

For most T, (including all T with $N(T)\le 4r^{1/2-\epsilon}$), this conditional probability agrees, to within a factor of $(1\pm O(r^{-2\epsilon}))$ with

$$\frac{\prod_{j=0}^{N-1}\langle b_j\rangle^{-2}\prod_{j=1}^{N} Rn_j^{-1}(n_j+1)^{-1}\cdot\langle b_N\rangle^{-2}(1+\{b_N\})^{-1}}{\prod_{j=0}^{N-1}\sum_{b\in V_R(k_j)}\langle b\rangle^{-2}\cdot\sum_{b\in V_R(k_N)}\langle b\rangle^{-2}(1+\{b\})^{-1}}.$$

As mentioned earlier, the sum of the small digits is a random variable with relatively little dispersion. Specifically, we shall see that $N(T)\ge 3r^{1/2-\epsilon}$ with probability $O(r^{-2\epsilon})$ (actually, far less), and that for all other T, the conditional probability in game 4 (T) that the sum of the small digits falls outside $\frac{r}{\log 2}\left((\frac{1}{2}+\epsilon)\log r-1\right)+O\left(r^{(3/4)+2\epsilon}\right)$ is $O(r^{-\epsilon})$. From the perspective of the theorem, this may as well say that the sum of the small digits is always $\frac{r}{\log 2}\left((\frac{1}{2}+\epsilon)\log r-1\right)$. There is a deep and fruitful connection [**7**, **8**, **12**] between the metric theory of continued fractions and the linear operator

$$G : f(t)\to\sum_{k=1}^{\infty}(k+t)^{-2}f(1/(k+t)).$$

This operator, and a variant L in which the upper limit of summation is $R-1$, will be at the center of the calculations ahead. At this point it will be convenient to introduce some more terminology.

Recall that $\langle b\rangle=\langle(b_1,b_2,\ldots,b_m)\rangle$ represents the denominator of $1/(b_1+1/(b_2+1/(\cdots+1/b_m)\ldots))$, that $[b]$ represents the fraction itself, and $\{b\}$, the fraction corresponding to the sequence $(b_m,b_{m-1}\cdots b_1)$. Let $V(k)$ denote the Cartesian product of k copies of Z^+, and $V_m(k)$ the Cartesian product of k copies of $\{1,2,\ldots,m\}$. Let $\|b\|$ denote $\sum_{j=1}^{k} b_j$ for $b\in V(k)$. Let L_m be the operator carrying f (a function from C to C) to the function $t\to\sum_{j=1}^{m}(j+t)^{-2}f(1/(j+t))$, also defined on C. For our purposes the important thing will be the action of powers of L_{R-1} on constant functions, and on functions of the form $t\to(1+\theta t)^{-2}$.

The basis of the connection between this operator and questions of continued fractions is the fact that

$$L_m^k(1)=\sum_{b\in V_m(k)}\langle b\rangle^{-2}(1+\{b\}t)^{-2}$$

and that $g = g_\infty = 1/(\log 2(1+t))$ is the leading eigenvalue of L_∞. More generally,

$$L_m^k(1+\theta t)^{-2} = \sum_{b \in V_m(k)} \langle \theta + b \rangle^{-2}(1 + \{\theta + b\}t)^{-2}$$

where $\theta + b$ denotes the sequence $(\theta + b_1, b_2, \dots b_k)$ and where $0 \leq \theta \leq 1$. We quote a result [**7**, Lemma 2.1 etc.], adapted and specialized to the case at hand:

Lemma 1. *There is a sequence of probability density functions* $g_m : [0,1] \to R^+$*, and numbers* $\lambda_m \in (0,1)$*, so that*

(a) $$L_m g_m = \lambda_m g_m$$

(b) $$g_m = \frac{1}{\log 2} \frac{1}{1+t} \left(1 + O(m^{-1/2})\right)$$

(c) $$\lambda_m = 1 - \frac{6}{\pi^2 m} + O(m^{-2} \log m)$$

(d) $$L_m^k(1+\theta t)^{-2} = \frac{\lambda_m^k g(t)}{1+\theta} \left(1 + O(m^{-1/2}) + O(3^{-k})\right)$$

The functions g_m *have the form* $\int_0^1 (1+\theta)(1+\theta t)^{-2} d\mu_m(\theta)$ *where* μ_m *is a probability measure on* $[0,1]$.

With this machinery we now set about getting estimates for the quantities

$$\sum_{b \in V_m(k)} \langle b \rangle^{-2}, \qquad \sum_{b \in V_m(k)} \langle b \rangle^{-2}(1 + \{b\})^{-1}$$

$$\sum_{b \in V_m(k)} \langle b \rangle^{-2} \|b\|, \quad \text{and} \quad \sum_{b \in V_m(k)} \langle b \rangle^{-2}(1 + \{b\})^{-1} \|b\|$$

with particular emphasis on the case $m = R - 1$, $k \in [r^\epsilon, r^{1/2+2\epsilon}]$. Apart from a negligible fraction (no more than $O(r^{-\epsilon})$) of the mass of a's, weighted according to the probabilities in any of the games $G_1 \dots G_4$, every interval k_j between consecutive large digits is of length at least r^ϵ. Indeed, prob[$\exists$ gap $< r^\epsilon$] $\leq \sum_{j=1}^{r} \sum_{k=1}^{r^\epsilon}$ prob[both a_j and $a_{j+k} \geq R-1$]. A long run of small digits is also unlikely: As we have seen, the conditional probability, in all games and for all antecedent digit strings a, that the next digit is large is always between $1/R$ and $2/R$. Thus in games 3 and 4, the probability of a string of k or more consecutive small digits in positions $j \dots j+k-1$ say, is $\leq (1 - 1/R)^k$ and so the probability of any such string is $\leq r(1 - 1/R)^k$. Setting $k = [R^{1/2+2\epsilon}]$ gives the claimed inequality.

The estimates from Lemma 1, applied in this case, lead to

(i) $$\sum_{b \in V_m(k)} \langle b \rangle^{-2}(1 + \{b\}t)^{-2} = (1 + O(\exp(-r^\epsilon)))\lambda_m^k g_m(t),$$

(ii) $$\sum_{b\in V_m(k)} \langle b\rangle^{-2} = (1+O(\exp(-r^{\epsilon})))\lambda_m^k g_m(0),$$

(iii) $$\sum_{b\in V_m(k)} \langle b\rangle^{-2}(1+\{b\})^{-1} = (1+O(\exp(-r^{\epsilon})))\lambda_m^k.$$

For the next sum

$$\sum_{b\in V_m(k)} \langle b\rangle^{-2}(1+\{b\})^{-1}\|b\|$$

we start with the observation that this is

$$\sum_{j=1}^{k}\sum_{c\in V_m(j-1)}\sum_{d\in V_m(k-j)}\sum_{n=1}^{m} n\langle cnd\rangle^{-2}(1+\{cnd\})^{-1}$$

where the juxtaposition cnd denotes the concatenation of the sequences c and d with the single entry n between them. Now $\{cnd\} = (1+O(\rho^{k-j}))\{d\}$ where $\rho = 2/(1+\sqrt{5})$. The sum here thus simplifies to

$$\sum_{j=1}^{k}(1+O(\rho^{k-j}))\sum_{c\in V_m(j-1)}\sum_{n=1}^{m}\langle c\rangle^{-2}\frac{n}{(n+\{c\})^2}$$

times an inner sum of

$$\int_0^1 \sum_{d\in V_m(k-j)} \langle d\rangle^{-2}\,(1+\theta[d])^{-2}\,(1+\{\theta+d\}t)^{-2}\,dt$$

where $\theta = 1/(n+\{c\})$. From the identities above, this inner sum is $\int_0^1 L_m^{k-j}(1+\theta t)^{-2}\,dt$. But by Lemma 1, this inner sum is $\lambda_m^{k-j}(1+1/(n+\{c\}))^{-1}(1+O(m^{-1/2})+O(3^{j-k}))$ so that

$$\begin{aligned}
&\sum_{b\in V_m(k)} \langle b\rangle^{-2}(1+\{b\})^{-1}\|b\| \\
&= \sum_{j=1}^{k}(1+O(m^{-1/2})+O(\rho^{k-j}))\lambda_m^{k-j} \\
&\qquad\cdot \sum_{c\in V_m(j-1)}\sum_{n=1}^{m}\langle c\rangle^{-2}\frac{n}{(n+\{c\})(n+1+\{c\})}.
\end{aligned}$$

The inner double sum can be simplified; with $h = j-1$ it is

$$\begin{aligned}
&\sum_{c\in V_m(h)} \langle c\rangle^{-2}\sum_{n=1}^{m} n\int_{1/(n+1)}^{1/n} L_m^h(1)(t)\,dt \\
&= \lambda_m^h(\log 2)^{-1}\sum_{n=1}^{m} n\log\left(\frac{(n+1)^2}{n(n+2)}\right)
\end{aligned}$$

$$
\begin{aligned}
&+ O(3^{-h}\lambda_m^h \log m) + O\left(\lambda_m^h \log m/\sqrt{m}\right) \\
&= \lambda_m^h \left(\frac{\log m - 1}{\log 2} + O\left(\log m/\sqrt{m}\right) + O(3^{-h}\log m)\right).
\end{aligned}
$$

and so

$$
\begin{aligned}
&\sum_{b\in V_m(k)} \langle b\rangle^{-2}(1+\{b\})^{-1}\|b\| \\
&= k\lambda_m^{k-1}\left(\frac{\log m - 1}{\log 2}\right)(1 + O(m^{-1/2}) + O(k^{-1})).
\end{aligned}
$$

In the same way we have

$$
\sum_{b\in V_m(k)} \langle b\rangle^{-2}\|b\| = k\lambda_m^{k-1}\left(\frac{\log m - 1}{\log^2 2}\right)(1 + O(m^{-1/2}) + O(k^{-1})).
$$

Now given a fixed set T of positions at which large digits are to occur, with $N(T) \ll r^{1/2-\epsilon}$ and all gaps between consecutive elements of T at least r^{ϵ}, the expected value in game 4(T) of the sum of the j^{th} sequence of k_j consecutive small digits (bordered by large digits on both ends or, in the last case, just on the left), is $(1 + O(R^{-1/2}) + O(k_j^{-1}))k_j(\log m - 1)/\log 2$ and summing this over all N gaps gives an expected value for the sum of the small digits of $(1+O(R^{-1/2}))r(\log R - 1)/\log 2$. Since this is the same, to within that margin of error, for all relevant values of T, the conditional expected value of the sum of the small digits, given that T satisifies the constraints $N(T) \ll r^{1/2-\epsilon}$,with all gaps $\geq r^{\epsilon}$, is $(1 + O(r^{-1/4}))r(\log R - 1)/\log 2$. We turn now to the question of the dispersion of the value of $\|b\|$ about this mean, and as one might expect, by way of an estimate of the second moment of $\|b\|$. For fixed T the events $\|b_0\|, \|b_1\| \ldots \|b_N\|$ are independent. Thus

$$
\operatorname{Var}\sum_{j=0}^{N}\|b_j\| = \sum_{j=0}^{N}\operatorname{Var}\|b_j\| \leq \sum_{j=0}^{N} E[\|b_j\|^2].
$$

For a string of k consecutive small digits, we can estimate $E[\|b\|^2]$. The idea of the estimate is this: We break the dependence of the digits of b by assuming at each new digit the worst: $\theta = 1$, or what is roughly equivalent, the previous two digits were a large one followed by 1. This will skew to the right the distribution of $\|b\|$ and increase its second moment, which latter second moment is easily calculated because we finally have independent random digits. The technicalities of the next few paragraphs are devoted to justifying the claim that changing the odds in this fashion really does shift the distribution of $\|b\|$ right. In this calculation we are concerned with the conditional probability, (in any of games 2 through 4) that $\|b\| = t$, given an initial value for θ, and given that for (at least) the next k digits, no digit as large as R occurs. With these conditions, the probability of a sequence

b is the product of the conditional probabilities for the consecutive digits of b, when those are given by the formulae

$$p(\theta, n) := \frac{\int_{s=1/(n+1)}^{1/n} (1+\theta s)^{-2}\, ds}{\int_{s=1/R}^{1} (1+\theta s)^{-2}\, ds}, \quad p(b) := \prod_{j=1}^{k} p(\{b_1, b_2, \dots, b_{j-1}\}, b_j).$$

If we use instead the worst-case odds, we have a different probability:

$$q(n) := \frac{\int_{s=1/(n+1)}^{1/n} (1+s)^{-2}\, ds}{\int_{s=1/R}^{1} (1+s)^{-2}\, ds}, \quad q(b) := \prod_{j=1}^{k} q(b_j)$$

in which the digits of b are independent. Now let $p_k(t) := \mathrm{prob}_p(\|b\| = t) = \sum_{b \in V_{R-1}(k)} p(b)$, and $q_k(t) := \mathrm{prob}_q(\|b\| = t) = \sum_{b \in V_{R-1}(k)} q(b)$. The sequences $p_k(t), (t \geq 1)$ and $q_k(t), (t \geq 1)$ are sequences of finitely many positive numbers followed by zeros, and with sum 1. The claim above about skewed distributions comes to this: $(p_k(t))$ is <u>more peaked</u> than $(q_k(t))$. Equivalently, $\sum_{j=1}^{t} p_k(j) \geq \sum_{j=1}^{t} q_k(j)$ for all $t \geq 1$ and all $k \geq 1$. For $k = 1$, this claim boils down to the readily verified proposition that for $t \geq 1$ and $0 \leq \theta \leq 1$,

$$\frac{\int_{1/(t+1)}^{1} (1+\theta s)^{-2}\, ds}{\int_{1/R}^{1} (1+\theta s)^{-2}} \geq \frac{\int_{1/(t+1)}^{1} (1+s)^{-2}\, ds}{\int_{1/R}^{1} (1+s)^{-2}}.$$

Assuming this peakedness inequality for $k = K$, we now consider the case $k = K + 1$. We have

$$\begin{aligned}
&\sum_{j=1}^{t} p_{K+1}(j) \\
&= \sum_{n=1}^{t} \sum_{\substack{a \in V_{R-1}(K) \\ \|a\| = t-n}} \langle a \rangle^{-2} (1 + \{a\})^{-1} \frac{\int \mathrm{max}_{[1/(n+1), 1/R]} (1 + s\{a\})^{-2}\, ds}{\int_{1/R}^{1} (1 + s\{a\})^{-2}\, ds} \\
&\geq \sum_{n=1}^{t} \sum_{\substack{a \in V_{R-1}(K) \\ \|a\| = t-n}} \langle a \rangle^{-2} (1 + \{a\})^{-1} \frac{\int \mathrm{max}_{[1/(n+1), 1/R]} (1 + s)^{-2}\, ds}{\int_{1/R}^{1} (1 + s)^{-2}\, ds} \\
&= \sum_{n=1}^{t} \sum_{\substack{a \in V_{R-1}(K) \\ \|a\| = t-n}} \langle a \rangle^{-2} (1 + \{a\})^{-1} \sum_{j=1}^{n} q_1(j) = \sum_{n=1}^{t} Q_1(n) p_k(t-n)
\end{aligned}$$

say, where $Q_1(n) := \sum_{j=1}^{n} q_1(j)$. On the other hand,

$$Q_{K+1}(t) = \sum_{j=1}^{t} q_{K+1}(j) = \sum_{n=1}^{t} Q_1(n) q_K(t-n).$$

We are now in a position to use a simple peakedness lemma. The proof is an exercise in summation by parts and is left to the reader.

Lemma 2. *If (c_n) is a positive, non-decreasing sequence, and if (d_n) and (e_n) are sequences of positive numbers with (d_n) more peaked than (e_n), then for all $t \geq 1$, $\sum_{n=1}^{t-1} c_n d_{t-n} \geq \sum_{n=1}^{t-1} c_n e_{t-n}$. From this lemma, then, for all $t \geq 1$ we have $\sum_{j=1}^{t} p_K(j) \geq \sum_{j=1}^{t} q_K(j)$ as claimed, which completes the induction.*

Choosing k small digits for b according to the probabilities q is a game of k independent trials. The mean or expected value of a single digit's contribution to $\|b\|$ is $\mathrm{E}_q(k) = \sum_{n=1}^{R-1} \frac{2n(R+1)}{(R-1)(n+1)(n+2)} \approx \log R$, and the variance is $\mathrm{Var}_q(k) = \sum_{n=1}^{R-1} \frac{2(R+1)(n-E_q(k))^2}{(R-1)(n+1)(n+2)} \ll R$. Hence for $\|b\|$ itself, the variance is dominated by kR. Therefore the second moment, with q probabilities, of $\|b\|$ is dominated by $kR + (k \log R)^2$, so that also for $\|b\|$ with the original conditional probabilities p, conditional probabilities the second moment is dominated by $kR + (k \log R)^2$. That is, we have shown that

$$\lambda_{R-1}^{-k} \sum_{b \in V_{R-1}} \|b\|^2 \langle b \rangle^{-2} (1 + \{b\})^{-1} \ll kR + (k \log R)^2.$$

Now in game 4, the various sequences of small digits, punctuated by digits $\geq R$, are themselves independent. Thus for all $T \subset \{1, 2, \ldots, r\}$ with maximum gap $k_j \leq 2r^{1/2+\epsilon}$, the conditional expected second moment, given that large digits are found at all positions corresponding to T and only there, is

$$\sum_{j=0}^{N(T)} \sum_{b_0 \in V_{R-1}(k_0), \ldots b_{N(T)} \in V_{R-1}(k_{N(T)})} \|b_j\|^2 \mathrm{prob}(b_0 \wedge b_1 \cdots \wedge b_{N(T)})$$
$$\ll \left[r^{1/2+\epsilon} R + r^{1+2\epsilon} \log^2 r \right] \cdot r^{1/2+\epsilon} \ll r^{3/2+\epsilon} \log^2 r.$$

Therefore, in all but intervals of aggregate length $O(r^{-\epsilon})$ in $[0,1]$, the sum of the small digits amongst the first r digits is $(r/\log 2)((1/2+\epsilon)\log r - 1) + O(r^{3/4+2\epsilon})$.

This brings us to the question of the statistics of the large digits. While the small digits do contribute to the determination of the set-point of the distribution of $\|a\|$, they have essentially nothing to do with its shape. The sum of the large digits is the chancy part; the big plays decide the game.

3. The number of large digits. For fixed T, and assuming that $N(T) \leq r^{1/2-\epsilon}$, the sequence of large digits in game 4 (T) is a sequence of $N(T)$ independent, identically distributed random variables. The probability of a given digit taking value $n \geq R$ is $R/(n(n+1))$; these probabilities sum to 1. We begin the analysis of the distribution of the sum of these digits by observing that the distribution will be essentially the same if we use the continuous probability density function Rt^{-2} on the interval $[R, \infty)$ in place of the original discrete measure: the effect is the same as if, after choosing a digit n, we add a random fractional quantity with probability density $n(n+1)(n+t)^{-2}$ on $[n, n+1]$. There is enough leeway in the statement of the main result that this change cannot affect the conclusion.

Let $Y_j, 1 \le j \le N$ be a sequence of N independent random variables with probability density function $f_R(t) := Rt^{-2}$ on $[R, \infty)$ (and 0 otherwise). The random variable $\mathrm{LDSum}_N := \sum_{j=1}^{N} Y_j$ has a probability density function $\rho_{R,N}(x)$ which is the N-fold convolution of f_R. Before getting deeply involved in the estimation of $\hat{\rho}$ and eventually ρ itself, we need to pin down N more tightly.

In game 4, the probability of an initial run of exactly m consecutive small digits is

$$\sum_{b \in V_R(m)} \langle b \rangle^{-2} \int_0^{1/R} (1 + \{b\}t)^{-2}\, dt = \int_0^{1/R} (L^m 1)(t)\, dt$$
$$= (1 + O(R^{-1/2}) + O(3^{-m})) \frac{\lambda_R^m}{(R \log 2)}.$$

This is also the conditional probability of such a run immediately following a large digit. Thus the conditional probability of a run ending with the very next digit, given that there have been exactly k small digits since the most recent large digit or since the beginning, is $(1+O(3^{-k})+O(R^{-1/2}))/(R\log 2)$.

In game 4, $\mathrm{prob}[N(T) \ge 4r^{1/2-\epsilon}] \ll r^{-\epsilon}$, and $\mathrm{prob}[\min[k_j] \le r^{\epsilon}] \ll r^{-\epsilon}$. For any other T, we have

$$\begin{aligned}
&\mathrm{prob}_4[\text{large digits occur exactly for indices in } T] \\
&= \prod_{j=0}^{N(T)} \prod_{i=1}^{k_j} \left(1 - (1 + O(3^{-i}))/R\log 2\right) \cdot \prod_{j=1}^{N(T)} \frac{1}{R \log 2} \left(1 + O(3^{-k_j})\right) \\
&= (1 + O(r^{-\epsilon}))(R \log 2)^{-N(T)} \left(1 - 1/R\log 2\right)^{r - N(T)}.
\end{aligned}$$

Apart from the error factor in front, this last is the probability of getting heads at exactly the tosses corresponding to elements in T, on r Bernoulli trials of a coin with probability of heads equal to $1/R\log 2$ and tails, $1 - 1/R\log 2$. Elementary calculations give $\mathrm{prob}_{\mathrm{Bernoulli}}[\#\text{heads} \notin \frac{r}{R\log 2} \pm O(r^{1/4})] \ll r^{-\epsilon}$ and in view of the previous estimates, the same holds for $\mathrm{prob}_4[N(T) \notin \frac{r}{R\log 2} \pm O(r^{1/4})]$. This brings us to the main issue of the next section: what do we get when we convolve N copies of the function $f_R(x) = Rx^{-2}$ for $x \ge R$, and zero otherwise? The precise value of N, between $(r^{1/2-\epsilon}/\log 2 \pm Cr^{1/4})$, is immaterial to our purposes since the convolution of $Cr^{1/4}$ copies of f_R is, as we shall see, a probability density function with all but $O(r^{-\epsilon})$ of its mass concentrated into the (sufficiently narrow) interval $[0, r^{(3/4)+2\epsilon}]$.

4. The sum of the reciprocals of random numbers in $(0,1)$. To study the convolution powers of f_R, it will suffice to investigate the convolution powers of $u(x) := 1/x^2 (x \geq 1), 0 (x < 1)$, as $f_R^{*n} = (1/R)u^{*n}(x/R)$. Note that $u^{*n}(x)$ is the probability density function of the sum of n independent, identically distributed random variables $1/X_j$ when X_j has uniform distribution on $(0,1)$. (Hence the title of this section.)

The distribution function corresponding to the density u is of course $U(x) := 1 - 1/x$ $(x \geq 1)$. According to a theorem of Gnedenko [**4**, Theorem 5, p. 182], U belongs to the 'domain of attraction' of a stable law with $\alpha = 1$. We need to know how fast this attraction works. With $\hat{} := f \to (t \to \int_{-\infty}^{\infty} f(z) e^{-izt}\, dz)$, we have

$$\begin{aligned} v(t) := \hat{u}(t) &= \int_1^{\infty} z^{-2} e^{-izt}\, dz \\ &= 1 + it \log|t| - \frac{1}{2}\pi |t| + it(\gamma - 1) - \sum_{k=2}^{\infty} \frac{(-it)^n}{(n-1)n!} \end{aligned}$$

which is the known function $\mathrm{Ei}(2,t)$. The series part of the expansion extends to an entire function, but the logarithmic term requires a branch cut, which we take along the negative real axis. This gives an extension of $v(t)$ to the rest of C. For completeness we extend v to all of C by taking limits from above the branch cut. Conversely, $u(z) := \check{v}(z) = \frac{1}{2\pi} \int_{-\infty}^{\infty} v(t) e^{+itz}\, dt$.

Let $u_1(z) := u(z)$, and for $n \geq 1$, let $u_n := u * u_{n-1}$. Then $\widehat{u_n}(z) = (v(t))^n$. From the series expansion of $v(t)$, for $n > 1$ we have

$$v^n(t) = \exp\left(-\frac{\pi}{2} n|t| + int \log|t| + int(\gamma - 1) + O(nt^2(1 + \log^2 |t|))\right)$$

uniformly in $|t| \leq 1/(\sqrt{n} \log n)$. Now let

$$\phi(z) := \frac{1}{2\pi} \int_{-\infty}^{\infty} \exp\left(-\frac{\pi}{2}|s| + is \log|s| + isz\right) ds$$

$$\text{so that } \hat{\phi}(t) = \exp\left(-\frac{\pi}{2}|t| + it \log|t|\right).$$

Note that for $z \in R$, $|\phi(z)| \leq 1$ and $\int_{-\infty}^{\infty} \phi(z)\, dz = 1$, while for $t \in R$, $|\hat{\phi}(t)| \leq 1$, with $\hat{\phi}(0) = 1$. Another observation is that the Fourier transforms of suitably shifted and scaled versions of ϕ are quite close to $v^n(t)$:

$$\left[\widehat{\frac{1}{n}\phi\left(\frac{z}{n} - \log n + \gamma - 1\right)}\right](t) = e^{-int(\log n + \gamma - 1)} \hat{\phi}(nt)$$

$$= \exp\left(-\frac{\pi}{2}n|t| + int\log|t| + int(\gamma-1)\right).$$

Thus for $n > 1$ and uniformly over $t \in R$ with $|t| \le 1/(\sqrt{n}\log^2 n)$,

$$v^n(t) = \left[\widehat{\frac{1}{n}\phi\left(\frac{z}{n} - \log n + \gamma - 1\right)}\right](t) \cdot \left(1 + O(nt^2(1+\log^2|t|))\right).$$

We now claim that uniformly for $z \in R$,

$$\left|\int_{-\infty}^{z} u_n(x) - \frac{1}{n}\phi\left(\frac{x}{n} - \log n + \gamma - 1\right)\,dx\right| \ll \frac{\log n}{\sqrt{n}} \qquad \text{and}$$

$$\left|\int_{-\infty}^{z} (f_R(x))^{*n} - \frac{1}{nR}\phi\left(\frac{x}{nR} - \log n + \gamma - 1\right)\,dx\right| \ll \frac{\log n}{\sqrt{n}}.$$

Proof. According to the Berry-Esseen inequality [**3**], 'If $F(x)$ and $G(x)$ are probability distribution functions with Fourier transforms $f(t)$ and $g(t)$ respectively, then for all $T > 0$,

$$\sup_{x\in R}|F(x) - G(x)|$$
$$\le c_1\left\{\sup_{y\in R}\frac{1}{2}T\int_0^{1/T}(G(y+u) - G(y-u))\,du + \int_{-T}^{T}\left|\frac{f(t)-g(t)}{t}\right|\,dt\right\}$$

where $c_1 > 0$ is an effectively computable constant.' (Remark: The constant may be computed directly from the proof given in [**3**]; I get $c_1 = 24.69....$)

We take $G(x) := \frac{1}{n}\int_{-\infty}^{x}\phi\left(\frac{z}{n} - \log n + \gamma - 1\right)\,dz$ and $F(x) := u_n(x)$, so that $f(t) = v^n(t)$ and $g(t) = \exp\left(-\frac{\pi}{2}n|t| + int\log|t| + int(\gamma-1)\right)$. We then take $T = 1/(\sqrt{n}\log n)$. Since $|\phi| \le 1$ on R, the first term in the Berry-Esseen bound is no more than $1/2nT = \log n/2\sqrt{n}$. For the second term, when $|t| \le T$ we have $|(f(t)-g(t))/t| \ll n|t|e^{-\pi n|t|/2}$. Thus $\int_{-T}^{T}|(f(t)-g(t))/t| \ll n\int_0^\infty t\exp(-\pi nt/2)\,dt = 4/\pi^2 n$. Both claims now follow. From the first claim our (much) earlier assertion that $\mathrm{prob}[\sum_1^n 1/X_k \le n(\log n + 1 - \gamma + Z)] \to \int_{-\infty}^{Z}\phi(x)\,dx$ is immediate, and with an error bound of $O(\log n/\sqrt{n})$ which is uniform in Z.

In view of our estimate for the sum of the small digits, it now follows that

$$\mathrm{prob}_{G1}\left[\sum_{k=1}^{r} a_k \le z\right] = O(r^{-\epsilon}\log^2 r) + \int_{-\infty}^{z-h(r)} R^{-1}u_n(x/R)\,dx$$

where $h(r) := (r/\log 2)((1/2+\epsilon)\log r - 1) + O(r^{3/4+2\epsilon})$, and where $n = [r^{1/2-\epsilon}/\log 2]$. Routine calculus with $\rho := r/\log 2$, $\epsilon := 1/12$, and taking into account the fact that $\phi(x) \le 1$ on R, reduces this to

$$\mathrm{prob}_{G1}\left[\sum_{k=1}^{r} a_k \le z\right] = \int_{-\infty}^{z-\rho(\log\rho+\gamma)} \rho^{-1}\phi(s/\rho)\,d\rho + O(E)$$

$$= \int_{-\infty}^{z/\rho - \log\rho + \gamma} \phi(s)\, ds + O(E)$$

where $E = r^{-1/12}\log^2 r$. This, apart from the characterisation of ϕ, is our main theorem.

Remark on Lévy's conjecture about this sum: For arbitrary positive λ, and $Y \in R$, prob $[\sum_{k=1}^{n} X_k < Y] < \int_{-\infty}^{Y} u^{*n}(z)\exp(\lambda(Y-z))\, dz$. The Fourier transform of $u(z)\exp(-\lambda z)$ is

$$1 + (\lambda + it)(\log(\lambda+it) + \gamma - 1) - \sum_{k=2}^{\infty} \frac{(-1)^k(\lambda+it)^k}{(k-1)k!}.$$

Setting $t = 0$ and using the inequality $(1-a)^n < \exp(-na)$ for $0 < a < 1$ we have

$$\text{prob}\left[\sum_{k=1}^{n} X_k < Y\right] < \exp\left(\lambda Y + n\lambda\log\lambda + n(\gamma-1)\right).$$

Setting $Y = n\log n + n\theta - n\log\log n$ and $\lambda = n^{-1}e^{-\gamma-\theta}\log n$ gives prob $<$ $\exp(-e^{-\gamma-\theta}\log n)$. The sum over n of this upper bound is finite if and only if $\theta < -\gamma$, or equivalently, $c < e^{-\gamma}$. The bounds obtained with this technique are usually quite good. The conjecture should thus be that the sum of the reciprocals of n $[0,1]$ uniform random variables is infinitely often less than $n\log n + \log c - \log\log n$ if and only if $c < e^{-\gamma}$.

5. Properties of $\phi(x)$. Recall that ϕ is defined as the inverse Fourier transform of $\exp\left(-\frac{\pi}{2}|t| + it\log|t|\right)$. We now put $\phi_\theta(x) := \theta^{-1}\phi(x/\theta - \log\theta)$ for $\theta > 0$. Then $\widehat{\phi_\theta}(t) = (\hat\phi(t))^\theta$. Thus $(\phi * \phi_\theta)\hat{}(t) = \hat\phi(t)\widehat{\phi_\theta} = (\hat\phi(t))^{1+\theta}$, so that $\phi * \phi_\theta = \phi_{1+\theta}$. In particular, and in contrast to the way the best-known stable function $\exp(-x^2)$ behaves, $(\phi * \phi)(x) = \frac{1}{2}\phi(x/2 - \log 2)$.

The factor $\exp(-(\pi/2)|t|)$ in the definition of ϕ ensures that ϕ extends to a function analytic at least in a strip $|\Im(z)| < \pi/2$. Thus

$$\phi^{(n)}(z) = \frac{1}{2\pi} i^n \int_{-\infty}^{\infty} s^n \exp\left(-\frac{\pi}{2}|s| + is\log|s| + isx\right) ds$$

$$\text{so that } |\phi^{(n)}(x)| \le n! 2^n \pi^{-n-2}.$$

Because ϕ is the limit of shifted and scaled versions of u_n's, it is non-negative on R. But since it is analytic on a strip including R, it is positive a.e. on R. Finally, since $\phi * \phi(x) = \phi(x/2 - \log 2)$, ϕ is strictly positive on all of R.

Remark. Not all stable probability density functions are positive on the entire real line. Those with $\alpha < 1$ and $|\beta| = 1$ have support an interval of the form $[c, \infty)$ or $(-\infty, c]$. From this one might say that ϕ barely escapes this fate. The rapid decay of ϕ to zero on the left fits in with this observation.

The asymptotic analysis of ϕ begins with the observation that for x real,

$$\phi(x) = \frac{1}{\pi}\Re\left[\int_0^\infty e^{-\pi s/2}\exp(i(s\log s + xs))\,ds\right].$$

We take an arbitrarily large M and deform the path of integration so that it runs in straight segments from 0 to Mi to $M(1+i)$ to M, and then to ∞ along R. The contributions from all but the first leg of this path tend to zero as $M\to\infty$, so (with $y = is$),

$$\phi(x) = \frac{1}{\pi}\int_0^\infty \sin(\pi y)e^{-y\log y - xy}\,dy.$$

From this formula it is clear that ϕ actually extends to an entire function.

Our immediate interest, though, is in the behavior as $x\to+\infty$. We have

$$\begin{aligned}\phi(x) &= \frac{1}{\pi}\int_0^{1/\sqrt{x}}(\pi y + O(y^3))e^{-y\log y - xy}\,dy + O\left(\int_{1/\sqrt{x}}^\infty e^{-xy+(1/2)y\log x}\,dy\right)\\ &= \int_0^{1/\sqrt{x}} ye^{-xy}(1 - y\log y + O(y^2\log^2 x))\,dy + O(x^{-4})\\ &= x^{-2} + O(x^{-3}\log x).\end{aligned}$$

Remark. The analysis may be carried to more terms, most conveniently with the aid of a computer algebra system. Mathematica gives $x^{-2} - 2x^{-3}\log x + (3-2\gamma)x^{-3} + O(x^{-4}\log^2 x)$ and higher-order expansions. The series does not appear to be convergent and x must be rather large before even the third term improves matters.

For $x\to-\infty$, we take $\alpha := e^{|x|-1}$ and deform the path $[0,\infty)$ to run instead from 0 to $-\alpha i$ and thence to $-\alpha i + M$ to M and on to ∞. Now if $s = \sigma - it$ with $0\le t\le\alpha$ and $\sigma > 0$, then the log of the integrand is

$$\begin{aligned}&-\frac{1}{2}\pi\sigma + \frac{1}{2}\sigma\log(\sigma^2+t^2) + tx + \sigma\tan^{-1}(t/\sigma)\\ &- it\tan^{-1}(t/\sigma) + \frac{1}{2}i\sigma\log(\sigma^2+t^2) + ix\sigma + \frac{1}{2}i\pi t.\end{aligned}$$

The real part of this tends to $-\infty$ uniformly in $0\le t\le\alpha$ which justifies omitting the return to the real axis, and instead proceeding to $-i\alpha+\infty$ after the initial down-leg. The contribution of that down-leg, though, to the real part of the integral, is zero. Now the series expansion of $-\pi s/2 + is\log s + isx$ about $-i\alpha$, is

$$-\alpha + \sum_{n=2}^\infty \frac{(-i)^n}{n(n-1)}e^{-\alpha(n-1)}(s+i\alpha)^n.$$

Thus for $x < 0$,

$$\begin{aligned}
\phi(x) &= \frac{1}{\pi} e^{-\alpha}\, \Re \left[\int_{\sigma=0}^{\infty} \exp\left(\alpha \sum_{n=2}^{\infty} \frac{(-i)^n \sigma^n}{n(n-1)} \right) d\sigma \right] \\
&= \frac{\alpha}{\pi} e^{-\alpha}\, \Re \left[\int_{\sigma=0}^{\infty} \exp\left(-\alpha \int_0^{\sigma} \tan^{-1} \nu \, d\nu + \frac{i\alpha}{2} \int_0^{\sigma} \log(1+\nu^2)\, d\nu \right) d\sigma \right].
\end{aligned}$$

The contribution to the double integral from $\sigma \geq 1$ is $\ll \exp(-\alpha \int_0^1 \tan^{-1} \nu \, d\nu) \ll e^{-\alpha\pi/8}$. For the interval $\alpha^{-2/5} \leq \sigma \leq 1$, the series expansion above shows that the contribution is $O(\exp(-\frac{1}{3}\alpha^{1/5}))$. For $0 < \sigma \leq \alpha^{-2/5}$, though, the series for the integrand reduces to $(1 + O(\alpha^{-1/5})) \exp(-\frac{1}{2}\alpha\sigma^2)$. Thus $\phi(x) = \sqrt{\frac{\alpha}{2\pi}} e^{-\alpha} \left(1 + O(\alpha^{-1/5})\right)$. This proves our second decay-rate claim for $\phi(x)$.

Directly from the integrals that define ϕ it is a relatively simple matter to compute and plot ϕ and its first two derivatives. The resulting plots are shown below, and they fit in nicely with Wolfe's theorem. (Our ϕ is one of the functions covered by his special case of Yamazato's theorem to the effect that all distributions in class L are unimodal.) From the graphs, it looks as though more might be true. Does the n^{th} derivative of ϕ have exactly n zeros?

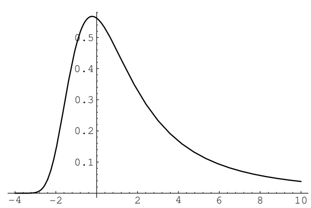

Fig. 4.

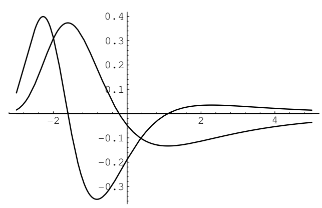

Fig. 5.

References

[1] S. Dharmadhikara and K. Joag-dev, *Unimodality, convexity, and applications*, Academic Press, San Diego, CA, 1988, [Chapter 5, Theorems 5.5 and 5.6].

[2] H. Diamond and J. Vaaler, *Estimates for partials sums of continued fraction partial quotients*, Pacific J. of Math., **122**(**1**) (1986), 73-82.

[3] P.D.T.A. Elliott, *Probabilistic number theory* I: *Mean-value theorems*, Grundlehren der Mathematischen Wissenschaften, **239**, Springer-Verlag, New York-Berlin, 1979.

[4] B.V. Gnedenko and A.N. Kolmogorov, *Limit distributions for sums of independent random variables*, Addison-Wesley, Cambridge Mass., 1954, Chapter 7 and Appendix 2.

[5] Y. Guivarc'h and Y.Le Jan, *Asymptotic winding of the geodesic flow on modular surfaces and continuous fractions*, Ann. Scient. Éc. Norm. Sup., **26** (1993), 23-50.

[6] S. Hartman, *Quelques propriétés ergodiques des fractions continues*, Studia Math., **12** (1951), 271-278.

[7] D. Hensley, *The number of steps in the Euclidean algorithm*, J. of Number Theory, **49** (1994), 142-182.

[8] ———, *Continued fraction Cantor sets, Hausdorff dimension, and functional analysis*, J. of Number Theory, **40** (1992), 336-358.

[9] A.Ya. Khinchin, *Continued Fractions*, U. of Chicago Press, Chicago, 1964.

[10] A.I. Lapin, *On some properties of stable laws*, Dissertation, 1947.

[11] P. Lévy, *Fractions continues aléatoires*, Rend. Ann., Palermo, 170-208.

[12] D. Mayer, *On the thermodynamic formalism for the Gauss map*, Commun. Math. Phys., **130** (1990), 311-333.

[13] A. Wintner, *Asymptotic Distribution and Infinite Convolution*, Edwards Brothers, Ann Arbor, MI, (1938), 30-32.

[14] S.J. Wolfe, *On the unimodality of L functions*, Annals of Math. Statistics, **42** (1971), 912-918.

[15] M. Yamazato, *Unimodality of infinitely divisible distribution functions of class L*, Ann. Probability, **6** (1978), 523-531.

Received January 16, 1996 and revised June 3, 1996

Texas A&M University
College station, TX 77843-3368
E-mail address: doug.hensley@math.tamu.edu

PACIFIC JOURNAL OF MATHEMATICS
Vol. 192, No. 1, 2000

SHARP SIZE ESTIMATES FOR CAPILLARY FREE SURFACES WITHOUT GRAVITY

Xi-nan Ma

For the equation of constant mean curvature with prescribed constant contact angle boundary condition, using the unique continuation of analytic function, we get a minimum principle for a combination of the solution and its gradient. Thus we get the endpoint case for P-function (Sperb, 1981) and in fact answer an open question which appeared twenty years ago in Payne & Philippin, 1977, 1979 and Sperb, 1981. As an application, sharp size and shape estimetes for capillary free surface without gravity are obtained.

1. Introduction and Results.

The capillary surface of a liquid contained in a vertical tube with arbitrary cross section Ω in the outer space has the shape of surface of constant mean curvature with constant contact angle θ_o against the wall of the tube. Let the capillary surface be expressed non-parametrically as the graph of a function u defined over the cross section Ω. How does the boundary geometry of Ω and the contact angle θ_o influence the size and shape of the capillary free surface?

For the convexity of the capillary free surface, in [**2**], Chen and Huang have shown if Ω is a bounded convex domain in the plane and $\theta_o = 0$, then the corresponding capillary surface is also convex. Finn [**3**] provided an example to show if $\theta_o \neq 0$ the result is in general false.

In [**1**, **10**], Chen and Sakaguchi showed if Ω be a bounded smooth convex domain in R^2, $0 < \theta_o < \frac{\pi}{2}$, the capillary free surface over Ω has only one minimal point. From the convexity of the surface as $\theta_o = 0$, we know for any θ_o $(0 \leq \theta_o < \frac{\pi}{2}$), the minimal point is unique.

In this paper we consider the influence of boundary geometry and the contact angle θ_o $(0 \leq \theta_o < \frac{\pi}{2})$ on the size and shape for the capillary free surface without gravity. Precisely, let Ω be a bounded convex domain in R^2 with smooth boundary $\partial\Omega$. Give a positive constant H, consider the following equations:

$$\sum_{i=1}^{2} D_i \left(\frac{u_i}{\sqrt{1+|Du|^2}} \right) = 2H \qquad \text{in} \qquad \Omega \tag{1.1}$$

$$\frac{u_n}{\sqrt{1+|Du|^2}} = \cos\theta_o \qquad \text{on} \qquad \partial\Omega. \tag{1.2}$$

Where $u_i, i = 1,2$ are partial derivatives of u, n denotes the unit outer normal to $\partial\Omega$, u_n denotes the direction derivative of u along n, and θ_o ($0 \le \theta_o < \frac{\pi}{2}$) is the constant with $2H|\Omega| = \cos\theta_o|\partial\Omega|$ ($|\Omega|$ is the area of Ω and $|\partial\Omega|$ is the length of $\partial\Omega$). The graph of solution u to (1.1)-(1.2) described a capillary free surface without gravity over the cross section Ω.

Let $A \in \partial\Omega$ be a point corresponding to a minimum boundary value of u, $B \in \partial\Omega$ be a point corresponding to a maximum boundary value of u, $C \in \Omega$ be the unique minimal (critical) point of u and $k(x)$ be the curvature of $\partial\Omega$ at $x \in \partial\Omega$. Now we state our theorems:

Theorem 1. *Let $u \in C^3(\bar\Omega)$ be a solution to* (1.1)-(1.2), *then the following inequalities hold*

1):

$$u(A) - u(C) \le \frac{1-\sin\theta_o}{H} \tag{1.3}$$

$$k(A) \le \frac{H}{\cos\theta_o}, \tag{1.4}$$

2):

$$u(B) - u(C) \ge \frac{1-\sin\theta_o}{H} \tag{1.5}$$

$$k(B) \ge \frac{H}{\cos\theta_o}. \tag{1.6}$$

If one of the equality signs of (1.3)-(1.6) *holds then Ω is a disk of radius $\frac{\cos\theta_o}{H}$ and*

$$u(x) - u(C) \equiv \frac{1-\sin\theta_o}{H} \qquad \text{on} \quad \partial\Omega \tag{1.7}$$

$$k(x) \equiv \frac{H}{\cos\theta_o} \qquad \text{on} \quad \partial\Omega. \tag{1.8}$$

Conversely, (1.7)-(1.8) *holds on $\partial\Omega$ if Ω is a disk of radius $\frac{\cos\theta_o}{H}$.*

The proof of Theorem 1 is based on Hopf maximum principle [**9**] and the following minimum principle.

Theorem 2. *Let $u \in C^3(\bar\Omega)$ be a solution to* (1.1)-(1.2), *then the function*

$$P(x) = 2 - 2(1+|Du|^2)^{-\frac{1}{2}} - 2Hu$$

attains its minimum on the boundary $\partial\Omega$, unless $P(x)$ is a constant on $\bar\Omega$.

In [7], Payne and Philippin had proved a similar maximum principle for the above function $P(x)$ that under the same condition it also attains its maximum on $\partial\Omega$ unless $P(x)$ is a constant in Ω.

Since our theorems concern only qualitative property of the solution, so only under the hypothesis of the existence of the solution we prove theorems. For the existence of solution and background details we refer the reader to the sources [4].

According to the work Nirenberg [5] or [4] we conclude that u is an analytic function in Ω, a feature which will be used in this paper.

In Section 2 we will give the the proof of Theorem 2 which is based on the unique continuation of analytic function. Section 3 contains a proof of Theorem 1 and a Corollary, which give the estimates of capillary free surface area using the volume of a liquid, $|\Omega|$, θ_o, $u(A)$ and $u(B)$.

We conclude the introduction with some notations and an identity for Equation (1.1). Let Ω be a bounded convex smooth domain in the plane. We introduce curvilinear coordinate system (r, s), where s represents arc length along $\partial\Omega$ and $r(x_1, x_2)$ is the distance from a point $x = (x_1, x_2)$ in Ω to $\partial\Omega$. As in [11], we denotes $n = (n^1, n^2)$ the unit outward normal to $\partial\Omega$, $T = (T^1, T^2)$ is the unit tangent vector of $\partial\Omega$. The summation convention over repeated indices (from 1 to 2) will be employed. Assume that a function $u(x_1, x_2)$ is smooth in $\bar{\Omega}$, the following abbreviations will be adopted

$$u_1 = \frac{\partial u}{\partial x_1}, \qquad u_2 = \frac{\partial u}{\partial x_2}, \qquad u_{ij} = \frac{\partial^2 u}{\partial x_i \partial x_j}, \dots, \qquad v = 1 + |Du|^2.$$

Following [11], we define the normal derivative $\frac{\partial u}{\partial n}$ of u by

$$u_n = \lim_{r \to 0} \frac{1}{r}(u(x) - u(x - rn)) = u_i n^i.$$

On $\partial\Omega$ we can also define a tagential derivative $\frac{\partial u}{\partial s}$ of u by

$$u_s = u_i T^i.$$

Then we have the following formulas on $\partial\Omega$

$$\begin{aligned} u_{ss} &= \frac{\partial^2 u}{\partial s^2} = u_{ij} T^i T^j - k u_i n^i \\ u_{nn} &= \frac{\partial^2 u}{\partial n^2} = u_{ij} n^i n^j \\ u_{ns} &= \frac{\partial}{\partial s}(\frac{\partial u}{\partial n}) = u_{ij} n^i T^j + k u_s \\ u_{sn} &= u_{ns} - k u_s. \end{aligned} \tag{1.9}$$

Using curvilinear coordinate system, Equation (1.1) implies the following formula on $\partial\Omega$

$$u_{nn}(1 + u_s^2) = 2Hv^{\frac{3}{2}} - (u_{ss} + k u_n)v + (u_{ss} u_s^2 + 2 u_s u_n u_{ns} - k u_n u_s^2), \tag{1.10}$$

which will be used in Section 3 to prove Theorem 1.

Remark. The formula (1.4) is implicit contained in [**7**].

2. A minimum principle.

We consider the boundary value problem (1.1)-(1.2) in a strict convex bounded domain Ω in R^2 with smooth boundary $\partial\Omega$, and define the following function:

$$P^\alpha(x) = 2 - 2(1+|Du|^2)^{-\frac{1}{2}} - 2\alpha Hu. \tag{2.1}$$

We know that $P^\alpha(x)$ takes its maximum value at the critical point for $\alpha \geq 2$ [**6**], and on the boundary $\partial\Omega$ for $\alpha \leq 1$ [**7**]. We concentrate now our attention on $\alpha \in [1,2]$, and state the following:

Lemma 2.1 ([**7**, **8**]). *The function $P^\alpha(x)$ defined in* (2.1) *satisfies the following elliptic differential equation:*

$$\begin{aligned}(\delta_{ij} - u_i u_j v^{-1})P^\alpha_{ij} &- [2Hv^{-\frac{1}{2}}u_i + 2v^{-1}|Du|^{-2}u_k u_{ki} \\ &+ 2(\alpha-2)Hv^{-\frac{1}{2}}|Du|^{-2}u_i - 2v^{-2}|Du|^{-2}u_k u_l u_{kl} u_i]P^\alpha_i \\ &= 4(\alpha-1)(\alpha-2)H^2 v^{-\frac{1}{2}},\end{aligned} \tag{2.2}$$

where δ_{ij} is Kronecker symbol.

For the proof of Lemma 2.1, we make use of Definition (2.1) and of the following identity (valid in R^2 only):

$$u_{ij}u_i u_j |Du|^2 \equiv |Du|^2(\Delta u)^2 + 2u_i u_{ij} u_k u_{kj} - 2\Delta u u_i u_j u_{ij}.$$

The details of the computations are omitted here since they had been given in [**7**].

From Lemma 2.1 and Hopf maximum principle [**9**] we conclude that $P^\alpha(x)$ takes its minimum value either on the boundary of $\partial\Omega$, or at the unique critical point $C \in \Omega$ for $\alpha \in [1,2]$. For $\alpha > 1$, the second alternative had been rejected by Philippin [**7**]. The purpose of this section is to show even $\alpha = 1$ the second alternative can also be rejected unless $P^\alpha(x)$ is a constant in Ω. This can be achieved as a consequence of the following:

Theorem 2.2. *Let $u \in C^3(\bar\Omega)$ is a solution to* (1.1)-(1.2), *if*

$$P(x) = 2 - 2v^{-\frac{1}{2}} - 2Hu$$

attains its minimum at the unique critical point $C \in \Omega$, then $P(x)$ is a constant on $\bar\Omega$.

For the proof of the Theorem 2.2, we use the strong unique continuation of analytic function, so our program is to show all order derivatives of $P(x)$

are vanishing at $C \in \Omega$. To this end, we choose the origin of the coordinate axes at the critical point $C \in \Omega$, then

$$u_1(C) = u_2(C) = 0, \tag{2.3}$$

and orient the axes x_1 and x_2 in such a way that

$$u_{12}(C) = 0. \tag{2.4}$$

From Chen , Huang [**2**] and Sakaguchi [**10**], we know

$$u_{11}(C) > 0, \qquad u_{22}(C) > 0, \tag{2.5}$$

which will be used essential in the following proof.

Proof of Theorem 2.2. Our proof is divided four steps.

Step 1: We show the derivatives of $P(x)$ up to order 2 are vanishing at C.

First we compute the first derivaties of $P(x)$ at $C \in \Omega$. Since at any point $x \in \Omega$

$$P_1 = v^{-\frac{3}{2}} v_1 - 2Hu_1 = 2v^{-\frac{3}{2}} u_i u_{i1} - 2Hu_1, \tag{2.6}$$

$$P_2 = v^{-\frac{3}{2}} v_2 - 2Hu_2 = 2v^{-\frac{3}{2}} u_i u_{i2} - 2Hu_2, \tag{2.7}$$

then from (2.3), we have

$$P_1(C) = P_2(C) = 0. \tag{2.8}$$

Now we compute the second derivatives of $P(x)$ at C. From (2.3)-(2.6), we have at C

$$P_{11} = -\frac{3}{2} v^{-\frac{5}{2}} v_1^2 + v^{-\frac{3}{2}} v_{11} - 2Hu_{11} = 2u_{11}^2 - 2Hu_{11} \tag{2.9}$$

$$P_{12} = -\frac{3}{2} v^{-\frac{5}{2}} v_1 v_2 + v^{-\frac{3}{2}} v_{12} - 2Hu_{12} = 0 \tag{2.10}$$

$$P_{22} = -\frac{3}{2} v^{-\frac{5}{2}} v_2^2 + v^{-\frac{3}{2}} v_{22} - 2Hu_{22} = 2u_{22}^2 - 2Hu_{22}. \tag{2.11}$$

Use the fact that $P(x)$ attains its minimum at C, we have

$$P_{11}(C)P_{22}(C) - P_{12}^2(C) \geq 0. \tag{2.12}$$

From (2.5),(2.9) and (2.11) we know

$$u_{11}(C) = u_{22}(C) = H \tag{2.13}$$

$$P_{11}(C) = P_{22}(C) = 0. \tag{2.14}$$

Now we will use the induction to show that all order derivatives of $P(x)$ at C are vanishing.

Step 2: As a first step for induction, we will show the derivatives of $P(x)$ of order $3, 4$ at C are vanishing.

First we claim

$$\frac{\partial^3 P}{\partial x_1^k \partial x_2^{3-k}}(C) = 0, \quad k = 0, 1, 2, 3. \tag{2.15}$$

Using (2.9)-(2.11), (2.4) and (2.13) we have

$$P_{x_1^3}(C) = 4Hu^3_{x_1}(C) \tag{2.16}$$

$$P_{x_1^2 x_2}(C) = 4Hu_{x_1^2 x_2}(C) \tag{2.17}$$

$$P_{x_1 x_2^2}(C) = 4Hu_{x_1 x_2^2}(C) \tag{2.18}$$

$$P_{x_2^3}(C) = 4Hu_{x_2^3}(C). \tag{2.19}$$

Now, by differentiating (1.1), we obtain

$$u_{x_1^3}(C) = -u_{x_1 x_2^2}(C) \tag{2.20}$$

$$u_{x_1^2 x_2}(C) = -u_{x_2^3}(C). \tag{2.21}$$

To this end, use (2.8), (2.10), (2.14) and (2.20)-(2.21), we expand the function $P(x)$ in a Taylor series in a neighborhood of C:

(2.22)

$$\begin{aligned} P(x_1, x_2) - P(C) = \frac{r^3}{3!}\Big\{ &\frac{\partial^3 P}{\partial x_1^3}(C) \times [\cos^3\varphi - 3\cos\varphi\sin^2\varphi] \\ &+ \frac{\partial^3 P}{\partial x_1^2 \partial x_2}(C) \times [3\cos^2\varphi\sin\varphi - \sin^3\varphi]\Big\} + O(r^4), \end{aligned}$$

where (r, φ) are polar coordinates: $x_1 = r\cos\varphi, \quad x_2 = r\sin\varphi$. Suppose

$$\sqrt{P^2_{x_1^3}(C) + P^2_{x_1^2 x_2}(C)} \neq 0,$$

then $P(x)$ is not a constant, so we are lead to the following representation of $P(x)$ in a neighborhood of the point C:

$$P(x) - P(C) = A_3 \cos[3\varphi - \beta_3] r^3 + O(r^4), \tag{2.23}$$

with

$$A_3 = \frac{\sqrt{P^2_{x_1^3}(C) + P^2_{x_1^2 x_2}(C)}}{3!},$$

$$\cos\beta_3 = \frac{P_{x_1^3}(C)}{\sqrt{P^2_{x_1^3}(C) + P^2_{x_1^2 x_2}(C)}},$$

and

$$\sin\beta_3 = \frac{P_{x_1^2 x_2}(C)}{\sqrt{P^2_{x_1^3}(C) + P^2_{x_1^2 x_2}(C)}}.$$

From (2.23) we conclude that $P(x)$ has at least 3 nodal lines forming equal angles at the point C, using Lemma 2.1 we know that $P(x)$ attains its minimum only on $\partial\Omega$ or at the critical point C, which is a contradiction. Thus $A_3(C) = 0$ or

$$\frac{\partial^3 P}{\partial x_1^k \partial x_2^{3-k}}(C) = 0 \quad \text{and} \quad \frac{\partial^3 u}{\partial x_1^k \partial x_2^{3-k}}(C) = 0, \qquad k = 0, 1, 2, 3. \tag{2.24}$$

Use the similar argument we can show

$$\frac{\partial^4 P}{\partial x_1^k \partial x_2^{4-k}}(C) = 0, \quad k = 0, 1, 2, 3, 4 \tag{2.25}$$

and

$$u_{x_1^4}(C) = u_{x_2^4}(C) = 3H^3 \tag{2.26}$$

$$u_{x_1^2 x_2^2}(C) = H^3 \tag{2.27}$$

$$u_{x_1^3 x_2}(C) = u_{x_1 x_2^3}(C) = 0. \tag{2.28}$$

Step 3: Now we assume all order derivatives of $P(x)$ up to n are vanishing at C, where $n \geq 5$. Use similar argument as in Step 2 we have the following relations.

If $n = 2l, \quad l \geq 3$. Then

$$u_{x_1^m x_2^{k-m}}(C) = u_{x_1^{k-m} x_2^m}(C) \tag{2.29}$$
for any $\quad m = 0, 1, 2, \dots, k,$ if $\quad k = 5, 6, \dots, 2l,$

$$u_{x_1^m x_2^{k-m}}(C) = 0 \tag{2.30}$$
for any $\quad m = 0, 1, 2, \dots, k, \quad$ if $\quad k = 5, 7, 9, \dots, 2l - 1,$

$$u_{x_1^m x_2^{2p-m}}(C) = 0 \tag{2.31}$$
for any $\quad m = 1, 3, 5, \dots, 2p - 1,$ if $\quad p = 3, 4, 5, \dots, l,$

$$u_{x_1^{2p}}(C) = u_{x_2^{2p}}(C) = (2p-1)[(2p-3)(2p-5)\dots 1]^2 H^{2p-1} \tag{2.32}$$
for any $\quad p = 3, 4, \dots, l.$

When l is even, we obtain for any $p = 4, 6, \dots l$

$$u_{x_1^{2p}}(C) \div u_{x_1^{2p-2} x_2^2}(C) = (2p-1) \div 1 \tag{2.33}$$

$$u_{x_1^{2p-2} x_2^2}(C) \div u_{x_1^{2p-4} x_2^4}(C) = (2p-3) \div 3 \tag{2.34}$$

$$\vdots$$

$$u_{x_1^{p+2} x_2^{p-2}}(C) \div u_{x_1^p x_2^p}(C) = (p+1) \div (p-1), \tag{2.35}$$

and for any $p = 3, 5, 7, \dots l - 1$, we have

$$u_{x_1^{2p}}(C) \div u_{x_1^{2p-2} x_2^2}(C) = (2p-1) \div 1 \tag{2.36}$$

$$u_{x_1^{2p-2}x_2^2}(C) \div u_{x_1^{2p-4}x_2^4}(C) = (2p-3) \div 3 \tag{2.37}$$

$$\vdots$$

$$u_{x_1^{p+3}x_2^{p-3}}(C) \div u_{x_1^{p+1}x_2^{p-1}}(C) = (p+2) \div (p-2). \tag{2.38}$$

When l is odd, we have the similar relations (2.36)-(2.38).

If $n = 2l+1, \quad l \geq 2$, a similar argument show (2.29)-(2.38) and

$$u_{x_1^m x_2^{2l+1-m}}(C) = 0, \quad \text{for any } \ m = 0, 1, 2, \dots, 2l+1 \tag{2.39}$$

hold.

Step 4: Now we show the derivatives of $P(x)$ of order $n+1$ are vanishing at C. We divided it two parts according to whether n is odd or even.

Part A: If $n = 2l+1, \quad l \geq 2$, so $n+1 = 2(l+1)$ is even, we first look for the relations among $P_{x_1^m x_2^{n+1-m}}(C)$, where $m = 0, 2, 4, \dots n+1$. Through calculating, we have

$$P_{x_1^{n+1}}(C) = 2nH\left\{u_{x_1^{n+1}}(C) - (2l+1)[(2l-1)(2l-3)\dots 1]^2 H^{2l+1}\right\} \tag{2.40}$$

$$P_{x_1^{n-1}x_2^2}(C) = 2nH\left\{u_{x_1^{n-1}x_2^2}(C) - [(2l-1)(2l-3)\dots 1]^2 H^{2l+1}\right\}. \tag{2.41}$$

Now, by differentiating (1.1), we obtain

$$\frac{\partial}{\partial x_1^{n-1}}(\Delta u - u_i u_j u_{ij} v^{-1})(C) = \frac{\partial}{\partial x_1^{n-1}}(2Hv^{\frac{1}{2}})(C), \tag{2.42}$$

and using the values of derivatives of u up to order n at C, this lead to

$$u_{x_1^{n+1}}(C) + u_{x_1^{n-1}x_2^2}(C) = (n+1)[(2l-1)(2l-3)\dots 1]^2 H^{2l+1}. \tag{2.43}$$

From (2.40)-(2.41) and (2.43) we obtain

$$P_{x_1^{n+1}}(C) = -P_{x_1^{n-1}x_2^2}(C). \tag{2.44}$$

A similar argument, it follows that

$$P_{x_1^{n-1}x_2^2}(C) = -P_{x_1^{n-3}x_2^4}(C) = \dots = (-1)^l P_{x_2^{n+1}}(C). \tag{2.45}$$

Now we will find the similar relations (2.44)-(2.45) among

$$P_{x_1^m x_2^{n+1-m}}(C), \quad \text{where} \qquad m = 1, 3, 5, \dots n.$$

Using the same argument, we have

$$P_{x_1^{n+1-m}x_2^m}(C) = 2nHu_{x_1^{n+1-m}x_2^m}(C), \quad \text{where } \ m = 1, 3, 5, \dots, n. \tag{2.46}$$

Now, by differentiating (1.1), we obtain as in (2.43) the following relations

$$u_{x_1^{n+1-m}x_2^m}(C) = -u_{x_1^{n-1-m}x_2^{m+2}}(C), \quad \text{for } \ m = 1, 3, 5, \dots, n-2. \tag{2.47}$$

From (2.46)-(2.47), it follows that

$$P_{x_1^n x_2}(C) = -P_{x_1^{n-2} x_2^3}(C) = \cdots = (-1)^l P_{x_1 x_2^n}(C). \tag{2.48}$$

Up to now we are able to show the derivatives of $P(x)$ of order $n+1$ are vanishing at C as Step 2. Using the induction assumption, (2.44)-(2.45) and (2.48), we expand $P(x)$ in a Tayor's series in a neighborhood of the point C:

(2.49)

$$\begin{aligned}
&P(x) - P(C) \\
&= \frac{r^{n+1}}{(n+1)!}\Bigg\{P_{x_1^{n+1}}(C) \times \left[\binom{n+1}{0}\cos^{n+1}\varphi\right. \\
&\quad \left. - \binom{n+1}{2}\cos^{n-1}\varphi\sin^2\varphi + \cdots + (-1)^{l+1}\binom{n+1}{n+1}\sin^{n+1}\varphi\right] \\
&\quad + P_{x_1^{n-1}x_2}(C) \times \left[\binom{n+1}{1}\cos^n\varphi\sin\varphi - \binom{n+1}{3}\cos^{n-1}\varphi\sin^3\varphi\right. \\
&\quad \left. + \cdots + (-1)^l\binom{n+1}{n}\cos\varphi\sin^n\varphi\right]\Bigg\} + O(r^{n+2}).
\end{aligned}$$

As in Step 2, we can show the derivatives of $P(x)$ of order $n+1$ are vanishing at C.

Part B: If $n = 2l, \quad l \geq 3$, a similar argument as in Part A, we have

$$P_{x_1^{n+1-m} x_2^m}(C) = 2nHu_{x_1^{n+1-m} x_2^m}(C) \tag{2.50}$$

for $m = 0, 1, 2, 3, \ldots, n+1$. The same analysis as in Part A leads to imply the derivatives of $P(x)$ of order $n+1$ are vanishing at C.

According to the unique continuation of analytic function, we know if the function $P(x)$ attains its minimum at C, then it must be a constant, this establishes Theorem 2.2. □

Combination of Theorem 2.2 and Lemma 2.1 implies Theorem 2.

3. The proof of Theorem 1.

From Section 2, we know if $u \in C^3(\bar{\Omega})$ is a solution to Equations (1.1)-(1.2), then the function

$$P(x) = 2 - 2(1 + |Du|^2)^{-\frac{1}{2}} - 2Hu$$

attains its maximum [**7**] and minimum on $\partial\Omega$ unless $P(x)$ is a constant in $\bar{\Omega}$. As an application of these maximum and minimum principle of the function $P(x)$, in this section we get the size estimates of capillary free surface without gravity to complete the proof Theorem 1 and a Corollary. The proof of Theorem 1 will be divided two parts to show the different applications for maximum (minimum) principle.

Proof of Theorem 1. Part A: Use the fact that the function $P(x)$ attains its maximum on $\partial\Omega$, we first prove (1.3)-(1.4).

Assume $P(x)$ attains its maximum at $x_o \in \partial\Omega$. We must have at x_o:

$$\frac{1}{2}P_s = \frac{1}{2}(-2)\left(-\frac{1}{2}\right)v^{-\frac{3}{2}}(u_n^2+u_s^2+1)_s - Hu_s = v^{-\frac{3}{2}}(u_nu_{ns}+u_su_{ss}) - Hu_s = 0, \tag{3.1}$$

from boundary condition (1.2), we have

$$\sin^2\theta_o u_n^2 = \cos^2\theta_o + \cos^2\theta_o u_s^2,$$

it follows that

$$u_nu_{ns} = \cot^2\theta_o u_su_{ss}, \tag{3.2}$$

we conclude from (3.2) and (3.1) that

$$\frac{1}{2}P_s = v^{-\frac{3}{2}}(\cot^2\theta_o u_su_{ss} + u_su_{ss}) - Hu_s = u_s\left[\frac{u_{ss}}{\sin^2\theta_o v^{\frac{3}{2}}} - H\right] = 0. \tag{3.3}$$

According to Hopf maximum principle [**9**], we also have at x_o:

$$\frac{1}{2}P_n = v^{-\frac{3}{2}}[u_nu_{nn} + u_su_{sn}] - Hu_n > 0, \tag{3.4}$$

unless $P(x)$ is a constant on $\bar{\Omega}$.

If $u_s(x_o) \neq 0$, then from (3.3)

$$u_{ss}(x_o) = H\sin^2\theta_o v^{\frac{3}{2}}(x_o). \tag{3.5}$$

Now we shall use (1.2), (1.9), (1.10), (3.4) and (3.5) to lead

$$-ku_n|Du|^2 > 0, \qquad \text{at} \quad x_o, \tag{3.6}$$

which is contradiction to the strictly convexity of $\partial\Omega$. The proof of (3.6) is a long calculation. Using (3.5), at x_o, we shall rewrite (1.10) as

$$u_{nn}(1+u_s^2) = 2Hv^{\frac{3}{2}} + u_{ss}u_s^2 + 2\cot^2\theta_o u_s^2u_{ss} - ku_nu_s^2 - (u_{ss}+ku_n)(1+u_n^2+u_s^2). \tag{3.7}$$

Using (1.9) and (3.4) we have at x_o

$$u_nu_{nn} + u_s(u_{ns} - ku_s) - Hu_nv^{\frac{3}{2}} > 0,$$

thus

$$u_n^2(1+u_s^2)u_{nn} + u_nu_s(1+u_s^2)(u_{ns}-ku_s) - Hu_n^2(1+u_s^2)v^{\frac{3}{2}} > 0, \quad \text{at} \quad x_o. \tag{3.8}$$

Combining (3.7) with (3.8) yields at x_o

$$\begin{aligned} &u_n^2\left[2Hv^{\frac{3}{2}} + u_{ss}u_s^2 + 2\cot^2\theta_o u_s^2u_{ss} - ku_nu_s^2 - u_{ss}(1+u_n^2+u_s^2) - ku_nv\right] \\ &\quad + u_nu_s(1+u_s^2)u_{ns} - ku_nu_s^2(1+u_s^2) - Hu_n^2(1+u_s^2)v^{\frac{3}{2}} > 0. \end{aligned} \tag{3.9}$$

From (3.2), (3.9) can be rewritten as

$$\begin{aligned}
&\left[Hu_n^2 v^{\frac{3}{2}} + u_{ss}(\cot^2\theta_o u_s^2 u_n^2 - u_n^2 - u_n^4)\right] \\
&+ \left[u_{ss}u_s^2\cot^2\theta_o(1+u_n^2+u_s^2) - Hu_n^2u_s^2v^{\frac{3}{2}}\right] \\
&- ku_n\left[u_n^2u_s^2 + u_n^2 v + u_s^2 + u_s^4\right] > 0.
\end{aligned} \tag{3.10}$$

Now we calculate (3.10), from (1.2) and (3.5), it follows that at x_o

$$\begin{aligned}
&Hu_n^2v^{\frac{3}{2}} + u_{ss}u_n^2(\cot^2\theta_o u_s^2 - 1 - u_n^2) \\
&= Hu_n^2v^{\frac{3}{2}} - H\sin^2\theta_o v^{\frac{3}{2}}u_n^2\sin^2\theta_o = 0,
\end{aligned} \tag{3.11}$$

and

$$\begin{aligned}
&u_{ss}u_s^2\cot^2\theta_o(1+u_n^2+u_s^2) - Hu_n^2u_s^2v^{\frac{3}{2}} \\
&= H\sin^2\theta_o v^{\frac{3}{2}}\cot^2\theta_o u_s^2 v - Hu_n^2u_s^2v^{\frac{3}{2}} \\
&= Hu_s^2v^{\frac{3}{2}}(\cos^2\theta_o v - u_n^2) = 0,
\end{aligned} \tag{3.12}$$

similarly we have

$$-ku_n(u_n^2u_s^2 + u_s^2 + u_s^4 + u_n^2 v) = -ku_n(u_s^2 v + u_n^2 v) = -ku_n|Du|^2 v. \tag{3.13}$$

Insertion (3.11)-(3.13) into (3.10) yields

$$-ku_n|Du|^2 v > 0, \qquad \text{at} \quad x_o,$$

now we complete the proof (3.6).

Thus we must have $u_s(x_o) = 0$, from the expression for $P(x)$, x_o must be a point A where u attains its minimum on $\partial\Omega$ and we may use the fact that $P_{ss}(A) \le 0$ also. It follows that

$$0 \le u_{ss}(A) \le H\sin^2\theta_o v^{\frac{3}{2}}(A). \tag{3.14}$$

Using the similar calculation to get (3.6), we conclude from (1.10), (1.2), (1.9) and (3.4) that

$$k(A)\cos\theta_o < H - u_{ss}(A). \tag{3.15}$$

Insert (3.14) into (3.15) to find

$$k(A) < \frac{H}{\cos\theta_o}. \tag{3.16}$$

Moreover from the maximum principle we have $P(A) > P(C)$, it yields

$$u(A) - u(C) < \frac{1-\sin\theta_o}{H}. \tag{3.17}$$

If $P(x)$ is a constant on $\bar{\Omega}$ then a similar argument as (3.15) we have

$$k(x) \equiv \frac{H}{\cos\theta_o}, \qquad \text{for any} \quad x \in \partial\Omega, \tag{3.18}$$

$$u(x) - u(C) \equiv \frac{1 - \sin\theta_o}{H}, \qquad \text{for any} \quad x \in \partial\Omega. \tag{3.19}$$

Which imply if $P(x)$ is a constant on $\bar{\Omega}$ then Ω is a disk with radius $\dfrac{\cos\theta_o}{H}$. Until now we complete the proof of (1.3)-(1.4).

Conversely, if at $A \in \partial\Omega$ we have $u(A) - u(C) = \dfrac{1 - \sin\theta_o}{H}$ or $k(A) = \dfrac{H}{\cos\theta_o}$, from strong maximum principle it follows that $P(x)$ must be a constant on $\bar{\Omega}$, so (1.7)-(1.8) hold.

Part B: Similar to Part A, use the fact that $P(x)$ attains its minimum on $\partial\Omega$, we will prove (1.5)-(1.6).

From the boundary condition (1.2), we can see that the minimum of $P(x)$ on $\partial\Omega$ must be a point $B \in \partial\Omega$ where u itself is a maximum. It follows that

$$u_s(B) = 0, \qquad u_{ss}(B) \leq 0. \tag{3.20}$$

A similar argument as in Part A, we have

$$u(B) - u(C) \geq \frac{1 - \sin\theta_o}{H},$$
$$k(B) \geq \frac{H}{\cos\theta_o},$$

this is (1.5)-(1.6).

Conversely if $u(B) - u(C) = \dfrac{1 - \sin\theta_o}{H}$ or $k(B) = \dfrac{H}{\cos\theta_o}$, then from strong maximum and Theorem 2.2 $P(x)$ must be a constant on $\bar{\Omega}$, as in Part A (1.7)-(1.8) holds and Ω is a disk with radius $\dfrac{\cos\theta_o}{H}$.

When Ω is a disk of radius $\dfrac{\cos\theta_o}{H}$, (1.7)-(1.8) hold obviousily. Thus we have proved Theorem 1. ☐

As an another application of the minimum principle of Section 2, we prove the following Corollary for the capillary free surface area S defined as

$$S = \int_\Omega \sqrt{1 + |Du|^2}\, dx.$$

Corollary. *Let A and B as in Theorem 1, $V = \int_\Omega u\, dx$ is the volume of a liquid in a vertical tube, then S satisfies the inequalities:*

$$[\sin\theta_o + 3Hu(A)]|\Omega| - 3HV \leq S \leq [\sin\theta_o + 3Hu(B)]|\Omega| - 3HV. \tag{3.21}$$

Proof. From the fact that $P(x)$ attains its maximum at $A \in \partial\Omega$. We must have

$$P(x) \leq P(A) \qquad \text{for any} \quad x \in \Omega.$$

So we are actually have

$$\frac{1}{\sqrt{1+|Du|^2}} \geq \sin\theta_o + H[u(A)-u]. \tag{3.22}$$

Since

$$\frac{1}{\sqrt{1+|Du|^2}} = (1+|Du|^2)^{\frac{1}{2}} - \frac{|Du|^2}{\sqrt{1+|Du|^2}},$$

we obtain from (3.22)

$$H[u(A)-u] + \sin\theta_o \leq \sqrt{1+|Du|^2} - \frac{|Du|^2}{\sqrt{1+|Du|^2}}. \tag{3.23}$$

Using the fact that $\dfrac{|Du|^2}{\sqrt{1+|Du|^2}} = \dfrac{u_i u_i}{\sqrt{1+|Du|^2}}$ and the divergence theorem in conjuction with (1.1)-(1.2), we find from (3.23) after an intergration over Ω that

$$S \geq [3Hu(A) + \sin\theta_o]|\Omega| - 3HV. \tag{3.24}$$

Similar using the fact that $P(x)$ attains its minimum at $B \in \partial\Omega$, we know that

$$S \leq [3Hu(B) + \sin\theta_o]|\Omega| - 3HV. \tag{3.25}$$

Inequalities (3.24)-(3.25) are optimal in the sense that the equality signs in (3.24)-(3.25) holds if and only if Ω is a disk with radius $\dfrac{\cos\theta_o}{H}$. This establishes the Corollary. □

Acknowledgement. The author thanks to Prof. Shen Chun-Li and Prof. Zhou Qing for many stimulating discussions.

References

[1] J.-T. Chen, *Uniqueness of minimal point and its location of capillary free surfaces over convex domain*, Societe Mathematiques de France, Asterique, **118** (1984), 137-143.

[2] J.-T. Chen and Wu-H. Huang, *Convexity of capillary surfaces in the outer space*, Invent. Math., **67** (1982), 253-259.

[3] R. Finn, *Existence criteria for capillary free surfaces without gravity*, Indiana Univ. Math. Jour., **32** (1983), 439-460.

[4] ———, *Equilibrium capillary surfaces*, Springer, 1986.

[5] L. Nirenberg, *On nonlinear partial differential equations and Holder continuity*, Comm. Pure Appl. Math., **6** (1953), 103-156.

[6] L.E. Payne and G.A. Philippin, *Some applications of the maximum principle in the problem of torsional creep*, SIAM J. Appl. Math., **33** (1977), 446-455.

[7] ———, *Some maximum principles for nonlinear elliptic equations in divergence form with application to capillary surfaces and to surfaces of constant mean curvature*, Nonlinear Analysis TMA, **3** (1979), 193-211.

[8] G.A. Philippin, *A minimum principle for the problem of torsional creep*, Jour. Math. Anal. Appl., **68** (1979), 526-535.

[9] M.H. Protter and H.F. Weinberger, *Maximum principles in differential equations*, Prentice-Hall, 1967.

[10] S. Sakaguchi, *Uniqueness of critical point of the solution to the prescribed constant mean curvature equation over convex domain in* R^2, in 'Recent Topics on Nonlinear PDE IV', Kyoto, (1988), 129-151.

[11] R.P. Sperb, *Maximum principle and their applications*, Academic Press, 1981.

Received April 13, 1998.

East China Normal University
Shanghai
P.R. China
E-mail address: xnma@math.ecnu.edu.cn

PACIFIC JOURNAL OF MATHEMATICS
Vol. 192, No. 1, 2000

ON INFINITE UNRAMIFIED EXTENSIONS

Christian Maire

In this article we construct number fields k which have a trivial class group, but an infinite unramified extension.

1. Introduction.

Let k be a number field. A natural question is: Does k admit an infinite unramified extension?

The answer is no, if the root discriminant of k is less than Odlyzko's bounds.

The answer is yes, if k fails the test of Golod-Shafarevic for a prime number p. In that case, we know that there exists an infinite unramified p-extension L over k.

But generally it is fairly difficult to determin whether k admits an infinite unramified extension.

For this problem we introduce the following unramified extensions of k:

i) k_∞ is the maximal unramified Galois extension of k.

ii) $k^{(1)}$ denotes the Hilbert field of k, i.e., the maximal unramified abelian extension of k; its Galois group over k is isomorphic to the class group of k via the Artin map. More generally, let $k^{(i)}$ be the Hilbert field of $k^{(i-1)}$, $i \geq 1$, where $k^{(0)} = k$. Write $k_H = \cup k^{(i)}$; k_H is the Hilbert tower of k. We say that the Hilbert tower is finite (or stops) if $[k_H/k] < \infty$, or infinite otherwise.

iii) For a prime number p, k_p will be the p-Hilbert tower of k, that is to say the maximal p-extension of k contained in k_H.

We have the following inclusions:

$$k \subset k_p \subset k_H \subset k_\infty.$$

We recall two facts:

1) There exist fields k for which k_H is different from k_∞. For instance, the field $k = \mathbb{Q}(\sqrt{3.883})$ has a trivial class group, but there exists an unramified A_5-extension over k.

2) Yamamura [**7**] has shown that for all imaginary quadratic fields k with a discriminant less than 420 (or 729 under GRH), $k_\infty = k_H$, and k_∞/k is finite (the tower stops at the first, second or third floor).

Using some refinements of a theorem of Golod-Shafarevic or genus theory, it is possible to obtain sufficient criteria for the existence of an infinite unramified extension of a given number field k. However, these results generally imply that the Hilbert tower of the considered field k is infinite.

In Theorem 3.1 we give various examples of fields, such as the biquadratic field $\mathbb{Q}(\sqrt{17601097}, \sqrt{17380678572159893})$, with a trivial class group and an infinite unramified Galois extension: For each field k of Theorem 3.1, k_H/k is finite and k_∞/k is infinite.

Moreover, we prove that there exists infinitely many quadratic fields with a finite 2-Hilbert tower, together with an infinite unramified extension of degre 2^∞ (an extension of degre 2^∞ is an infinite extension which is the compositum of extensions of degree a power of 2).

Note that the calculations were performed using PARI [**1**].

Acknowlegements. The author wishes to thank X. Roblot for providing Example 4.1.2. He also thanks the referee for comments and for Example 3 in Remark 3.4.1.

2. Preliminaries.

In this section, we recall four propositions that will be used in the sequel.

Proposition 2.1 (Genus theory and theorem of Golod-Shafarevich, see [**6**])**.** *Let K/k be a cyclic extension of degree p, and ρ the number of places of k which are ramified in this extension. Denote by E_K (resp. E_k) the units group of K (resp. k).*

Suppose that

$$\rho \geq 3 + 2\sqrt{d_p E_K + 1} + d_p E_k,$$

then the p-Hilbert tower of K is infinite.

From then on, we shall write $r = \lceil 3 + 2\sqrt{d_p E_K + 1} + d_p E_k \rceil + \delta$, where $\lceil . \rceil$ is the ceiling function and where $\delta = 0$ or 1 depending on whether $3 + 2\sqrt{d_p E_K + 1} + d_p E_k$ is an integer or not.

Proposition 2.2 (Kummer theory)**.** *Let $k/\mathbb{Q}$ be a cyclic extension of degree p unramified at p, and $K/\mathbb{Q}$ an extension of degree n. Suppose that for all places $\mathfrak{Q}$ of K, the ramification index of $\mathfrak{Q}$ in $K/\mathbb{Q}$ divides the ramification index of $\mathfrak{q} = \mathbb{Q} \cap \mathfrak{Q}$ in $k/\mathbb{Q}$.*

Then the extension Kk/k is unramified at the finite places. In particular, if C is the Galois closure of K, then Ck/k is unramified at the finite places.

Note that in this case, a place $\mathfrak{Q}$ of K which is unramified over $\mathbb{Q}$ will ramifiy in Kk/K whenever $\mathfrak{Q} \cap \mathbb{Q}$ is ramified in $k/\mathbb{Q}$.

A consequence of the above is the following well-known result (see [**4**]):

Proposition 2.3. *Let P be an irreducible real polynomial, with squarefree discriminant $D(P)$. Denote by C the Galois closure of P. Then $C/\mathbb{Q}(\sqrt{\mathrm{Disc}\,(P)})$ is unramified.*

Proposition 2.4 (Class group of biquadratic fields)**.** *Let $k_1 = \mathbb{Q}(\sqrt{d_1})$ and $k_2 = \mathbb{Q}(\sqrt{d_2})$ be two disctincts quadratic fields. Let $k_3 = \mathbb{Q}(\sqrt{d_1.d_2})$ and $F = k_1 k_2$. Assume that:*

1) *F/k_3 is uramified at all places;*
2) *The class groups of k_1 and k_2 are trivial;*
3) *The class group of k_3 is cyclic of order 2.*

Then the Hilbert tower of k_3 stops at F; in particular, the class group of F is trivial.

Proof. It is known that for $p \neq 2$ (see [**5**])

$$|cl_F(p)| = |cl_{k_1}(p)||cl_{k_2}(p)||cl_{k_3}(p)| = 1,$$

where $cl_M(p)$ denotes the p-primary component of the class group of a number field M.

Moreover, the 2-Hilbert tower of k_3 is exactly F because $cl_{k_3}(2)$ is cyclic of order 2. In addition $|cl_F(2)| = 1$, and $|cl_F| = 1$. □

3. Main result.

We propose to construct *quadratic fields* with a finite Hilbert tower, such that they have an infinite unramified extension. Note that in all our examples the tower will stop at the first floor. This point will be proved with the help of Proposition 2.4.

3.1. Construction of infinite unramified extension. We consider the following situation:

1) K is a totally real field of degree n over $\mathbb{Q}$ such that:
   i) The discriminant of K is equal to a prime number l. Hence the Galois group of the closure of K is S_n (see [**4**]). Let P be a defining polynomial of K.
   ii) There exist $n-2$ unramified places above l in $K/\mathbb{Q}$.
2) q_1 and q_2 are two primes different from l such that:
   i) $q_1 \equiv q_2 \equiv 3(4)$,
   ii) q_1 splits completely in K,
   iii) there exist $n-1$ places above q_2 in $K/\mathbb{Q}$.
3) We write $k = \mathbb{Q}(\sqrt{l.q_1.q_2})$, and $M = Kk$.

Using Proposition 2.2, we have that all places above q_1 and q_2 plus $n-2$ places above l are ramified in M/K. Therefore, in the notation of Proposition 2.1, in the extension M/K we have

$$\rho = n + (n-1) + (n-2).$$

Applying Proposition 2.1 to M/K, we conclude that M has an infinite 2-Hilbert tower, denoted by M_2, if n is at least 7.

We have from Proposition 2.2 that the extension M/k is everywhere unramified. The above facts imply the following inclusions:

$$k = \mathbb{Q}(\sqrt{l.q_1.q_2}) \subset M \subset M_2,$$

where M_2/k is an infinite unramified extension. Taking the Galois closure of M_2 over k, we obtain an infinite unramified Galois extension L over k. In particular, k_∞/k is infinite.

3.2. The finiteness of k_H/k. Suppose that we have found primes l, q_1, q_2 and a field K of degree n satisfying all the hypotheses of paragraph 3.1. Let k_1, k_2, and $k = k_3$ be the quadratic fields of discriminant l, $l.q_1$, and $l.q_1.q_2$, respectively, and put $F = k_1k_2$. Since $l \equiv q_1q_2 \equiv 1 \pmod 4$, F/k_3 is unramified. If, in addition, k_1, k_2 have trivial class groups and k_3 has class number 2, then the Hilbert tower stops at F.

We remark that cl_F is trivial, and that the extension LF/F is an infinite unramified Galois extension, because F/k is unramified.

3.3. Finding a good polynomial. We want to find a polynomial of degree 7 verifying condition 1.i and 1.ii of paragraph 3.1. Such a polynomial does not exist in the PARI tables. Consequently, we use the following method:

Pick 7 reals around 0. Construct the monic polynomial P' of degree 7 whose roots are the considered reals, and take $P \in \mathbb{Z}[X]$ the nearest polynomial to P'. Then by choosing many families of 7 reals, we finally hope to find a good polynomial, i.e., such that:

i) $\mathrm{Disc}\,(P) = l$;

ii) $P \equiv (X-\alpha_1)^2 \prod_{i=2,6} (X-\alpha_i) \pmod l$.

Indeed, we rapidly obtain the polynomial

$$P(X) = X^7 - 3X^6 - 13X^5 + 28X^4 + 42X^3 - 47X^2 - 31X + 12,$$

with

$$l = \mathrm{Disc}\,(P) = 17380678572159893.$$

Conditions 1.i and 1.ii of paragraph 3.1 are verified.

3.4. Examples. We find one potential prime q_1 less than 300000 ($q_1 = 16747$), and 17 potential primes q_2 less than 100000.

Finally, eight of these primes q_2 are such that $(l, q_1 = 16747, q_2)$ satisfies the conditions of Proposition 2.4 (or paragraph 3.2).

Hence we have:

Theorem 3.1. *For the quadratic fields* $\mathbb{Q}(\sqrt{17380678572159893.16747.q_2})$, *with* $q_2 = 1051, 11863, 24659, 31583, 74527, 77339, 86579, 93491$, *one has:*

i) k_H/k *is finite;*

ii) k_∞/k *is infinite.*

3.4.1. Remarks.

1) If we choose q_2 such that P is totally decomposed modulo q_2, then q_2 is totally decomposed in the Galois closure of P, and so q_2 is decomposed in the extension $\mathbb{Q}(\sqrt{l})/\mathbb{Q}$. Consequently the class group of $\mathbb{Q}(\sqrt{l.q_1.q_2})$ contains the cyclic group (4) [**2**]. In particular, the Hilbert field of this field is not a biquadratic field so it would be more difficult to control its class group.

2) For each field of Theorem 3.1, there exists an unramified infinite extension L over F with Galois group G such that we have the extension sequence:

$$1 \longrightarrow H \longrightarrow G \longrightarrow A_7 \longrightarrow 1,$$

where H is a non-analytic pro-2-group [**3**]. Note that $G = [G, G]$.

As a result, the 2-rank of the class group of any nested sequence of extensions M of F contained in L tends to infinity (see [**3**]). One shall bear in mind that $cl_F = (1)$.

3) It would be interesting to find the field of smallest discriminant satisfying the hypotheses of Theorem 3.1. The referee gives the following example:

$$P(X) = X^7 + 9X^6 + 13X^5 - 57X^4 - 86X^3 + 120X^2 - 1,$$

with $q_1 = 20411$ and $q_2 = 787$.

This example has the desired properties and the biquadratic field F has slighty smaller discrimant than the field in the introduction.

4. A remark.

One may try to apply Proposition 2.3 in order to find fields satisfying the conditions of Theorem 3.1.

More precisely, let P be an irreducible totally real polynomial with degree n such that the discriminant of P is equal to $l_1.l_2.l_3$ with $l_1 \equiv 1(4)$, and $l_2 \equiv l_3 \equiv 3(4)$, where l_1, l_2, l_3 are primes.

Let $K = \mathbb{Q}(\theta)$, where θ is a root of P, $k = \mathbb{Q}(\sqrt{l_1.l_2.l_3})$, and $M = Kk$. Assume that l_1, l_2 and l_3 are sufficiently decomposed in $K/\mathbb{Q}$. Then for larger n, using Propositions 2.1 and 2.3 we obtain that the field $k = \mathbb{Q}(\sqrt{l_1.l_2.l_3})$ has an infinite unramified Galois extension.

To show that the Hilbert tower of this field stops we can use an argument similar to the one involved in the previous section.

Yet, the problem becomes: How to find a polynomial satisfying the hypotheses above.

4.1. A method. Let $l_1 \equiv 1(4)$ and $l_2 \equiv 3(4)$ be two primes, and put

$$P(X) = X^2 \prod_{i=1}^{n-2} (X - \alpha_i) \pm l_1.l_2,$$

where α_i are different integers. Changing α_i, l_1, and l_2, we hope to find some polynomials P such that:

1) P is irreducible;
2) P is totally real;
3) $\mathrm{Disc}\,(P) = l_1.l_2.l_3$, for some prime l_3;
4) P is sufficiently decomposed modulo l_3;
5) For $i = 2$ or 3, $\left(\dfrac{l_1}{l_i}\right) = -1$.

Note that the last condition is necessary to get $cl_{k_3} = (2)$ [**2**].

4.1.1. Case $n = 17$; $r = 32$.

For the definition of r, see the paragraph of following Proposition 2.1. In that case, we should have at least two places above l_3.

We have produced one polynomial:

$$\begin{aligned} P(X) = X^{17} &- 8X^{16} - 140X^{15} + 1120X^{14} + 7472X^{13} - 59696X^{12} \\ &- 191620X^{11} + 1532960X^{10} + 2475473X^{9} - 19803784X^{8} \\ &- 15291640X^{7} + 122333120X^{6} + 38402064X^{5} \\ &- 307216512X^{4} - 25401600X^{3} + 203212800X^{2} - 4819 \end{aligned}$$

with $\mathrm{Disc}\,(P) = l_1.l_2.l_3$, where $l_1 = 61$, $l_2 = 79$ and
$l_3 = 6638290055279332101085152678390469043164905703667099721264428979858829342643629881078901319043598924174484237159571111873134668406985800955450434299964685089696459$.

Note that $\log(l_3)/\log(10) \approx 163$, and that

$$\begin{aligned} P(X) = X^2&(X-1)(X-2)(X-3)(X-4)(X-5)(X-6) \\ &\cdot (X-7)(X-8)(X+1)(X+2)(X+3)(X+4)(X+5) \\ &\cdot (X+6)(X+7) - 61.79 \end{aligned}$$

There are 5 places above l_3 which are ramified in K/k. So $\rho = 15+15+5 = 35 > r$.

4.1.2. Case $n = 11$; $r = 24$.

X. Roblot comes up with the polynomial:

$P(X) = X^{11} - 45X^{10} + 870X^{9} - 9450X^{8} + 63273X^{7} - 269325X^{6} + 723680X^{5} - 1172700X^{4} + 1026576X^{3} - 362880X^{2} + 2483$
with $\mathrm{Disc}\,(P) = l_1.l_2.l_3$, where $l_1 = 13$, $l_2 = 191$ and
$l_3 = 1975697671490152075520432855935517362161188018903$.
Here $\rho = 9 + 9 + 7 = 25 > r$.

4.1.3. Case $n = 9$; $r = 21$.

We found three polynomials ; for all these cases $\rho = 7 + 7 + 7 = 21 \geq r$.

- $X^9 - 16X^8 - 122X^7 + 1100X^6 + 3709X^5 - 20524X^4 - 28068X^3 + 95760X^2 - 18467$
  with $l_1 = 313$, $l_2 = 59$, and
  $l_3 = 2625511568033074168604133553850114168723695438468422 7$.
- $X^9 + 35X^8 + 51X^7 - 1535X^6 - 2476X^5 + 18660X^4 + 17424X^3 - 69120X^2 + 13067$
  with $l_1 = 73$, $l_2 = 179$, and
  $l_3 = 7154637362578369142642781505295690026075933100681339$.
- $X^9 + 41X^8 + 69X^7 - 1805X^6 - 3046X^5 + 22044X^4 + 21096X^3 - 82080X^2 + 18511$
  with $l_1 = 173$, $l_2 = 107$, and
  $l_3 = 9538450397783460513336573925700848923329715205666 1447$.

The problem is to calculate the class group of the fields $\mathbb{Q}(\sqrt{l_1})$ and $\mathbb{Q}(\sqrt{l_1.l_2.l_3})$.

5. Infinite non Galois unramified extension.

In this section, we prove:

Theorem 5.1. *There exists infinitely many quadratic fields (imaginary and real) with a finite 2-Hilbert tower, but with an infinite unramified extension of degre 2^∞.*

Proof (only for the real case) and examples.

Let K be a totally real extension over $\mathbb{Q}$ of degree 8, such that Disc $(K) = l$ where l is prime number. Let q_1 and q_2 be two primes, satisfying:

1) $q_1 \equiv q_2 \equiv 3(4)$;
2) q_1 and q_2 are totally decomposed in $K/\mathbb{Q}$.

Put $N = K\mathbb{Q}(\sqrt{l})$.

Then q_1 and q_2 are totally decomposed in the Galois closure of P, in particular in $N/\mathbb{Q}$. This extension is of degree 16.

Using Proposition 2.1 applied to the extension $N(\sqrt{q_1.q_2})/N$, we conclude that $E = N(\sqrt{q_1.q_2})$ has an infinite 2-Hilbert tower, noted E_2.

Thus, we have the following inclusions:

$$\mathbb{Q}(\sqrt{l.q_1.q_2}) \subset \mathbb{Q}(\sqrt{l}, \sqrt{q_1.q_2}) \subset E \subset E_2,$$

where all extensions are unramified 2-extensions. Accordingly, $E_2/\mathbb{Q}(\sqrt{l.q_1.q_2})$ is an unramified extension of degree 2^∞.

Moreover, the choice of q_1 and q_2 implies that the 2-part of class group of $\mathbb{Q}(\sqrt{l.q_1.q_2})$ is cyclic. As consequence the 2-Hilbert tower stops at the first floor.

If $K/\mathbb{Q}$ exists, Cebotarev's density criterion asserts that there exist infinitely many such fields $\mathbb{Q}(\sqrt{l.q_1.q_2})$.

For the imaginary quadratic case, one may start with the same K and q_1 as above and take $q_2 = -1$. Then $\mathbb{Q}(\sqrt{l.q_1.q_2})$ is an imaginary quadratic field with the desired properties.

5.1. Examples. Consider the following polynomial, found in the PARI tables (E-mail: megrez.math.u-bordeaux.fr, directory pub/numberfields):

$$P(X) = X^8 - X^7 - 7X^6 + 5X^5 + 14X^4 - 6X^3 - 9X^2 + X + 1,$$

where

$$\mathrm{Disc}(P) = 1318279381.$$

Then the fields

$$\mathbb{Q}\left(\sqrt{1302839.4503991.1318279381}\right)$$

and

$$\mathbb{Q}\left(\sqrt{-643.1318279381}\right)$$

satisfy the conditions of Theorem 5.1.

References

[1] C. Batut, K. Belabas, D. Bernardi, H. Cohen and M. Olivier, *User's guide to PARI-GP*, 1997.

[2] G. Gras, *Sur les classes d'idéaux dans les extensions cycliques relatives de degré premier l*, Ann. Inst. Fourier, **23**(**4**) (1973), 1-48.

[3] F. Hajir, *On the growth of p-class groups in p-class field towers*, J. Algebra, **188** (1997), 256-271.

[4] T. Kondo, *Algebraic number fields with the discriminant equal to that of a quadratic number field*, J. Math. Soc. Japan, **42** (1995), 31-36.

[5] T. Kubota, *Über den Bizyklischen Biquadratischen Zahlkörper*, Nagoya Math. J., **10** (1955), 65-85.

[6] J. Martinet, *Tours de corps de classes et estimations de discriminants*, Invent. Math., **44** (1978), 65-73.

[7] K. Yamamura, *Maximal unramified extensions of imaginary quadratic number fields of small conductors*, preprint, 1997.

Received April 15, 1998 and revised June 2, 1998.

Laboratoire A2X - Université Bordeaux I
351, Cours de la Libération
33405 Talence Cedex
France
E-mail address: maire@math.u-bordeaux.fr

PACIFIC JOURNAL OF MATHEMATICS
Vol. 192, No. 1, 2000

FINITE QUOTIENTS OF THE ALGEBRAIC FUNDAMENTAL GROUP OF PROJECTIVE CURVES IN POSITIVE CHARACTERISTIC

Amílcar Pacheco and Katherine F. Stevenson

Let X be a smooth connected projective curve defined over an algebraically closed field k of characteristic $p > 0$. Let G be a finite group whose order is divisible by p. Suppose that G has a normal p-Sylow subgroup. We give a necessary and sufficient condition for G to be a quotient of the algebraic fundamental group $\pi_1(X)$ of X.

1. Introduction.

Let X be a smooth projective connected algebraic curve of genus g defined over an algebraically closed field k of characteristic $p > 0$. In this paper we study necessary and sufficient conditions for a finite group G to be a quotient of the algebraic fundamental group $\pi_1(X)$ of X. We denote by $\pi_A(X)$ the set of isomorphism classes of finite groups which are quotients of $\pi_1(X)$. Recall that a group $G \in \pi_A(X)$ will occur as a Galois group of an étale Galois cover $Z \to X$. In this paper we will call $Z \to X$ a Galois G-cover.

Let G be a finite group and suppose that its order is not divisible by p. In [**Groth71**, Corollary 2.12] Grothendieck showed that $G \in \pi_A(X)$ if and only if G is a quotient of the topological fundamental group Γ_g of a compact Riemann surface of genus g.

We consider next a finite p-group G. Denote by $\Phi(G) = [G,G]G^p$ its Frattini subgroup and let $\mathcal{G} = G/\Phi(G)$. This group is an elementary p-abelian group. The p-torsion subgroup $J_X[p]$ of the Jacobian variety J_X of X is an $\mathbb{F}_p$-vector space whose dimension γ_X is called the Hasse-Witt invariant of X. It follows from [**Ser56**, §11] that $\mathcal{G} \in \pi_A(X)$ if and only if $\mathcal{G}$ has p-rank at most γ_X. Suppose now that $G \in \pi_A(X)$, then $\mathcal{G} \in \pi_A(X)$, therefore the p-rank of G (the minimal number of generators of its maximal p-quotient) is at most γ_X. Actually, this condition is also sufficient. This follows from the fact that the p-cohomological dimension $\mathrm{cd}_p(\pi_1(X))$ of $\pi_1(X)$ is at most 1 (cf. end of proof of Theorem 1.3).

Now these two situations are understood, the next step to study is the case of a finite group G whose order is divisible by p. Consider the case where G has a normal p-Sylow subgroup P. Let $H = G/P$. The main theorem (Theorem 1.3) addresses the question of when a Galois P-cover $Z \to Y$ and a Galois H-cover $Y \to X$ can be composed to give a Galois G-cover $Z \to X$ (recall that in general the cover may not be Galois). Roughly speaking the theorem says that $G \in \pi_A(X)$ if and only if the action of H on P is compatible with the action of H on $J_Y[p]$. The result fits nicely with the fact (implied above) that the p-torsion of the Jacobian variety of Y regulates the Galois P-covers of Y. In order to state the main theorem precisely we need to introduce some notation.

1.1. Group theory. Let G be a finite group with normal p-Sylow subgroup P and quotient $H = G/P$. A theorem of Schur and Zassenhaus assures that G is isomorphic to the semi-direct product $P \rtimes H$ taken with respect to the the action $\eta : H \to \mathrm{Aut}(P)$ defined by conjugation. Let $\Phi(P) = [P, P]P^p$ be the Frattini subgroup of P. The quotient $\mathcal{P} = P/\Phi(P)$ is the maximal elementary abelian quotient of P, hence it is an $\mathbb{F}_p$-vector space. This action induces an $\mathbb{F}_p$-representation $\rho : H \to \mathrm{Aut}(\mathcal{P})$.

Let $Z(H)$ be the set of irreducible characters χ of H with values in the algebraically closed field k of characteristic $p > 0$. Let χ^0 be the trivial character of H and $\rho_\chi : H \to \mathrm{GL}(V_\chi)$ an irreducible k-representation of H of character χ of degree n_χ. The canonical decomposition of $\mathcal{P} \otimes_{\mathbb{F}_p} k$ as a $k[H]$-module is given by

$$\mathcal{P} \otimes_{\mathbb{F}_p} k = \bigoplus_{\chi \in Z(H)} V_\chi^{m_\chi}. \tag{1.1}$$

1.2. Generalized Hasse-Witt invariants. Let $Y \to X$ be a Galois cover with $\mathrm{Gal}(Y/X) \cong H$ and g_Y the genus of Y. Let J_Y be the Jacobian variety of Y and $J_Y[p]$ its p-torsion subgroup. Suppose that $\mathbb{F}_q = \mathbb{F}_{p^m}$ is a finite field large enough to contain the $|H|$-th roots of unity. Let $e_\chi = \frac{\chi(1)}{|H|} \sum_{h \in H} \chi\left(h^{-1}\right) h \in k[H]$ be the idempotent corresponding to χ. Denote by

$$J_Y[p] \otimes_{\mathbb{F}_p} \mathbb{F}_q = \bigoplus_{\chi \in Z(H)} J_Y[p]_\chi \tag{1.2}$$

the canonical decomposition of $J_Y[p] \otimes_{\mathbb{F}_p} \mathbb{F}_q$, where $J_Y[p]_\chi = e_\chi \cdot (J_Y[p] \otimes_{\mathbb{F}_p} \mathbb{F}_q)$.

Definition 1.1. The generalized Hasse-Witt invariant $\gamma_{Y,\chi}$ of Y with respect to χ is defined as the dimension of $J_Y[p]_\chi$ as an $\mathbb{F}_q$-vector space (cf. [**Ruc86**, §2]). A surjection $\phi : \pi_1(X) \twoheadrightarrow H$ corresponds to a Galois H-cover $Y \to X$, and in the case that the cover $Y \to X$ is not named, we will use $\gamma_{\phi,\chi}$ to denote the generalized Hasse-Witt invariant $\gamma_{Y,\chi}$ of Y with respect to χ.

The notation $\gamma_{\phi,\chi}$ has the advantage that it emphasizes that the generalized Hasse-Witt invariants are invariants of the *cover* $Y \to X$ (corresponding to the surjection $\phi : \pi_1(X) \twoheadrightarrow H$) rather than of the curve Y alone. Also, the main result of this paper is phrased in terms of embedding problems involving ϕ. Thus, the notation $\gamma_{\phi,\chi}$ eases the exposition in that case. However, in the literature the notation $\gamma_{Y,\chi}$ is standard. Moreover, in this paper when we are dealing directly with the cover $Y \to X$, as opposed to the surjection ϕ, we use the notation $\gamma_{Y,\chi}$.

A consequence of (1.2) is

$$\gamma_Y = \sum_{\chi \in Z(H)} \gamma_{Y,\chi}. \tag{1.3}$$

1.3. Embedding problems.

Definition 1.2 ([**Har95**, p. 366])**.** An embedding problem for a profinite group Λ is a pair of surjective profinite group homomorphisms $(\alpha : \Lambda \to \mathcal{K}_2, \delta : E \to \mathcal{K}_2)$. The embedding problem is finite, if E is a finite group, and trivial, if δ is an isomorphism. A weak, respectively proper, solution to the embedding problem is a homomorphism, respectively a surjective homomorphism, $\beta : \Lambda \to E$ such that $\alpha = \delta \circ \beta$.

1.4. Main theorem. Here we give the statement of the main result. Since a surjection $\phi : \pi_1(X) \twoheadrightarrow H$ corresponds to a unique Galois H-cover $Y \to X$, embedding problems $(\phi : \pi_1(X) \twoheadrightarrow H, G \twoheadrightarrow H)$ relate to Galois theory. Specifically, a proper solution to such an embedding problem corresponds to the existence of a Galois G-cover $Z \to X$ dominating the Galois H-cover $Y \to X$. Thus we use the language of embedding problems to state the main theorem.

Again we assume throughout this paper that all curves are smooth connected projective k-curves. For such a curve X we make the following notation.

Notation. Given a group G with normal p-Sylow subgroup P and quotient $H = G/P$, and given $\phi : \pi_1(X) \twoheadrightarrow H$, let m_χ, n_χ, and $\gamma_{\phi,\chi}$ be as in Sections 1.1 and 1.2. By *Condition A for the curve X* we will mean that for every $\chi \in Z(H)$ the following inequality holds: $m_\chi n_\chi \leq \gamma_{\phi,\chi}$.

Theorem 1.3. *Let G be a finite group having a normal p-Sylow subgroup P. Let $H = G/P$. An embedding problem $(\phi : \pi_1(X) \twoheadrightarrow H, G \twoheadrightarrow H)$ has a proper solution if and only if Condition A holds for the curve X.*

The necessity of Condition A for the curve X in Theorem 1.3 was previously obtained in [**Ste96a**, Proposition 3.4]. In this paper we show that it is also sufficient. Rephrasing this in terms of covers we obtain the following immediate corollary.

Corollary 1.4. *Let G be a finite group having a normal p-Sylow subgroup P. Let $H = G/P$. Then, $G \in \pi_A(X)$ if and only if there exists a Galois H-cover $Y \to X$ such that $m_\chi n_\chi \leq \gamma_{Y,\chi}$, for every $\chi \in Z(H)$.*

Remark 1.5. In the case that ϕ corresponds to a Galois H-cover $Y \to X$ where Y is an ordinary curve (namely, that the genus of Y is equal to γ_Y) we show (Theorem 7.1) that an embedding problem $(\phi : \pi_1(X) \twoheadrightarrow H, G \twoheadrightarrow H)$ has a proper solution if and only if $m_\chi \leq g$, when χ is the trivial character of H, and $m_\chi \leq (g-1)n_\chi$, otherwise. The advantage here is that we eliminate the generalized Hasse-Witt invariant notation from the condition. This result appears in Section 7 where we discuss it and other consequences of Theorem 1.3. The existence or non-existence of 'ordinary Galois H-covers' is a difficult and open problem. However, in the case that H is abelian and X is 'generic' (cf. Section 7) a great deal of progress has been made by Nakajima [**Nak83**] and Zhang [**Zha92**]. We use their theorems in Section 7 to obtain some interesting results and examples (cf. Theorem 7.4 and Example 7.11).

We start with some Preliminaries which allow us to compute the generalized Hasse-Witt invariants in terms of differentials and to estimate how big they are. Next, in Section 3, we determine when $\mathcal{P} \rtimes H \in \pi_A(X)$ and develop some elementary representation theory tools which will be used in Section 6 to prove Theorem 1.3. In Section 4, some useful results regarding solutions of embedding problems are given. In Section 5, we prove that the p-cohomological dimension of $\pi_1(X)$ is at most 1. The main result is proved in Section 6, and in Section 7 we discuss some consequences of the main theorem and make some comparisons to previous work of Nakajima [**Nak87**] and Stevenson [**Ste96a**].

2. Preliminaries.

Let Y be a smooth projective connected algebraic curve of genus g_Y defined over an algebraically closed field k of characteristic $p > 0$.

Definition 2.1. Let Ω^1_Y be the space of differentials of Y and $\Omega^1_Y(0) \subset \Omega^1_Y$ the subspace of regular differentials. Let L be the function field of Y and t a separating variable of L. Given $\omega = f\,dt \in \Omega^1_Y$, the Cartier operator is defined by $\mathcal{C}(\omega) = (-d^{p-1}f/dt^{p-1})^{1/p}\,dt$. This is a $1/p$-linear operator, i.e., $\mathcal{C}(a^p\omega) = a\mathcal{C}(\omega)$, for any $a \in K$. Moreover, $\mathcal{C}$ acts on $\Omega^1_Y(0)$ (cf. [**Ser56**, §10, p. 39]).

It also follows from [**Ser56**, §10, p. 39] that there exists an $\mathbb{F}_p$-isomorphism between $J_Y[p]$ and $\mathrm{Ker}(1 - \mathcal{C} \,|\, \Omega^1_Y(0))$ given by $\mathrm{class}(D) \mapsto df/f$, where $p.\mathrm{class}(D) = \mathrm{div}(f)$. In particular, $\gamma_Y \leq g_Y$.

Definition 2.2. The curve Y is called ordinary if $\gamma_Y = g_Y$.

In order to understand how big the generalized Hasse-Witt invariants are we recall that a theorem of Nakajima ([**Nak84**] Corollary, one can also follow the proof in characteristic 0 of Chevalley and Weil [**CheWei34**]) which says that if $Y \to X$ is étale and $\mathrm{Gal}(Y/X) \cong H$, then we have an isomorphism of $k[H]$-modules

$$\Omega^1_Y(0) \cong k \oplus k[H]^{g-1}. \tag{2.1}$$

Given $\chi \in Z(H)$, let $\Omega^1_Y(0)_\chi = e_\chi.\Omega^1_Y(0)$ and $g_\chi = \dim_k \Omega^1_Y(0)_\chi$. Note that (2.1) implies that $g_{\chi^0} = g$ and $g_\chi = (g-1)n_\chi^2$ for every $\chi \in Z(H)$, $\chi \neq \chi^0$. It is a result due to Rück [**Ruc86**, Proposition 2.3] that the $\mathbb{F}_q[H]$-modules $J_Y[p]_\chi$ and $\mathrm{Ker}(1 - \mathcal{C}^m \mid \Omega^1_Y(0)_\chi)$ are isomorphic (this generalizes the above result of Serre). Hence, for each $\chi \in Z(H)$ we have

$$\gamma_{Y,\chi} \leq g_\chi. \tag{2.2}$$

Remark 2.3. In particular, by (1.3), we conclude that Y is ordinary if and only if for each $\chi \in Z(H)$ we have

$$\gamma_{Y,\chi} = \begin{cases} g, & \text{if } \chi = \chi^0 \text{ and} \\ (g-1)\, n_\chi^2, & \text{if } \chi \neq \chi^0. \end{cases} \tag{2.3}$$

3. Unramified covers and Galois modules.

Let $Y \to X$ be a Galois cover with $\mathrm{Gal}(Y/X) \cong H$ and $Z \to Y$ an étale Galois cover with $\mathrm{Gal}(Z/Y) \cong (\mathbb{Z}/p\mathbb{Z})^r$, for some $1 \leq r \leq \gamma_Y$. In [**Pac95**, Propositions 2.4 and 2.5] the first author determined a necessary and sufficient condition for $Z \to X$ to be also Galois. We review these results and as a consequence we obtain a necessary and sufficient condition for $\mathcal{P} \rtimes H \in \pi_A(X)$.

Denote by $\mathcal{S}_1$ the set of all étale Galois covers $Z \to Y$ with $\mathrm{Gal}(Z/Y) \cong (\mathbb{Z}/p\mathbb{Z})^r$ for some $1 \leq r \leq \gamma_Y$. This set corresponds bijectively to the set $\mathcal{S}_2$ of $\mathbb{F}_p$-vector subspaces of $\mathrm{Hom}(\pi_1(Y), \mathbb{Z}/p\mathbb{Z})$ by $(Z \to Y) \mapsto \mathrm{Hom}(\mathrm{Gal}(Z/Y), \mathbb{Z}/p\mathbb{Z})$, where we identify $\mathrm{Hom}(\mathrm{Gal}(Z/Y), \mathbb{Z}/p\mathbb{Z})$ with the $\mathbb{F}_p$-vector space of $\psi \in \mathrm{Hom}(\pi_1(Y), \mathbb{Z}/p\mathbb{Z})$ such that $\pi_1(Z) \subset \mathrm{Ker}(\psi)$. Its inverse is equal to $V \mapsto (Z \to Y)$, where $\bigcap_{\psi \in V} (L^{\mathrm{un}})^{\mathrm{Ker}(\psi)}$ is the function field of Z, L is the function field of Y and L^{un} is the maximal unramified Galois extension of L. An element $(Z \to Y)$ of $\mathcal{S}_1$ is explicitly described as follows.

For each $Q \in Y$, let L_Q be the completion of L at Q, $U_L = \bigcap_{Q \in Y} (\wp(L_Q) \cap L)$, where $\wp$ denotes the operator $\wp(x) = x^p - x$. Let $W_L = U_L/\wp(L)$ and for each $a \in U_L - \wp(L)$, let $\langle a + \wp(L) \rangle$ be the cyclic subgroup of order p of W_L generated by $a + \wp(L)$. Denote by $\wp^{-1}(a)$ a solution of $\wp(T) = a$ in the algebraic closure of L.

Lemma 3.1 ([**Pac95**, Proposition 2.4])**.** *Let $(Z \to Y) \in \mathcal{S}_1$ with $\mathrm{Gal}(Z/Y) \cong (\mathbb{Z}/p\mathbb{Z})^r$ for some $1 \leq r \leq \gamma_Y$. There exist $\mathbb{F}_p$-linearly independent $a_1 + \wp(L), \ldots, a_r + \wp(L) \in W_L$ such that $k(Z) = k(\wp^{-1}(a_1), \cdots, \wp^{-1}(a_r))$. Moreover, the cover $(Z \to Y)$ is uniquely determined by the $\mathbb{F}_p$-vector subspace $A_{Z/Y} = \bigoplus_{j=1}^r \langle a_j + \wp(L) \rangle$ of W_L and $\mathrm{Gal}(Z/Y) = \mathrm{Hom}(A_{Z/Y}, \mathbb{Z}/p\mathbb{Z})$.*

Lemma 3.2 ([**Pac95**, Proposition 2.5])**.** *With hypothesis and notation as in Lemma* 3.1*, $Z \to X$ is Galois if and only if $A_{Z/Y}$ is an $\mathbb{F}_p[H]$-module. In this case,* $\mathrm{Gal}(Z/X) \cong \mathrm{Gal}(Z/Y) \rtimes H$ *and the action of H on $A_{Z/Y}$ is contragradient to the natural action of H on* $\mathrm{Gal}(Z/Y)$*.*

Our goal now is to describe the $\mathbb{F}_p[H]$-module structure of $\mathcal{P}$ and compare it with the $\mathbb{F}_p[H]$-module structure of $\mathrm{Ker}(1-\mathcal{C} \mid \Omega^1_Y(0))$. In order to do this we introduce some basic facts on representation theory.

Definition 3.3. Let $\chi \in Z(H)$ and denote by $\rho_\chi : H \to \mathrm{GL}(V_\chi)$ an irreducible representation with character χ. Given $h \in H$, let $(a_{ij}(h))$ be the matrix of $\rho_\chi(h)$ with respect to some fixed basis of V_χ. For each $m \geq 0$, let $\rho_{\chi^{p^m}} : H \to \mathrm{GL}(V_\chi)$ be the map defined by $\rho_{\chi^{p^m}}(h) = \rho_\chi(h)^{p^m}$.

Lemma 3.4 ([**Isa76**, p. 151])**.** *The map $\rho_{\chi^{p^m}}$ is an irreducible k-representation of H with character χ^{p^m} defined by $\chi^{p^m}(h) = \chi(h)^{p^m}$.*

Definition 3.5 ([**Isa76**, p. 152])**.** Denote by $\mathbb{F}_{p^{l_\chi}}$ the field $\mathbb{F}_p(\chi)$ generated by $\mathbb{F}_p$ and the character values $\{\chi(h)\,;\, h \in H\}$. Given $\chi, \psi \in Z(H)$, define $\chi \sim \psi$ if and only if there exists $0 \leq m < l_\chi$ such that $\psi = \chi^{p^m}$. Let $[\chi]$ be the class of χ in $\mathcal{Z}(H) = Z(H)/\sim$. Let $\mathcal{F}$ be the set of $\mathbb{F}_p$- irreducible representations $\rho : H \to \mathrm{GL}(U)$ of H.

Lemma 3.6 ([**Isa76**, Theorem 9.21])**.** *There is a bijection between the sets $\mathcal{F}$ and $\mathcal{Z}(H)$ given by $\rho \longmapsto [\chi]$, where $\rho \otimes_{\mathbb{F}_p} k : H \to \mathrm{GL}(U \otimes_{\mathbb{F}_p} k)$ is isomorphic to $\rho_{[\chi]} = \bigoplus_{j=0}^{l_\chi - 1} \rho_{\chi^{p^j}}$.*

The action $\eta : H \to \mathrm{Aut}(P)$ given by conjugation induces an $\mathbb{F}_p$-representation $\rho : H \to \mathrm{Aut}(\mathcal{P})$. By Lemma 3.6, $\rho \otimes_{\mathbb{F}_p} k$ is a sum of the representations $\rho_{[\chi]}$ with multiplicities m_χ (note that since ρ is defined over $\mathbb{F}_p$, $m_\psi = m_\chi$, for $\psi \sim \chi$). Denote $V_{[\chi]} = \bigoplus_{j=0}^{l_\chi-1} V_{\chi^{p^j}}$. Hence,

$$\mathcal{P} \otimes_{\mathbb{F}_p} k \cong \bigoplus_{[\chi] \in \mathcal{Z}(H)} V_{[\chi]}^{m_\chi}. \tag{3.1}$$

Let $\mathcal{V}_{[\chi]}$ be the irreducible $\mathbb{F}_p[H]$-module such that $\mathcal{V}_{[\chi]} \otimes_{\mathbb{F}_p} k \cong V_{[\chi]}$. It follows from (3.1) that

$$\mathcal{P} \cong \bigoplus_{[\chi] \in \mathcal{Z}(H)} \mathcal{V}_{[\chi]}^{m_\chi}. \tag{3.2}$$

Let $n_\chi = \dim_k V_\chi$, $g_\chi = \dim_k \Omega^1_Y(0)_\chi$ and $\Omega^1_Y(0)_{[\chi]} = \bigoplus_{i=0}^{l_\chi-1} \Omega^1_Y(0)_{\chi^{p^i}}$. The Cartier operator $\mathcal{C}$ induces a k-isomorphism between $\Omega^1_Y(0)_{\chi^p}$ and $\Omega^1_Y(0)_\chi$ given by $\omega \mapsto \mathcal{C}(\omega)$. In particular, $\Omega^1_Y(0)_{[\chi]} \cong V_{[\chi]}^{g_\chi/n_\chi}$. Clearly $\mathcal{C}$ acts on $\Omega^1_Y(0)_{[\chi]}$. Hence, $\mathrm{Ker}(1-\mathcal{C} \mid \Omega^1_Y(0)_{[\chi]}) \cong \mathcal{V}_{[\chi]}^{t_\chi}$, for some $1 \leq t_\chi \leq g_\chi/n_\chi$.

The canonical decomposition of $\Omega^1_Y(0)$ into irreducible $k[H]$-modules is given by

$$\Omega^1_Y(0) = \bigoplus_{\chi\in Z(H)} \Omega^1_Y(0)_\chi = \bigoplus_{[\chi]\in\mathcal{Z}(H)} \Omega^1_Y(0)_{[\chi]}.$$

As a consequence we obtain the canonical decomposition

$$\begin{aligned}\text{(3.3)}\qquad \operatorname{Ker}(1-\mathcal{C}\mid\Omega^1_Y(0)) &= \bigoplus_{[\chi]\in\mathcal{Z}(H)} \operatorname{Ker}(1-\mathcal{C}\mid\Omega^1_Y(0)_{[\chi]})\\ &\cong \bigoplus_{[\chi]\in\mathcal{Z}(H)} \mathcal{V}_{[\chi]}^{t_\chi}\end{aligned}$$

of $\operatorname{Ker}(1-\mathcal{C}\mid\Omega^1_Y(0))$ into irreducible $\mathbb{F}_p[H]$-modules.

4. Cohomological dimension and embedding problems.

In this section we describe one tool from Galois cohomology which we use to prove that if $\mathcal{P}\rtimes H\in\pi_A(X)$ and $\operatorname{cd}_p(\pi_1(X))\le 1$, then $G\in\pi_A(X)$. This result is expressed in terms of embedding problems (cf. Remark 4.4). This concept is also reviewed here.

Definition 4.1 ([**Ser86**, I-17])**.** A profinite group Λ has p-cohomological dimension at most $d\ge 1$, if for every Λ-module M and for every integer $e>d$ the p-primary component of $H^e(\Lambda,M)$ is trivial. The infimum $\operatorname{cd}_p(\Lambda)$ of all such d is called the p-cohomological dimension of Λ.

Definition 4.2 ([**Ser86**, I-23, 3.4])**.** Let

$$\text{(4.1)}\qquad 1\to\mathcal{K}_1\to E\xrightarrow{\delta}\mathcal{K}_2\to 1$$

be an extension of profinite groups. A profinite group Λ has the lifting property for this extension, if for every homomorphism $\alpha:\Lambda\to\mathcal{K}_2$ there exists a homomorphism $\beta:\Lambda\to E$ such that $\alpha=\delta\circ\beta$.

Proposition 4.3 ([**Ser86**, Proposition 16, I-23])**.** *The inequality* $\operatorname{cd}_p(\Lambda)\le 1$ *holds if and only if the extension* (4.1) *has the lifting property, when* $\mathcal{K}_1$ *is a pro-p group.*

Remark 4.4. In the case where $\operatorname{cd}_p(\Lambda)\le 1$, it follows from Proposition 4.3 and Definition 1.2 that there exists a weak solution to the embedding problem

$$(\delta:E\twoheadrightarrow\mathcal{K}_2,\Lambda\twoheadrightarrow\mathcal{K}_2).$$

Let G be a finite group having a normal p-Sylow subgroup P, $H=G/P$ and $\mathcal{P}=P/\Phi(P)$. Recall that $G\cong P\rtimes H$. Define $\delta_G:G\to\mathcal{P}\rtimes H$ by $\delta_G((a,b))=(a \bmod \Phi(P),b)$. This function is a surjective group homomorphism and $\operatorname{Ker}(\delta_G)=\Phi(P)$.

In particular, if $\mathrm{cd}_p(\pi_1(X)) \leq 1$ and $\mathcal{P} \rtimes H \in \pi_A(X)$, then there exists a weak solution $\pi_1(X) \to G$ to the embedding problem

$$(\delta_G : G \twoheadrightarrow \mathcal{P} \rtimes H, \pi_1(X) \twoheadrightarrow \mathcal{P} \rtimes H).$$

Furthermore, this weak solution is indeed a proper one, because $\Phi(P) \subset \Phi(G)$ and the latter set is exactly the set of "non-generators" of G, thus $\pi_1(X) \to G$ must be surjective.

5. Cohomological dimension at most one.

In this section we prove that the p-cohomological dimension $\pi_1(X)$ is at most 1. The proof follows the argument sketched out by Serre in [**Ser90**, Proposition 1] where he proves a similar result for an affine curve U (sf. also [**Kat88**]).

Definition 5.1. Let X be a smooth projective connected curve defined over k. Denote by $\mathbf{FEt}/X$ the category of finite étale covers of X. Given a closed point $\overline{x}$ of X define the functor $\mathfrak{F} : \mathbf{FEt}/X \to \mathbf{Sets}$ by $Y \mapsto \mathrm{Hom}_X(\overline{x}, Y)$.

Remark 5.2. It follows from [**Mil80**, Chapter I, §5, p. 39] that $\mathfrak{F}$ is strictly pro-representable, i.e., there exists a projective system $(X_\nu, \phi_{\nu\mu})$ in $\mathbf{FEt}/X$ where the transition morphisms $\phi_{\nu\mu} : X_\nu \to X_\mu$ are epimorphisms for $\nu \geq \mu$ and the elements $f_\nu \in \mathrm{Hom}_X(\overline{x}, X_\nu)$ satisfy

1) $f_\nu = \phi_{\nu\mu} \circ f_\mu$; and
2) for any $Y \in \mathbf{FEt}/X$ the natural map $\varinjlim_\nu \mathrm{Hom}_X(X_\nu, Y) \to \mathrm{Hom}_X(\overline{x}, Y)$ is an isomorphism.

Notation. Given a morphism $Y \to X$ and $\mathcal{F}$ an étale sheaf on X (cf. [**Mil80**, Chapter II]), we denote by $\mathcal{F}_{|Y}$ the pullback of $\mathcal{F}$ to Y. For any $n \geq 0$ and $\alpha \in H^n_{\text{ét}}(X, \mathcal{F})$ denote by $\alpha_{|Y} \in H^n_{\text{ét}}(Y, \mathcal{F}_{|Y})$ the pullback of α to Y.

Definition 5.3 ([**Mil80**, p. 155 and 220])**.** An étale sheaf $\mathcal{F}$ on X is called finite if for every quasi-compact $U \subset X$, $\mathcal{F}(U)$ is finite. $\mathcal{F}$ has finite stalks if for every geometric point $\overline{x}$ of X, $\mathcal{F}_{\overline{x}}$ is finite. $\mathcal{F}$ is called locally constant if there exists a covering $(U_\xi \to X)_{\xi \in \Xi}$ such that for every $\xi \in \Xi$, $\mathcal{F}_{|U_\xi}$ is constant. $\mathcal{F}$ is called a p-torsion sheaf if for every quasi-compact $U \subset X$, $\mathcal{F}(U)$ is killed by a power of p.

Proposition 5.4 ([**Mil80**, Proposition 1.1, Remark 1.2 (b)])**.** *Each locally constant sheaf $\mathcal{F}$ on X with finite stalks is finite and represented by a group scheme $\widetilde{\mathcal{F}}$ that is finite and étale over X. Furthermore, there exists a finite étale morphism $X' \to X$ such that $\widetilde{\mathcal{F}} \times_X X'$ is a disjoint union of copies of X' and $\mathcal{F}_{|X'}$ is constant.*

Convention. From this point till the end of this section, unless otherwise stated, $\mathcal{F}$ will denote a p-torsion locally constant sheaf on X with finite stalks.

Remark 5.5. It follows from Definition 5.3 and Proposition 5.4 that

$$\mathcal{F}_{|X'} \cong \bigoplus_{i=1}^{r} (\mathbb{Z}/p^{n_i}\mathbb{Z})^{m_i}, \tag{5.1}$$

where the n_i's and m_i's are positive integers.

Proposition 5.6. *For each $Y \in \mathbf{FEt}/X$ and $\beta \in H^1_{\mathrm{et}}(Y, \mathcal{F}_{|Y})$ there exists $Z \in \mathbf{FEt}/X$ such that Z factors through Y and $\beta_{|Z} \in H^1_{\mathrm{et}}(Z, \mathcal{F}_{|Z})$ is trivial.*

Proof. We start with the case where $Y = X$. Given $\beta \in H^1_{\mathrm{et}}(X, \mathcal{F})$, let x' be as in Prop. 5.4 and $\beta' = \beta_{|X'} \in H^1_{\mathrm{et}}(X', \mathcal{F}_{|X'})$. By (5.1)

$$H^1_{\mathrm{et}}(X', \mathcal{F}_{|X'}) \cong \bigoplus_{i=1}^{r} H^1_{\mathrm{et}}(X', \mathbb{Z}/p^{n_i}\mathbb{Z})^{m_i}.$$

So, we denote $\beta' = (\beta_{1,1}, \dots, \beta_{1,m_1}, \dots, \beta_{r,1}, \dots, \beta_{r,m_r})$ with $\beta_{i,j} \in H^1_{\mathrm{et}}(X', \mathbb{Z}/p^{n_i}\mathbb{Z})$. Let $\mathcal{W}_n$ be the sheaf of Witt vectors of length n on X [**Ser56**, §2], $F_{\mathrm{abs}} : X \to X$ the absolute Frobenius morphism and $\wp$ the operator $\wp(x) = x^p - x$. The exact sequence

$$1 \to \mathbb{Z}/p^n\mathbb{Z} \to \mathcal{W}_n \xrightarrow{\wp} \mathcal{W}_n \to 1,$$

gives an isomorphism $H^1_{\mathrm{et}}(X, \mathbb{Z}/p^n\mathbb{Z}) \cong H^1(X, \mathcal{W}_n)^{F_{\mathrm{abs}}}$, as in the usual Artin-Schreier theory [**Mil80**, p. 127-128]. Hence, by [**Ser56**, Proposition 13], we conclude that $\beta_{i,j}$ parametrizes a cyclic étale cover $X_{i,j} \to X'$ of degree p^{n_i}. Given $\alpha \in H^1_{\mathrm{et}}(X', \mathbb{Z}/p^n\mathbb{Z})$ and $V \to X'$ any finite étale cover, let $X'' \to X'$ be the cyclic étale cover of degree p^n defined by α and let $\alpha' = \alpha_{|V} \in H^1_{\mathrm{et}}(V, \mathbb{Z}/p^n\mathbb{Z}_{|V})$. Thus α' parametrizes the covering $W = V \times_{X'} X'' \to V$. In the case where $\alpha = \beta_{i,j}$, the covering $X_{i,j} \to X'$ plays the role of both $V \to X'$ and $X'' \to X'$. Therefore $\beta_{i,j|X_{i,j}}$ is trivial. Let $Z \to X'$ be a finite étale cover such that for every $i \in \{1, \dots, r\}$ and $j \in \{1, \dots, m_i\}$. The cover $Z \to X'$ factors through $X_{i,j} \to X'$. Therefore $\beta_{|Z} = \beta'_{|Z} \in H^1_{\mathrm{et}}(Z, \mathcal{F}_{|Z})$ is trivial.

In the case where $Y \neq X$, let $Y' = Y \times_X X'$. We have

$$\mathcal{F}_{|Y'} \cong \bigoplus_{i=1}^{r} (\mathbb{Z}/p^{n_i}\mathbb{Z})^{m_i}.$$

It follows from the above argument that there exists a finite étale cover $Z \to Y'$ such that $\beta_{|Z} = (\beta_{|Y'})_{|Z} \in H^1_{\mathrm{et}}(Z, \mathcal{F}_{|Z})$ is trivial. $\square$

Proposition 5.7. *For each $Y \in \mathbf{FEt}/X$ there exists $Z \in \mathbf{FEt}/X$ which factors through Y such that $H^0_{\mathrm{et}}(Z, \mathcal{F}_{|Z}) \cong \bigoplus_{i=1}^{r} (\mathbb{Z}/p^{n_i}\mathbb{Z})^{m_i}$.*

Proof. As in the proof of Proposition 5.6 it suffices to take $Z = Y \times_X X'$. $\square$

Remark 5.8 ([**Mil80**, Chapter I, 5.4])**.** Given $Y \in \mathbf{FEt}/X$ denote by $\mathrm{Aut}_X(Y)$ the set of X-automorphisms of Y. There exists $Z \in \mathbf{FEt}/X$ such that $Z \to X$ is Galois and $Z \to Y$ is an X-morphism. In this case $\mathrm{Hom}_X(\overline{x}, Z)$ is isomorphic to $\mathrm{Aut}_X(Z)$. In particular the elements of the projective system $(X_\nu, \phi_{\nu\mu})$ can be taken so that for each ν the cover $X_\nu \to X$ is Galois. Furthermore, $\pi_1(X, \overline{x}) = \varprojlim_\nu \mathrm{Aut}_X(X_\nu)$.

Remark 5.9. Since for each ν the map $X_\nu \to X$ is finite, hence affine, it follows from [**SGA 4**, VII, §5] that the projective limit of schemes $\widehat{X} = \varprojlim_\nu X_\nu$ exists. Moreover, by [**Mil80**, Chapter III, Lemma 1.16], for any étale sheaf $\mathcal{F}$ on X and for any integer $n \geq 0$ we have

$$H^n_{\mathrm{et}}(\widehat{X}, \mathcal{F}_{|\widehat{X}}) \cong \varinjlim_\nu H^n_{\mathrm{et}}(X_\nu, \mathcal{F}_{|X_\nu}).$$

Corollary 5.10. *$H^1_{\mathrm{et}}(\widehat{X}, \mathcal{F}_{|\widehat{X}}) = 0$ and $H^0_{\mathrm{et}}(\widehat{X}, \mathcal{F}_{|\widehat{X}}) \cong \bigoplus_{i=1}^r (\mathbb{Z}/p^{n_i}\mathbb{Z})^{m_i}$.*

Proof. This is an immediate consequence of Propositions 5.6 and 5.7 and Remark 5.9. $\square$

Theorem 5.11. *Let X be a smooth projective connected algebraic curve defined over an algebraically closed field of characteristic $p > 0$. For any closed point $\overline{x}$ of X we have $\mathrm{cd}_p(\pi_1(X, \overline{x})) \leq 1$.*

Proof. It follows from [**Sha72**, p. 55, Theorem 11] that it suffices to show that $H^2(\pi_1(X, \overline{x}), F) = 0$ for any finite simple $\pi_1(X, \overline{x})$-module F of p-power order. By [**Mil80**, p. 155-156], any such F is associated uniquely to a p-torsion locally constant étale sheaf $\mathcal{F}$ with finite stalks. Proposition 5.4 shows that there exists $X' \in \mathbf{FEt}/X$ such that

$$F \cong \mathcal{F}_{|X'} \cong \bigoplus_{i=1}^{r} (\mathbb{Z}/p^{n_i}\mathbb{Z})^{m_i}. \tag{5.2}$$

Furthermore, by [**SGA 4**, X, Corollary 5.2], since X is a smooth projective connected algebraic curve defined over k, we conclude that

$$H^n_{\mathrm{et}}(X, \mathcal{F}) = 0 \text{ for any } n \geq 2. \tag{5.3}$$

For every ν we consider the Hochschild-Serre spectral sequence [**Mil80**, p. 105, Theorem 2.20] $E_\nu^{r,s} \Rightarrow E^{r+s}$, where $E_\nu^{r,s} = H^r(\mathrm{Aut}_X(X_\nu), H^s_{\mathrm{et}}(X_\nu, \mathcal{F}_{|X_\nu}))$ and $E^{r+s} = H^{r+s}_{\mathrm{et}}(X, \mathcal{F})$. Also, as in [**Mil80**, p. 106 (b)], taking the projective limit we obtain a spectral sequence $E_\infty^{r,s} \Rightarrow E^{r+s}$, where $E_\infty^{r,s} = H^r(\pi_1(X, \overline{x}), H^s_{\mathrm{et}}(\widehat{X}, \mathcal{F}_{|\widehat{X}}))$. Furthermore, it follows from [**Mil80**, p. 309, l.8] that there exists an exact sequence

(5.4)

$$0 \to H^1(\pi_1(X,\overline{x}), H^0_{\text{et}}(\widehat{X}, \mathcal{F}_{|\widehat{X}})) \to H^1_{\text{et}}(X,\mathcal{F}) \to H^0(\pi_1(X,\overline{x}), H^1_{\text{et}}(\widehat{X}, \mathcal{F}_{|\widehat{X}}))$$
$$\to H^2(\pi_1(X,\overline{x}), H^0_{\text{et}}(\widehat{X}, \mathcal{F}_{|\widehat{X}})) \to H^2_{\text{et}}(X,\mathcal{F}) \to H^1(\pi_1(X,\overline{x}), H^1_{\text{et}}(\widehat{X}, \mathcal{F}_{|\widehat{X}})).$$

Finally, we conclude from Corollary 5.10, (5.2), (5.3) and (5.4) that $H^2(\pi_1(X,\overline{x}), F) = 0$. Thus, $\text{cd}_p(\pi_1(X,\overline{x})) \leq 1$. □

In the next two corollaries we assume that X has genus $g \geq 2$. In this case it follows from [**Ray82**, Corollaire 4.3.2] that the p-Sylow subgroups of $\pi_1(X,\overline{x})$ are non-trivial.

Corollary 5.12. *For every finite simple p-power order $\pi_1(X,\overline{x})$-module F we have $H^1(\pi_1(X,\overline{x}), F) \cong H^1_{\text{et}}(X,\mathcal{F})$.*

Proof. The result is a consequence of Corollary 5.10 and (5.4). □

Corollary 5.13. *The p-Sylow subgroups of $\pi_1(X,\overline{x})$ are non-trivial and pro-p-free.*

Proof. Recall that [**Ser86**, p. I-20, Proposition 14 (i)] implies $\text{cd}_p(P) = \text{cd}_p(\pi_1(X,\overline{x}))$, for any p-Sylow subgroup P of $\pi_1(X,\overline{x})$. Moreover, it follows from Theorem 5.11 that $\text{cd}_p(\pi_1(X,\overline{x})) \leq 1$. But, for a pro-$p$-group P this is equivalent to P being pro-p-free. □

6. Galois covers.

Proof of Theorem 1.3. Let $\pi_1(X) \twoheadrightarrow G$ be a proper solution for the embedding problem $(\phi : \pi_1(X) \twoheadrightarrow H, G \twoheadrightarrow H)$. Let $Y \to X$ be the Galois H-cover corresponding to ϕ. Thus, $\gamma_{\phi,\chi} = \gamma_{Y,\chi}$. Recall that $\Phi(P)$ is the Frattini subgroup of P and $\mathcal{P} = P/\Phi(P)$. Observe that $\mathcal{P} \in \pi_A(Y)$. It follows from the correspondence described in the second paragraph of Section 3 that $\text{Hom}(\mathcal{P}, \mathbb{Z}/p\mathbb{Z})$ is an $\mathbb{F}_p$-subspace of $\text{Hom}(\pi_1(Y), \mathbb{Z}/p\mathbb{Z})$. This latter space is $\mathbb{F}_p$-isomorphic to $\text{Hom}(\text{Ker}(1-\mathcal{C} \,|\, \Omega^1_Y(0)), \mathbb{F}_p)$ by Serre's duality [**Ser56**, §9]. Therefore, (3.2) and (3.3) imply $m_\chi \leq t_\chi$, for every $\chi \in Z(H)$. Note that $\text{Ker}(1-\mathcal{C} \,|\, \bigoplus_{j=0}^{l_\chi-1} \Omega^1_Y(0)_{\chi^{p^j}})$ and $\text{Ker}(1-\mathcal{C}^{l_\chi} \,|\, \Omega^1_Y(0)_\chi)$ are $\mathbb{F}_p[H]$-isomorphic via $\omega = \sum_{j=0}^{l_\chi-1} \omega_j \mapsto \omega_0$ (cf. [**Pac95**, Lemma 2.14]). Moreover, $\dim_{\mathbb{F}_p} \text{Ker}(1-\mathcal{C}^{l_\chi} \,|\, \Omega^1_Y(0)_\chi) = \gamma_{Y,\chi} l_\chi$, therefore $t_\chi = \gamma_{Y,\chi}/n_\chi$ (cf. [**Pac95**, Corollary 3.6]), hence $m_\chi n_\chi \leq \gamma_{Y,\chi}$ (cf. [**Ste96a**, Proposition 3.4]). Conversely, suppose that $m_\chi \leq \gamma_{Y,\chi}/n_\chi$, for every $\chi \in Z(H)$. Since $t_\chi = \gamma_{Y,\chi}/n_\chi$, it follows from (3.3) there exists an $\mathbb{F}_p[H]$-submodule $\mathcal{B}_\chi$ of $\text{Ker}(1-\mathcal{C} \,|\, \bigoplus_{j=0}^{l_\chi-1} \Omega^1_Y(0)_{\chi^{p^j}})$ such that $\mathcal{B}_\chi \cong \mathcal{V}_{[\chi]}^{m_\chi}$. Let $\mathcal{B} = \bigoplus_{\chi \in Z(H)} \mathcal{B}_\chi$ and remark that there exists an $\mathbb{F}_p[H]$-isomorphism between $\mathcal{B}$ and $\mathcal{P}$. Once again by the the correspondence described in the second paragraph of Section 3, $\mathrm{Hom}(\mathcal{B}, \mathbb{F}_p)$ is $\mathbb{F}_p[H]$-isomorphic to $\mathrm{Hom}(\mathrm{Gal}(Z/Y), \mathbb{Z}/p\mathbb{Z})$ for some étale cover $Z \to Y$ and $\mathrm{Gal}(Z/Y) \cong \mathcal{P}$. Therefore, Lemma 3.2 implies that $Z \to X$ is Galois and $\mathrm{Gal}(Z/X) \cong \mathcal{P} \rtimes H$. Hence $\mathcal{P} \rtimes H \in \pi_A(X)$. It follows from Theorem 5.11 that $\mathrm{cd}_p(\pi_1(X, \overline{x})) \leq 1$ for any closed point $\overline{x}$ of X. Therefore, the argument of Remark 4.4 implies that $G \in \pi_A(X)$. □

7. A generic condition.

Theorem 1.3 tells us that if we are given a finite group G with a normal p-Sylow subgroup P and quotient H, then whether or not G lies in $\pi_A(X)$ depends not only on the size of P, but also on the specific action of H on P. The role that the action of H on P plays in this question was examined previously in the work of Nakajima [**Nak87**, Theorem A], Pacheco [**Pac95**, Propositions 2.4 and 2.5] and Stevenson [**Ste96a**, Proposition 3.5]. However, for the groups we are considering, Theorem 1.3 is stronger. In particular, it gives us a necessary and sufficient condition which is reasonably easy to compute. We begin this section with some consequences of Theorem 1.3. These involve situations where the generalized Hasse-Witt invariants can be most easily computed. At the end of this section we compute Condition A for the curve X_g (which represents the generic geometric point of the coarse moduli scheme $\mathcal{M}_g$ of curves of genus g) under the assumption that H is abelian. This situation is sufficient to demonstrate the strengths of our results while also distinguishing it from previous work.

As a preliminary step, we will prove the result mentioned in Remark 1.5, which deals with "ordinary Galois H-covers". The advantage in this case is that Condition A can be rephrased in a way that is independent of the H-cover. Given a finite group G having a normal p-Sylow subgroup P, recall that $H = G/P$, $Z(H)$ denotes the set of irreducible characters χ of H defined over the algebraically closed field k of characteristic $p > 0$ and χ^0 is the trivial character of H.

Theorem 7.1. *Let G be a finite group having a normal p-Sylow subgroup P. Let $H = G/P$. Suppose that $\phi : \pi_1(X) \twoheadrightarrow H$ corresponds to a Galois H-cover $Y \to X$ where Y is an ordinary curve. An embedding problem $(\phi : \pi_1(X) \twoheadrightarrow H, G \twoheadrightarrow H)$ has a proper solution if and only if $m_{\chi^0} \leq g$, and $m_\chi \leq (g-1)n_\chi$, for $\chi \neq \chi^0$.*

Proof. Notice that by Remark 2.3, the Galois H-cover Y is ordinary if and only if we have

$$\gamma_{Y,\chi} = \begin{cases} g, & \text{if } \chi = \chi^0 \text{ and} \\ (g-1)\, n_\chi^2, & \text{if } \chi \neq \chi^0. \end{cases} \tag{7.1}$$

Thus condition A is equivalent to the condition of Theorem 7.1. □

Let $g \geq 2$ be an integer and $\pi_A(g)$ the set of isomorphism classes of finite groups G such that $G \in \pi_A(X)$ for some smooth projective connected curve X of genus g.

Remark 7.2. Suppose that there exists some smooth projective connected curve X defined over k such that a finite group $G \in \pi_A(X)$. Denote by $x \in \mathcal{M}_g$ the point corresponding to X. In [**Ste96**, Proposition 4.2] Stevenson showed that in this case there exists an open subset U of $\mathcal{M}_g$ containing x such that for every $z \in U$ we have $G \in \pi_A(Z)$, where Z denotes the curve corresponding to z. In particular, $G \in \pi_A(X_g)$, therefore $\pi_A(X_g) = \pi_A(g)$.

Remark 7.3. It is an immediate consequence of the definition of $\pi_A(g)$ that a finite group G satisfying the hypothesis of Theorem 1.3 lies in $\pi_A(g)$ if and only if there exists a smooth projective connected curve X of genus g for which Condition A holds.

Notation. Let G be a finite group. Denote by $d(G)$ the minimum number of generators of G.

Now we can prove another consequence of Theorem 1.3.

Theorem 7.4. *Let G be a finite group having a normal p-Sylow subgroup P. Suppose that $H = G/P$ is abelian and $g \geq 2$. A necessary and sufficient condition for $G \in \pi_A(g)$ is $d(H) \leq 2g$, $m_{\chi^0} \leq g$ and $m_\chi \leq g-1$ for each $\chi \in Z(H)$ and $\chi \neq \chi^0$.*

Proof. Suppose that $G \in \pi_A(g)$. It follows from Remark 7.3 that there exists a smooth projective connected curve X and an étale Galois cover $Y \to X$ with $\mathrm{Gal}(Y/X) \cong H$ such that for every $\chi \in Z(H)$ we have $m_\chi \leq \gamma_{Y,\chi}$. By (2.2) we conclude that $\gamma_{Y,\chi^0} \leq g$ and $\gamma_{Y,\chi} \leq g-1$ for every $\chi \in Z(H)$, $\chi \neq \chi^0$. Moreover, since $H \in \pi_A(X)$, [**Groth71**, Corollary 2.12] implies that $d(H) \leq 2g$. In particular, the condition of Theorem 7.4 is satisfied. Conversely, suppose that $d(H) \leq 2g$, $m_{\chi^0} \leq g$ and $m_\chi \leq g-1$ for each $\chi \in Z(H)$ and $\chi \neq \chi^0$. Since H is abelian and $d(H) \leq 2g$, it follows from [**Groth71**, Corollary 2.12] that $H \in \pi_A(X_g)$, i.e., there exists an étale covering $Y_g \to X_g$ such that $\mathrm{Gal}(Y_g/X_g) \cong H$. It is a result due to Nakajima [**Nak83**, Theorem 2] that every étale cyclic covering $Z_g \to X_g$ of degree prime to p is ordinary. (It is essential here that X_g is generic.) This result was extended to all abelian prime to p groups by Zhang [**Zha92**, Théorème 3.1] (again for X_g). Hence Y_g is ordinary. So, by Theorem 7.1, $\gamma_{Y_g,\chi^0} = g$ and $\gamma_{Y_g,\chi} = g-1$ for every $\chi \in Z(H)$, $\chi \neq \chi^0$. Furthermore, by hypothesis, $m_{\chi^0} \leq g$ and $m_\chi \leq g-1$ for every $\chi \in Z(H)$, $\chi \neq \chi^0$. Therefore, Condition A holds for X_g and by Theorem 1.3, $G \in \pi_A(X_g)$. Finally, Remark 7.2 shows that this is equivalent to $G \in \pi_A(g)$. $\square$

Another result in this direction is the following one from [**Ste96a**].

Theorem 7.5 ([**Ste96a**, Propositions 3.1 and 3.2])**.** *Let G be a finite group having a normal p-Sylow subgroup P and $H = G/P$. Suppose that $g \geq 2$ and $d(H) \leq g$. A necessary and sufficient condition for $G \in \pi_A(g)$ is $m_{\chi^0} \leq g$ and $m_\chi \leq (g-1)n_\chi$ for each $\chi \in Z(H)$ and $\chi \neq \chi^0$.*

Remark 7.6. Notice that for an abelian group H such that $d(H) \leq 2g$, Theorem 7.4 is stronger than Theorem 7.5 since the latter requires that $d(H) \leq g$. However, for arbitrary H with $d(H) \leq g$, Theorem 7.5 is stronger than Theorem 7.4.

Now we can compare these results to a result of Nakajima. Let G be a finite group, $I_G = \{\sum_{\sigma \in G} a_\sigma \sigma \in \mathbb{Z}[G]\,;\, \sum_{\sigma \in G} a_\sigma = 0\}$ its augmentation ideal and $t(G)$ the minimum number of generators of I_G. Suppose that there exists a smooth projective curve X of genus g such that $G \in \pi_A(X)$, i.e., $G \cong \mathrm{Gal}(Y/X)$ for some étale Galois cover $Y \to X$.

Theorem 7.7 (Nakajima, [**Nak84**, Theorem 4])**.** *There exists a short exact sequence of $k[G]$-modules*

$$1 \to \Omega^1_Y(0) \to k[G]^g \to I_G \to 1.$$

Corollary 7.8 (Nakajima, [**Nak87**, Theorem A])**.** *$t(G) \leq g$.*

Notation. We call Condition B the inequality of Corollary 7.8.

Remark 7.9. From the definition of $\pi_A(g)$, Theorem 7.7 and Corollary 7.8, we see that Condition B is necessary for $G \in \pi_A(g)$.

Corollary 7.10. *Let G be a finite group having a normal p-Sylow subgroup P, $H = G/P$. Suppose that either:* (a) *H is abelian and $d(H) \leq 2g$; or* (b) *$d(H) \leq g$. Under either hypothesis* (a) *or* (b) *Condition* A *is equivalent to Condition* B.

Proof. By [**Ste96a**, Proposition 3.5] Condition A implies Condition B without any restrictions on H. Conversely, by [**Ste96**, Proposition 3.1] Condition B implies that $m_{\chi^0} \leq g$ and $m_\chi \leq (g-1)n_\chi$ for each $\chi \in Z(H)$ and $\chi \neq \chi^0$. Under hypothesis (a) (resp. (b)) Theorem 7.4 (resp. 7.5) show that the latter condition implies that $G \in \pi_A(g)$. Now by Theorem 1.3 this implies Condition A. $\square$

In order to obtain a converse in the case where H is a non-abelian finite quotient of Γ_g we need to generalize the Nakajima-Zhang result ([**Nak83**, Theorem 2] and [**Zha92**, Théorème 3.1]) to non-abelian Galois étale covers of degree prime to p of X_g. Another option is to show that there exists an ordinary Galois H-cover of a curve X of genus g and apply [**Ste96**] (cf. Remark 1.5). Very recently M. Raynaud has found a counter example to both these approaches.

Example 7.11. Theorem 7.4 gives a result which is not covered by [**Ste96a**, Theorem 3.2] in the case where H is abelian and $g < d(H) \leq 2g$. Let $n \geq 1$ be an integer and let $g \geq 2$ be an integer. Let $H = (\mathbb{Z}/n\mathbb{Z})^{2g}$ and label the elements τ_j for $j = 1, \ldots, n^{2g}$. For each $i = 1, 2, \ldots, g-1$ let $P_i = (\mathbb{Z}/p\mathbb{Z})^{n^{2g}}$ and $P_g = \mathbb{Z}/p\mathbb{Z}$. Pick a basis $a_{i,\tau_1}, \ldots, a_{i,\tau_{n^{2g}}}$ for P_i for $i = 1, \ldots, g-1$ and let a_g be a basis of P_g. Then we define an action of H on each P_i for $i = 1, \ldots, g-1$ as follows: $\rho_i : H \to \mathrm{Aut}(P_i)$ by $\rho_i(\tau_j)a_{i,\tau_l} = a_{i,\tau_j\tau_l}$. With this action each P_i is isomorphic to $\mathbb{F}_p[H]$, which is the H-module defined over $\mathbb{F}_p$ corresponding to the regular representation of H. Let H act on P_g trivially. Now let $P = \bigoplus_{i=1}^{g} P_i$ with the induced action of H on P. Then P is isomorphic to $(\mathbb{Z}/p\mathbb{Z})^{n^{2g}(g-1)+1}$ as a group and to $\mathbb{F}_p[H]^{g-1} \oplus \mathbb{F}_p$ as an $\mathbb{F}_p[H]$-module. Let G be defined as the semi-direct product $P \rtimes H$ with respect to this action.

By construction $P \otimes_{\mathbb{F}_p} k$ is isomorphic as a $k[H]$-module to $k[H]^{g-1} \oplus k$. Let $Z(H)$ be the set of irreducible characters of H defined over k and let χ^0 be the trivial character of H. Then using the notation of Section 1.1, $m_{\chi^0} = g$ and $m_\chi = g - 1$ for $\chi \neq \chi^0$. Note that since H is abelian, by Zhang's theorem [**Zha92**, Théorème 3.1], any Galois H-cover Y_g of X_g is ordinary, thus $\gamma_{Y_g,\chi^0} = g$, and $\gamma_{Y_g,\chi} = g - 1$ for $\chi \neq \chi^0$. In particular, Condition A is satisfied for the curve X_g, therefore $G \in \pi_A(X_g) = \pi_A(g)$.

Remark 7.12. In the set-up of Example 7.11, as the rank of H is greater than g, Theorem 7.5 does not apply. If we keep P the same but change the action of H on P in any way, then G will not lie in $\pi_A(g)$, because for some character $\chi \neq \chi^0$ we would have $m_\chi > g - 1$ or $m_{\chi^0} > g$. Finally, if we replace P by any p-group Q with Frattini quotient isomorphic to P and extend the action of H in P of Example 7.11 to an action of H in Q, then we get (as in the end of the proof of Theorem 1.3) $Q \rtimes H \in \pi_A(X_g) = \pi_A(g)$.

References

[CheWei34] C. Chevalley and A. Weil, *Über das Verhalten der Integrale* 1. *Gattung bei Automorphismen des Funktionenkörpers*, A. Weil Collected Papers, **I**, 68-71.

[Groth71] A. Grothendieck, *Rêvetements étales et groupe fondamental* (SGA 1), Lecture Notes in Math., **244**, Springer-Verlag, 1971.

[SGA 4] A. Grothendieck, M. Artin and J.-L. Verdier, *SGA* 4: *Théorie des topos et cohomologie étale des schémas*, Lecture Notes in Mathematics, **269**, **270** and **305**, Springer-Verlag, 1972-1973.

[Har95] D. Harbater, *Fundamental groups and embedding problems in characteristic p*, Contemporary Mathematics, **186** (1995), 353-369.

[Isa76] I.M. Isaacs, *Character Theory of Finite Groups*, Academic Press, 1976.

[Kat88] N. Katz, *Local-to-global extensions of representations of fundamental groups*, Ann. Inst. Fourier, **36** (1986), 69-106.

[Mil80] J.S. Milne, *Étale cohomology*, Princeton University Press, 1980.

[Pac95] A. Pacheco, *Unramified Galois coverings of algebraic curves*, J. Number Theory, **53** (1995), 211-228.

[Nak83] S. Nakajima, *On generalized Hasse-Witt invariants of an algebraic curve*, Adv. Studies Pure Math., **2** (1983), 69-88.

[Nak84] ______, *On the Galois module structure of the cohomology groups of an algebraic variety*, Invent. Math., **75** (1984), 1-8.

[Nak87] ______, *On the quotients of the fundamental group of an algebraic curve*, Adv. Stu. Pure Math., **12** (1987), 259-264.

[Ray82] M. Raynaud, *Sections des fibrés véctoriels sur une courbe*, Bull. Soc. Math. France, **110** (1982), 103-125.

[Ruc86] H.G. Rück, *L-series and class groups of function fields*, J. Number Theory, **22** (1986), 177-189.

[Ser56] J.-P. Serre, *Sur la topologie des variétés algébriques en caractéristique p*, in 'Symp. Topologia Algebraica' (Mexico City), 1956, 24-53.

[Ser86] ______, *Cohomologie Galoisienne*, Lecture Notes in Mathematics, **5** (1986), Springer-Verlag.

[Ser88] ______, *Algebraic Groups and Class Fields*, Springer-Verlag, 1988.

[Ser90] ______, *Constructions de revêtements étales de la droite affine en caractéristique p*, C. R. Acad. Sci. Paris, **311**, Série I, (1990), 341-346.

[Sha72] S.S. Shatz, *Profinite Groups, Arithmetic, and Geometry*, Princeton University Press, 1977.

[Ste96] K. Stevenson, *Galois groups of étale covers of projective curves in characteristic p*, J. Algebra, **182** (1996), 770-804.

[Ste96a] ______, *Conditions related to π_1 of projective curves*, J. Number Theory, (March, 1998), 62-79.

[Zha92] B. Zhang, *Revêtements étales abéliens de courbes génériques et ordinarité*, Ann. Fac. Sci. Toulouse Sér. 6, **1** (1992), 133-138.

Received May 18, 1998. Amílcar Pacheco was partially supported by CNPq research grant number 300896/91-3 and Pronex, Brazil. Section 5 was partially written during a visit of Amílcar Pacheco to the University of Campinas, São Paulo, Brazil. He would like to thank this Institution for its hospitality and Antônio Engler for many conversations. Katherine F. Stevenson was supported by NSA grant MDA 9049710017. Both authors would like to thank David Harbater and his student Rachel Pries for their careful reading and many suggestions on previous versions of this paper.

Universidade Federal do Rio de Janeiro
Rio de Janeiro RJ
Brasil
E-mail address: amilcar@impa.br

California Institute of Technology
Pasadena, CA 91125
E-mail address: kfs@vision.caltech.edu

PACIFIC JOURNAL OF MATHEMATICS
Vol. 192, No. 1, 2000

SHAPE EQUIVALENCE, NONSTABLE K-THEORY AND AH ALGEBRAS

Cornel Pasnicu

We give several necessary and sufficient conditions for an AH algebra to have its ideals generated by their projections. Denote by $\mathcal{C}$ the class of AH algebras as above and in addition with slow dimension growth. We completely classify the algebras in $\mathcal{C}$ up to a shape equivalence by a K-theoretical invariant. For this, we show first, in particular, that any C^*-algebra in $\mathcal{C}$ is shape equivalent to an AH algebra with slow dimension growth and real rank zero (generalizing so a result of Elliott-Gong); then, we use a classification result of Dadarlat-Gong. We prove that any AH algebra in $\mathcal{C}$ has stable rank one (i.e., in the unital case, that the set of the invertible elements is dense in the algebra), generalizing results of Blackadar-Dadarlat-Rørdam and of Elliott-Gong. Other nonstable K-theoretical results for C^*-algebras in $\mathcal{C}$ are also proved, generalizing results of Dadarlat-Némethi, Martin-Pasnicu and Blackadar.

1. Introduction.

An AH algebra is an amenable C^*-algebra of the form $A = \varinjlim(A_n, \Phi_{n,m})$, with $A_n = \oplus_{i=1}^{k_n} P_{n,i} C(X_{n,i}, M_{[n,i]}) P_{n,i}$, where $X_{n,i}$ are finite, connected CW complexes, k_n, $[n,i]$ are positive integers and $P_{n,i} \in C(X_{n,i} M_{[n,i]})$ are projections. The problem of finding suitable invariants for AH algebras was raised by Effros [**Eff**] (after the results obtained by Glimm [**Gl**], Dixmier [**Di**], Bratteli [**Br**] and especially after Elliott's classification of the AF algebras by their ordered K_0 group [**Ell 1**]) and is now a part of Elliott's project of classification of the separable, amenable C^*-algebras by invariants including K-theory (for details see e.g. [**Ell 6**]).

In this paper we shall deal with AH algebras B whose ideals are generated (as ideals) by their projections (here by an ideal we mean a closed, two-sided ideal). The study of these amenable C^*-algebras was suggested by Elliott and they provide a common generalization of two very important classes of algebras: The real rank zero AH algebras (i.e., with the property that any self-adjoint element can be approximated as close as we want by self-adjoint elements with finite spectrum [**BP**]) and the simple AH algebras.

Recently, successful classification results have been obtained for certain classes of AH algebras. On one hand is the remarkable classification of the real rank zero AH algebras with slow dimension growth (see [**Ell 3**], [**EGLP 1**, **EGLP 2**], [**EG 1**, **EG 2**], [**Da 2**], [**Go 1**, **Go 2**], [**DL 1**], [**DL 2**], [**Ei**] and [**DG**], the last paper containing the general classification result). On the other hand is the very remarkable classification obtained in [**Go 3**] and [**EGL**] for the simple, unital AH algebras with very slow dimension growth (also, see [**Ell 4**, **Ell 5**] and [**L**]). It would be important to generalize and unify these classification results for the above two classes of AH algebras. Our paper can be seen as the first step in this attempt. We believe that the results of this paper, combined with the techniques from [**Go 3**] and [**EGL**], will play an essential role in the classification of the AH algebras whose ideals are generated by projections and which have slow dimension growth. A particular class of AI algebras (i.e., inductive limits of sequences of C^*-algebras of the form $C([0,1], F)$, with F a finite dimension C^*-algebra) whose ideals are generated by projections has been classified in [**St**].

We have been able to give several necessary and sufficient conditions for an AH algebra to have its ideals generated by their projections (see Theorem 3.1 and Remark 3.2 b)). Let $\mathcal{C}$ be the class of AH algebras whose ideals are generated by their projections and with slow dimension growth. Also, we classified the C^*-algebras in $\mathcal{C}$ up to a shape equivalence by a K-theoretical invariant (see Theorem 2.15). To do it, we generalized first - relying heavily on part of Theorem 3.1, [**DN**] and [**EG 2**] - a result of Elliott-Gong ([**EG 2**]) proving, in particular, that any AH algebra in $\mathcal{C}$ is shape equivalent to an AH algebra with slow dimension growth and real rank zero (see Theorem 2.6); then, we used the classification result of the AH algebras with real rank zero and slow dimension growth obtained by Dadarlat-Gong in [**DG**].

We have been able also to obtain nonstable K-theoretical results for the algebras in $\mathcal{C}$. One of the main results of this paper says that any C^*-algebra in $\mathcal{C}$ has stable rank one (that means, in the unital case, that the set of the invertible elements is dense in the algebra [**R**]) (see Theorem 4.1). This Theorem extends a result of Blackadar-Dadarlat-Rørdam ([**BDR**]) (and hence, also one in [**DNNP**]) in the case when the C^*-algebra is simple and a result of Elliott-Gong in [**EG 2**] in the case when the AH algebra has real rank zero. It is important to mention that Theorem 4.1 contains more information than one might think at a first sight; e.g., it allows us to compute the real rank for the AH algebras in $\mathcal{C}$. Indeed, this theorem together with the inequality [**BP**], Proposition 1.2 imply that the real rank for a C^*-algebra in C could be 0 or 1. But the real rank zero case was characterized by us in [**P 2**]. Note that if we drop the slow dimension growth condition, then Theorem 4.1 is no more true. Indeed, a recent result of Villadsen [**V**] shows that for any positive integer n there is a simple AH algebra of stable rank n and which does not have slow dimension growth. Note also that the C^*-algebras A with stable rank one have many "nice" and important properties (e.g., cancellation and, in the unital case, the canonical map $U(A)/U_0(A) \to K_1(A)$ is a group isomorphism [**R**]).

We have proved also that for any AH algebra A in $\mathcal{C}, K_0(A)$ is weakly unperforated in the sense of Elliott [**Ell 2**] and, in the unital case, that if two projections p and q in A satisfy $\sigma(p) < \sigma(q)$ for any tracial state σ of A, then p is Murray-von Neumann equivalent to a proper subprojection of q (see Theorem 5.1). This theorem generalizes results of Dadarlat-Némethi [**DN**], Martin-Pasnicu [**MP**] and Blackadar [**Bl 2**].

We are grateful to Guihua Gong for useful discussions concerning the subject of our investigation.

2. The shape equivalence type.

In this section we shall describe the shape equivalence type for a large class of AH algebras by a K-theoretical invariant (see Theorem 2.15). The proof of Theorem 2.15 combines Theorem 2.6 (which is based on part of Theorem 3.1, [**DN**] and [**EG 2**]) with the classification result of the AH algebras with real rand zero and slow dimension growth obtained by Dadarlat-Gong in [**DG**].

We begin with some notations and definitions.

Notation 2.1. Let $\mathcal{H}$ be the set of homogeneus C^*-algebras of the form $\oplus_{i=1}^{k} P_i C(X_i, M_{n_i}) P_i$ where k and n_i are positive integers, X_i is a finite, connected CW complex and $P_i \in C(X_i, M_{n_i})$ is a projection.

Definition 2.2. An AH algebra is a C^*-algebra $A = \varinjlim(A_n, \Phi_{n,m})$, where $(A_n, \Phi_{n,m})$ is an inductive system of C^*-algebras $A_n \in \mathcal{H}$ and homomorphisms.

Definition 2.3 ([**Go 1**])**.** We recall the following definition from [**Go 1**]. An inductive system $(A_n, \Phi_{n,m})$ of C^*-algebras and homomorphisms will be said to have *slow dimension growth* if $A_n = \oplus_{i=1}^{k_n} P_{n,i} C(X_{n,i}, M_{[n,i]}) P_{n,i}, X_{n,i}$ are finite, connected CW complexes, $P_{n,i} \in C(X_{n,i}, M_{[n,i]})$ are projections, $[n,i]$ and k_n are positive integers and for any n, there is a positive integer M such that:

$$\lim_{m\to\infty} \min_{\substack{\dim X_{m,j} > M \\ \operatorname{rank}(\Phi_{n,m}^{i,j}(P_{n,i})) \neq 0}} \left\{ \frac{\operatorname{rank}(\Phi_{n,m}^{i,j}(P_{n,i}))}{\dim X_{m,j} + 1} \right\} = +\infty$$

(we use the convention that the minimum of the empty set is $+\infty$).

An AH algebra will be said to have *slow dimension growth* if there is an inductive system $(A_n, \Phi_{n,m})$ with slow dimension growth such that $A = \varinjlim(A_n, \Phi_{n,m})$.

Note that the above definition gives a more general slow dimension growth condition than that defined in [**EG 2**] and one of its advantages is that any system $(\oplus_{i=1}^{k_n} P_{n,i} C(X_{n,i}, M_{[n,i]}) P_{n,i}, \Phi_{n,m})$ with $\sup_{n,i} \dim X_{n,i} < +\infty$ has slow dimension growth (take for any $n, M = \sup_{n,i} \dim X_{n,i}$) even though the limit is not simple.

The first notion of slow dimension growth for some simple AH algebras was introduced in [**BDR**]. For other extensions in the non-simple case see e.g. [**Goo**], [**MP**].

Notation 2.4. Let $\mathcal{S}$ be the category of separable C^*-algebras and homomorphisms.

Definition 2.5 ([**EK**], [**Bl 1**])**.** Two inductive systems $(A_n, \Phi_{n,m})$ and $(B_n, \Psi_{n,m})$ in $\mathcal{S}$ are said to be *shape equivalent* if there are positive integers $k_i < \ell_i < k_{i+1}$ $(i \geq 1)$ and homomorphisms $\xi_i : A_{k_i} \to B_{\ell_i}$ and $\eta_i : B_{\ell_i} \to A_{k_{i+1}}$ such that:

$$\eta_i \circ \xi_i \sim_h \Phi_{k_i, k_{i+1}} : A_{k_i} \to A_{k_{i+1}}$$

and

$$\xi_{i+1} \circ \eta_i \sim_h \Psi_{\ell_i, \ell_{i+1}} : B_{\ell_i} \to B_{\ell_{i+1}}$$

where $\sim_h$ means homotopy between homomorphisms.

The following theorem generalizes a result of Elliott and Gong in [**EG 2**].

Theorem 2.6. *Let $(A_n, \Phi_{n,m})$ be an inductive system in $\mathcal{S}$ with $A_n \in \mathcal{H}$, $n \geq 1$. Suppose that $(A_n, \Phi_{n,m})$ has slow dimension growth and that any ideal of its limit $A = \varinjlim(A_n, \Phi_{n,m})$ is generated by its projections.*

Then $(A_n, \Phi_{n,m})$ is shape equivalent to some inductive system $(B_n, \Psi_{n,m})$ in $\mathcal{S}$ with $B_n \in \mathcal{H}, n \geq 1$, such that $(B_n, \Psi_{n,m})$ has slow dimension growth and its limit $B = \varinjlim(B_n, \Psi_{n,m})$ has real rank zero.

Remarks 2.7. a) Observe that in the above Theorem 2.6 the AH algebra B can not be always simple. Indeed, let A be an AH algebra with real rank zero, stable rank one [**R**] which is not simple and which has slow dimension growth. Suppose that the AH algebra B given by Theorem 2.6 is in addition simple. The shape equivalence induces an isomorphism between $(K_0(A), K_0(A)_+)$ and $(K_0(B), K_0(B)_+)$. But since B is simple it follows that $(K_0(B), K_0(B)_+)$ is simple and hence $(K_0(A), K_0(A)_+)$ is also simple. But since A has real rank zero and stable rank one, by a general result in [**GL**] it follows that A is simple, a contradiction.

b) Observe that by [**Da 2**], [**Go 1**] and [**EG 2**], Theorem 2.2, the AH algebra B in the above Theorem is unique up to an isomorphism.

In the proof of the above theorem we shall need the following six lemmas. First, let us recall shortly some definitions. For a C^*-algebra $A = \oplus_{i=1}^{k} P_i C(X_i, M_{n_i}) P_i$, where $P_i \in C(X_i, M_{n_i})$ are nonzero projections and the X_i's are compact, connected spaces, define $SP(A) = \sqcup X_i$. If B is another C^*-algebra of the same type as A, $\Phi : A \to B$ is a homomorphism and $y \in SP(B)$, we want to define $SP(\Phi)_y$, the spectrum of Φ in the point y. The definition reduces practically (in a natural manner) (see e.g. [**EG 2**]) to the case when we have $\Phi : C(X) \to PC(Y, M_n)P$ and $y \in Y$. Since the map $C(X) \ni f \mapsto \Phi(f)(y) \in M_n$ is a finite dimensional $*$-representation, it follows that it is a direct sum of some irreducible $*$-representations of $C(X)$- given by the evaluation maps in $x_1(y), x_2(y), \cdots, x_m(y) \in X$- and of a zero $*$-representation of $C(X)$. Then, by definition $SP(\Phi)_y = \{x_1(y), x_2(y), \ldots, x_m(y)\}$, where we count multiplicity.

The first lemma is inspired by [**DNNP**], Proposition 2.1 (see also [**St**]).

Lemma 2.8. *Let* $A = \varinjlim(A_n, \Phi_{n,m})$, *with*

$$A_n = \oplus_{i=1}^{k_n} A_n^i, A_n^i = P_{n,i} C(X_{n,i}, M_{[n,i]}) P_{n,i},$$

where $X_{n,i}$ *are finite, connected* CW *complexes and* $P_{n,i} \in C(X_{n,i}, M_{[n,i]})$ *are nonzero projections. Suppose that any ideal of* A *is generated by projections.*

Then, for any fixed n *and fixed* $F = \overline{F} \subset U = \overset{\circ}{U} \subset \sqcup_{i=1}^{k_n} X_{n,i} = SP(A_n)$ *there is* $m_o > n$ *such that for any* $m \geq m_o$ *any partial map* $\Phi_{n,m}^j : A_n \to A_m^j$ *satisfies either:*

$$SP(\Phi_{n,m}^j)_y \cap F = \phi \quad \textit{for all} \;\; y \in X_{m,j}$$

or

$$SP(\Phi_{n,m}^j)_y \cap U \neq \phi \quad \textit{for all} \;\; y \in X_{m,j}.$$

Proof. We may suppose that $n = 1$ and that $F \subset U \subset X_{1,1}$. Define $f \in C(X_{1,1})$ such that $f(x) = 1$ if $x \in F$ and $f(x) = 0$ if $x \notin U$. Consider $\tilde{f} = \oplus_{i=1}^{k_n} f_i \in \oplus_{i=1}^{k_n} A_n^i$ where $f_1 = f \cdot P_{1,1}$ and $f_i = 0$ if $i \neq 1$.

Define $G_k = \{y \in SP(A_k) : SP(\Phi_{1,k})_y \cap U = \phi\}, k \in \mathbb{N}$. But then each G_k is a closed subset of $SP(A_k)$ (see e.g. [**Da 1**], [**P 1**]). As in [**DNNP**] define $J_k = \{g \in A_k : g\,|_{G_k} = 0\},\;\; k \geq 1$. Obviously J_k is an ideal in A_k and let J be the ideal of A defined by $J_k,\;\; k \geq 1$. Define by G_k^j the component of G_k in $X_{k,j}$. It is not difficult to see that if we define $p_m = \oplus_{j=1}^{k_m} p_m^j \in \oplus_{j=1}^{k_m} A_m^j = A_m,\;\; m \geq 1$ by:

$$p_m^j = \begin{cases} P_{m,j} & \text{if } \; G_m^j = \phi \\ 0 & \text{if } \; G_m^j \neq \phi \end{cases}$$

then each p_m is a projection in $A_m, p_m \in J_m$ and:

$$p_m \cdot aeb = aeb \cdot p_m = aeb$$

for any projection e of J_m and any $a, b \in A_m$ (each $X_{m,j}$ is a connected space). Since, by hypothesis, J is generated by its projections, the above equalities imply that $(\Phi_{m,\infty}(p_m))_{m=1}^{\infty}$ is an approximate unit for J. Here $\Phi_{m,\infty} : A_m \to A = \varinjlim(A_n, \Phi_{n,k})$ is the canonical homomorphism. Since obviously $\widetilde{f} \in J_1 (f(x) = 0$ for any $x \in X_{1,1} \setminus U)$, these facts imply that there is $m_0 > n$ such that:

$$\|p_m \Phi_{1,m}(\widetilde{f}) p_m - \Phi_{1,m}(\widetilde{f})\| < 1, \quad m \geq m_0. \tag{1}$$

Fix $m \geq m_0$. Then:

a) *Let j be such that $p_m^j = 0$.*
In this case, (1) implies that:

$$\|p_m^j \Phi_{1,m}^j(\widetilde{f}) p_m^j - \Phi_{1,m}^j(\widetilde{f})\| = \|\Phi_{1,m}^j(\widetilde{f})\| = \|\Phi_{1,m}^{1,j}(f_1)\| < 1.$$

But since $\|\Phi_{1,m}^{1,j}(f_1)\| = \sup_{y \in X_{m,j}} \|f_1 \mid_{SP(\Phi_{1,m}^{1,j})_y} \| = \sup_{y \in X_{m,j}} \|f \mid_{SP(\Phi_{1,m}^{1,j})_y} \|$, one can deduce that

$$SP(\Phi_{1,m}^{1,j})_y \cap F = \phi \;\text{ for all } y \in X_{m,j}$$

$(f(x) = 1$ if $x \in F)$.

b) *Let j be such that $p_m^j = P_m^j$.*
By the definition of p_m it follows that $G_m^j = \phi$, which, by the definition of G_m, is equivalent to:

$$SP(\Phi_{1,m}^j)_y \cap U \neq \phi \;\text{ for all }\; y \in X_m^j.$$

□

Now we shall prove the following result:

Lemma 2.9. *Let $A = \varinjlim(A_n, \Phi_{n,m})$ be as in Lemma 2.8. Then, for any fixed n, i and $\delta > 0$ there is $m_0 > n$ such that the following is true:*

For any $F = \overline{F} \subset X_{n,i}$ and any $m \geq m_0$ we have that any partial map $\Phi_{n,m}^{i,j}$ satisfies either

$$SP(\Phi_{n,m}^{i,j})_y \cap F = \phi \;\textit{ for all } y \in X_{m,j}$$

or

$$SP(\Phi_{n,m}^{i,j})_y \cap B_\delta(F) \neq \phi \;\textit{ for all } y \in X_{m,j}.$$

(*Here we use the standard notation $B_\delta(M) = \{x \in X_{n,i} : \operatorname{dist}(x, M) < \delta\}$ for any subset M of $X_{n,i}$. Note that the result applies to any metric defining the topology of $X_{n,i}$* (*of course, m_0 depends on the chosen metric*).)

Proof. The proof will use the "test functions" introduced in [**Su**] and, of course, the above lemma. Let us denote $X = X_{n,i}$ and $H = \{\chi_{T,\frac{4}{\delta}} : T$ is a closed subset of $X\}$. We recall that for any closed subset ω of $X$and any $\epsilon > 0$ the continuous function $\chi_{\omega,\epsilon} : X \to \mathbb{R}$ is defined in [**Su**] by:

$$\chi_{\omega,\epsilon}(x) = \begin{cases} 1 & \text{if } x \in \omega \\ 1 - \epsilon \cdot \operatorname{dist}(x,\omega) & \text{if } \operatorname{dist}(x,\omega) \leq \frac{1}{\epsilon} \\ 0 & \text{if } \operatorname{dist}(x,\omega) \geq \frac{1}{\epsilon}. \end{cases}$$

Notice that $\operatorname{supp}(\chi_{\omega,\epsilon}) = \overline{B_{\frac{1}{\epsilon}}(\omega)}$.

As observed in the proof of [**L**], Lemma 7.15, H is an equicontinuous family of functions and hence there is a finite subset $H_1 \subset H$ such that $\operatorname{dist}(h, H_1) < \frac{1}{8}$ for any $h \in H$ (see also [**Su**] for a direct proof). Using this, we shall prove now that:

$$(*) \qquad \begin{cases} \text{For any closed subset } C \text{ of } X \text{ there} \\ \text{is } f \in H_1 \text{ such that } f|_C > \frac{3}{4} \text{ and} \\ f|_{X \setminus B_{\frac{3\delta}{4}}(C)} < \frac{1}{8}. \end{cases}$$

Let $D = \overline{B_{\frac{\delta}{2}}(C)}$. By the above observation, for $\chi_{D,\frac{4}{\delta}}$ there is $f \in H_1$, say $f = \chi_{D_1,\frac{4}{\delta}}$ for some $D_1 = \overline{D_1} \subset X$ such that:

$$\text{(i)} \qquad |\chi_{D,\frac{4}{\delta}}(x) - \chi_{D_1,\frac{4}{\delta}}(x)| < \frac{1}{8}, x \in X.$$

Let $x \in C \subset D$. Then $(\chi_{D,\frac{4}{\delta}})(x) = 1$ and hence (i) implies:

$$|1 - \chi_{D_1,\frac{4}{\delta}}(x)| < \frac{1}{8} => f(x) = \chi_{D_1,\frac{4}{\delta}}(x) > \frac{3}{4}.$$

Now let $t \in X \setminus B_{\frac{3\delta}{4}}(C) \subset X \setminus B_{\frac{\delta}{4}}(D)$. Since $\chi_{D,\frac{4}{\delta}}(t) = 0$, (i) implies that:

$$f(t) < \frac{1}{8}.$$

Hence the above statement $(*)$ is proved.

Now, for any $h \in H_1$ let us introduce the notations:

$$F_h = \left\{x \in X : h(x) \geq \frac{3}{4}\right\}$$
$$U_h = \left\{x \in X : h(x) > \frac{1}{8}\right\}.$$

Obviously $F_h = \overline{F_h} \subset U_h = \overset{\circ}{U}_h \subset X$.

Applying Lemma 2.8 for the sets F_h and U_h corresponding to any h in the *finite* set H_1, we get a number $m_0 > n$ such that for all $m \geq m_0, j$ and $h \in H_1$, we have either

$$\text{(ii)} \qquad \begin{cases} SP(\Phi^{i,j}_{n,m})_y \cap F_h = \phi & \text{for all } \ y \in X_{m,j} \\ \quad \text{or} & \\ SP(\Phi^{i,j}_{n,m})_y \cap U_h \neq \phi & \text{for all } \ y \in X_{m,j}. \end{cases}$$

Fix a closed subset F of X. By $(*)$ there is $f \in H_1$ such that $f\mid_F > \frac{3}{4}$ and $f\mid_{X \setminus B_{\frac{3\delta}{4}}(F)} < \frac{1}{8}$.

Obviously:

$$\text{(iii)} \qquad F \subset F_f \subset U_f \subset B_\delta(F).$$

The conclusion follows now from (ii) and (iii). □

The following lemma is a generalization of [**EG 2**], Lemma 2.3.

Lemma 2.10. *Let* $A = \lim\limits_{\rightarrow}(A_n, \Phi_{n,m})$ *be as in Lemma* 2.8. *Then, for any* n, *any finite subset* $F^i_n \subset A^i_n \subset A_n$, *any positive integer* N *and any* $\epsilon > 0$ *there is* $m_0 > n$ *such that any partial map* $\Phi^{i,j}_{n,m}$ *with* $m \geq m_0$ *satisfies either:*

a) rank $(\Phi^{i,j}_{n,m}(P_{n,i})) \geq N \cdot$ rank $(P_{n,i})$ *or*

b) $\Phi^{i,j}_{n,m}$ *is homotopic within the projection* $\Phi^{i,j}_{n,m}(P_{n,i})$ *to a homomorphism* $\Psi^{i,j}_{n,m}$ *with finite dimensional range and:*

$$\|\Phi^{i,j}_{n,m}(f) - \Psi^{i,j}_{n,m}(f)\| < \epsilon$$

for any $f \in F^i_n$.

Proof. Choose a $\delta > 0$ as in the proof of [**EG 2**], Lemma 2.3; that is such that any closed ball in $X_{n,i}$ of radius a is contractible for any $a \leq 2N\delta$ and such that whenever $x, x' \in X_{n,i}$ satisfy $d(x,x') \leq 2N\delta$ ($d(\cdot,\cdot)$ is the canonical metric on $X_{n,i}$) it follows that $\|f(x) - f(x')\| < \epsilon$ for all $f \in F^i_n$. Suppose that rank $(\Phi^{i,j}_{n,m}(P_{n,i})) < N \cdot$ rank $(P_{n,i})$.

Let $m_0 > n$ be the number obtained applying Lemma 2.9 for n, i and $\delta > 0$. We shall use the following standard notation: If $M \subset X$ and (X,d) is a metric space, then for any $\epsilon > 0, B_\epsilon(M) = \{x \in X : d(x,M) < \epsilon\}$. For any fixed $m \geq m_0$ and fixed $y_j \in X_{m,j}$ let $F = X_{n,i} \setminus B_\delta(SP(\Phi^{i,j}_{n,m})_{y_j})$. Obviously F is closed. By Lemma 2.9 it follows that either:

$$\text{(i)} \qquad SP(\Phi^{i,j}_{n,m})_y \cap F = \phi \quad \text{for all } \ y \in X_{m,j}$$

or

$$\text{(ii)} \qquad SP(\Phi^{i,j}_{n,m})_y \cap B_\delta(F) \neq \phi \quad \text{for all } \ y \in X_{m,j}.$$

But since for $y = y_j$ the relation from (ii) is false, it follows that we have (i), which implies that:

$$SP(\Phi^{i,j}_{n,m})_y \subset B_\delta(SP(\Phi^{i,j}_{n,m})_{y_j})$$

for all $y \in X_{m,j}$. Now the proof continues as in the proof of [**EG 2**], Lemma 2.3. $\square$

The next lemma is a generalization of [**Go 1**], Lemma 2.23.

Lemma 2.11. *Suppose that A is the C^*-inductive limit of an inductive system*

$$\Big(A_n = \oplus_{i=1}^{k_n} P_{n,i} C(X_{n,i}, M_{[n,i]}) P_{n,i}, \Phi_{n,m}\Big)$$

with slow dimension growth. Assume that any ideal of A is generated by its projections.

Then, for any n, any finite subset $F_n^i \subset P_{n,i} C(X_{n,i}, M_{[n,i]}) P_{n,i} \subset A_n$, any $\epsilon > 0$ and any integer N there is $m_0 > n$ such that each partial map $\Phi^{i,j}_{n,m}$ with $m \geq m_0$ satisfies either:

1) $\operatorname{rank}(\Phi^{i,j}_{n,m}(P_{n,i})) \geq N \cdot (\dim \quad X_{m,j} + 1)$ *or*
2) $\Phi^{i,j}_{n,m}$ *is homotopic within the projection $\Phi^{i,j}_{n,m}(P_{n,i})$ to a homomorphism $\Psi^{i,j}_{n,m}$ with finite dimensional range and:*

$$\|\Phi^{i,j}_{n,m}(f) - \Psi^{i,j}_{n,m}(f)\| < \epsilon$$

for any $f \in F_{n,i}$.

The proof of the above lemma follows easily from Lemma 2.10.

The next lemma is implicitely contained in the proof of [**EG 2**], Lemma 3.27:

Lemma 2.12 ([**EG 2**]). *Let $\Phi : PC(X, M_k)P \to QC(Y, M_\ell)Q$ be a unital homomorphism between homogeneous C^*-algebras. Let $\epsilon > 0$. Assume that X is a path connected space and that $\{x_1, x_2, \dots, x_n\}$ is $\frac{\epsilon}{2}$-dense in X. Suppose that $\Psi : PC(X, M_k)P \to RC(Y, M_p)R$ is a unital homomorphism between homogenous C^*-algebras defined by:*

$$\Psi(f) = \begin{pmatrix} \Phi(f) & & & \\ & f(x_1) \otimes 1_{m_1} & & \\ & & \ddots & \\ & & & f(x_n) \otimes 1_{m_n} \end{pmatrix}$$

$f \in PC(X, M_k)P$, where $f(x) \in M_{\operatorname{rank}(P)} \cong P(x) M_k P(x)$ with $m_i \geq |SP(\Phi)_y|$, $y \in Y$ (we count with multiplicities). (Here 1_{m_i} is the identity in M_{m_i}.) Then:

$$SPV(\Psi) < \epsilon.$$

The following lemma is essentially contained in the proof of [**Go 1**], Lemma 4.6 (see also [**Da 2**], [**EG 2**]) and it is based on [**DN**].

Let $A = \oplus_{i=1}^{k} P_i M_{n_i}(C(X_i))P_i$, where X_i is a finite, connected CW complex and $P_i \in M_{n_i}(C(X_i))$ is a projection. Let $A' = \oplus_{i=1}^{k} M_{n_i}(C(X_i))$ and $\tilde{A}' = (\oplus_{i=1}^{k} M_{\ell n_i}(C(X_i))) \oplus (\oplus_{i=1}^{k} M_{n_i}) = (\oplus_{i=1}^{k} C'^{i}) \oplus (\oplus_{i=1}^{k} D'^{i})$, where ℓ is a positive integer. Let $\alpha' \oplus \beta' : A' \to \tilde{A}'$ be a unital homomorphism where $\alpha' : A' \to \oplus_{i=1}^{k} C'^{i}, \beta' : A' \to \oplus_{i=1}^{k} D'^{i}$ are homomorphisms such that: $\alpha'^{i,j} = 0, \beta'^{i,j} = 0$ if $i \neq j, [\alpha'^{i,i}] = id \in kk(X_i, X_i)$ (see [**DN**]), $\alpha^{i,i}$ takes trivial projections to trivial projections and $\beta'^{i,i}(f) = f(x_0^i)$, where x_0^i is the base point of X_i. Let $(\alpha' \oplus \beta')(A) \subseteq \tilde{A} = (\oplus_{i=1}^{k} C^i) \oplus (\oplus_{i=1}^{k} D^i) = (\oplus_{i=1}^{k} Q_i M_{\ell n_i}(C(X_i))Q_i) \oplus (\oplus_{i=1}^{k} M_{\mathrm{rank}\,(P_i)}) \subset \tilde{A}'$ where $Q_i = \alpha'^{i,i}(P_i)$ and define $\alpha \oplus \beta : A \to \tilde{A}$ by $(\alpha \oplus \beta)(a) = (\alpha' \oplus \beta')(a), \quad a \in A$.

Lemma 2.13. *With the above notation, let $\Phi : A \to B$ be a homomorphism, where $B = \oplus_{j=1}^{\ell} B_j, B_j = R_j M_{m_j}(C(Y_j))R_j$, where Y_j is a finite, connected CW complex and $R_j \in M_{m_j}(C(Y_j))$ is a projection.*

Suppose that each partial map $\Phi^{i,j} : A_i \to B_j$ satisfies either

1) $\Phi^{i,j} : A_i \to \Phi^{i,j}(P_i)B_j\Phi^{i,j}(P_i)$ *is homotopic to a homomorphism $A_i \to \Phi^{i,j}(P_i)B_j\Phi^{i,j}(P_i)$ with finite dimensional range or*
2) $\mathrm{rank}\,(\Phi^{i,j}(P_i)) \geq 3(\dim Y_j + 1)(\ell + 1) \cdot \mathrm{rank}\,(P_i)$.

Then, there is a homomorphism $\gamma : \tilde{A} \to B$ such that $\gamma \circ (\alpha \oplus \beta)$ is homotopic to Φ and $\Phi(1_A) = (\gamma \circ (\alpha \oplus \beta))(1_A)$.

Proof. It follows from [**EG 2**], Lemma 2.13 and the proof of [**Go 1**], Lemma 4.6 (see also [**Da 2**]) and it is based on [**DN**]. □

Proof of Theorem 2.6. Suppose that A is the C^*-inductive limit of a system:

$$\left(A_n = \oplus_{i=1}^{k_n} P_{n,i} C(X_{n,i} M_{[n,i]}) P_{n,i}, \Phi_{n,m}\right)$$

with slow dimension growth. Combining Lemma 2.11 with Lemma 2.12 and Lemma 2.13 it follows that there is a sequence of positive integers $\ell_1 < \ell_2 < \cdots < \ell_n < \ell_{n+1} < ...$ and homomorphisms $\Psi_n : A_{\ell_n} \to A_{\ell_{n+1}}$ such that for any n:

a) $\Phi_{\ell_n,\ell_{n+1}}$ is homotopic to Ψ_n.
b) $SPV(\Psi_n) < 2^{-n}$.

Since the inductive system $(A_n, \Phi_{n,m})$ has slow dimension growth it follows obviously that $(A_{\ell_n}, \Phi_{\ell_n,\ell_{n+1}})$ has slow dimension growth and, by the above condition a), it follows easily that the system (A_{ℓ_n}, Ψ_n) has also slow dimension growth (note that the spaces $X_{n,i}$ are connected). If we denote $B = \varinjlim(A_{\ell_n}, \Psi_n)$ then b) implies as in the proofs of [**EG 2**], Corollary 2.25 and [**EG 2**], Remark 2.26 (see also the comment after [**Go 1**], Proposition 2.19) that the AH algebra B with slow dimension growth has real rank zero.

To conclude the proof observe that the systems $(A_n, \Phi_{n,m})$ and (A_{ℓ_n}, Ψ_n) are shape equivalent (see a)) (by unital homomorphisms in the unital case since if the $\Phi_{n,m}$'s are unital then the above Ψ_n's can be chosen unital). $\square$

Definition 2.14 ([**EK**], [**Bl 1**]). Two separable C^*-algebras A and B are said to be *shape equivalent* if there are two inductive systems $(A_n, \Phi_{n,m})$ and $(B_n, \Psi_{n,m})$ in $\mathcal{S}$ (the category of separable C^*-algebras and homomorphisms) which are shape equivalent (see Definition 2.5) and $A = \varinjlim(A_n, \Phi_{n,m})$, $B = \varinjlim(B_n, \Psi_{n,m})$.

If A and B are shape equivalent we shall write it in the following way:

$$Sh(A) = Sh(B).$$

The next theorem classifies completely, up to a shape equivalence, a large class of *AH* algebras with slow dimension growth. The invariant is the graded, preordered, scaled group $(\underline{K}(\cdot), \underline{K}(\cdot)^+, \sum(\cdot))$ together with the action of the Bockstein operations on $\underline{K}(\cdot)$ defined in ([**DG**], [**DL 1**]) (see also [**DL 2**], [**Ei**]), where:

$$\underline{K}(A) = K_*(A) \oplus \oplus_{p=2}^{\infty} K_*(A; \mathbb{Z}_p)$$

for any σ-unital C^*-algebra A.

Theorem 2.15. *Let A and A' be AH algebras with slow dimension growth and such that any of their ideals is generated by its projections.*

Then, the following are equivalent:

a) $Sh(A) = Sh(A')$.

b) *There is a graded isomorphism of ordered, scaled groups*

$$(\underline{K}(A), \underline{K}(A)^+, \textstyle\sum(A)) \cong (\underline{K}(A'), \underline{K}(A')^+, \textstyle\sum(A'))$$

which commutes with the Bockstein operations.

Proof. a) $\Rightarrow$ b). This implication follows easily from the definition of $(\underline{K}(\cdot), \underline{K}(\cdot)^+, \sum(\cdot))$ and from the definition of the Bockstein operations ([**DG**], [**DL 2**]) and it is true for more general C^*-algebras A and A'.

b) $\Rightarrow$ a). By Theorem 2.6, it follows that there are *AH* algebras B and B' with slow dimension growth and real rank zero such that:

(i) $$Sh(A) = Sh(B)$$

(ii) $$Sh(A') = Sh(B').$$

Using the hypothesis, the implication a) $\Rightarrow$ b) and [**DG**] it follows that:

$$B \cong B'$$

which obviously implies:

(iii) $$Sh(B) = Sh(B').$$

Results in [**Bl 1**] allow us to deduce from (i), (ii) and (iii) that:

$$Sh(A) = Sh(A').$$

□

3. A characterization theorem.

We shall give now several characterizations of the AH algebras whose ideals are generated by their projections.

Theorem 3.1. *Let* $A = \varinjlim(A_n, \Phi_{n,m})$ *be an* AH *algebra, with* $A_n = \oplus_{i=1}^{k_n} A_n^i$, $A_n^i = P_{n,i}C(X_{n,i,}, M_{[n,i]})P_{n,i}$, *where* $X_{n,i}$ *are connected, finite CW complexes and* $P_{n,i} \in C(X_{n,i}M_{[n,i]})$ *are projections. Then the following are equivalent:*

a) *Any ideal of* A *is generated by its projections.*

b) *For any fixed* n *and any fixed* $F = \overline{F} \subset U = \overset{\circ}{U} \subset SP(A_n) = \sqcup_{i=1}^{k_n} X_{n,i}$ *there is* $m_0 > n$ *such that for any* $m \geq m_0$ *any partial map* $\Phi_{n,m}^j : A_n \to A_m^j$ *satisfies either:*

$$SP(\Phi_{n,m}^j)_y \cap F = \phi \quad \text{for all } \ y \in X_{m,j}$$

or

$$SP(\Phi_{n,m}^j)_y \cap U \neq \phi \quad \text{for all } \ y \in X_{m,j}.$$

c) *For any fixed* n, i *and* $\delta > 0$ *there is* $m_0 > n$ *such that the following is true:*

For any $F = \overline{F} \subset X_{n,i}$ *and any* $m \geq m_0$ *we have that any partial map* $\Phi_{n,m}^{i,j}$ *satisfies either:*

$$SP(\Phi_{n,m}^{i,j})_y \cap F = \phi \quad \text{for all } \ y \in X_{m,j}$$

or

$$SP(\Phi_{n,m}^{i,j})_y \cap B_\delta(F) \neq \phi \quad \text{for all } \ y \in X_{m,j}.$$

(*Here we used the standard notation* $B_\delta(M) = \{x \in X_{n,i} : \mathrm{dist}\,(x, M) < \delta\}$ *for any subset* M *of* $X_{n,i}$.)

d) *Any ideal of* A *has a countable approximate unit consisting of projections.*

e) *For any ideal* I *of* A *we have:*

For any integer n, *any* $\epsilon > 0$ *and any* $x \in A_n \cap I$ *there is* $m > n$ *and a projection* $p \in A_m \cap I$ *such that:*

$$\|\Phi_{n,m}(x) - p \cdot \Phi_{n,m}(x)\| \leq \epsilon.$$

f) *For any ideal I of A we have:*

For any integer n, any $\epsilon > 0$ and any $x \in A_n \cap I$ there is $m > n$ and a projection $p \in A_m \cap I$ such that:

$$\|\Phi_{n,m}(x) - p \cdot \Phi_{n,m}(x) \cdot p\| \leq \epsilon.$$

(*Above we used the notation $A_k \cap I = \{y \in A_k : \Phi_{k,\infty}(y) \in I\}$.*)

Proof. a) $\Rightarrow$ b) follows from Lemma 2.8.

a) $\Rightarrow$ c) follows from Lemma 2.9.

b) $\Rightarrow$ d) Fix I an ideal in $A, n, \epsilon > 0$ and an element $x \in A_n \cap I$. Define:

$$F = \{t \in SP(A_n) : \|x(t)\| \geq \epsilon\}$$
$$U = \{t \in SP(A_n) : \|x(t)\| \neq 0\}.$$

Obviously $F = \overline{F} \subset U = \overset{\circ}{U} \subset SP(A_n)$. Then, by hypothesis, there is $m_0 > n$ such that for any $m \geq m_0$ any partial map $\Phi^j_{n,m} : A_n \to A^j_m$ satisfies either

(i) $$SP(\Phi^j_{n,m})_y \cap F = \phi \quad \text{for all} \;\; y \in X_{m,j}$$

or

(ii) $$SP(\Phi^j_{n,m})_y \cap U \neq \phi \quad \text{for all} \;\; y \in X_{m,j}.$$

Let $G_k = \overline{G}_k \subset SP(A_k)$ be such that $A_k \cap I = \{f \in A_k : f|_{G_k} = 0\}$, $k \geq 1$. Define $p_m = \oplus_{j=1}^{k_m} p^j_m \in \oplus_{j=1}^{k_m} A^j_m = A_m$ for any $m \geq 1$ by:

$$p^j_m = \begin{cases} P_{m,j} & \text{if } \; G^j_m = \phi \\ 0 & \text{if } \; G^j_m \neq \phi \end{cases}$$

(here, obviously, G^j_m is the component of G_m in $X_{m,j}$). It is clear that p_m is a projection in $A_m \cap I$. Denote for $m \geq m_0$:

$$d^j_m = \Phi^j_{n,m}(x) - p^j_m \Phi^j_{n,m}(x)$$

for any $1 \leq j \leq k_m$. We shall prove that $\|d^j_m\| \leq \epsilon$ for all j and all $m \geq m_0$. Let us fix now $m \geq m_0$.

1) *Let j be such that $p^j_m = P_{m,j}$.*

In this case $d^j_m = 0$ and hence $\|d^j_m\| = 0$.

2) *Let j be such that $p^j_m = 0$.*

Then $d^j_m = \Phi^j_{n,m}(x)$. We have two possibilities:

If (i) is true, then:

$$\|d^j_m\| = \|\Phi^j_{n,m}(x)\| = \sup_{y \in X_{m,j}} \left\| x \mid_{SP(\Phi^j_{n,m})_y} \right\| \leq \epsilon.$$

If (ii) is true, then, since obviously:

$$U \subset SP(A_n) \setminus G_n := G^c_n$$

it follows that:

$$SP(\Phi^j_{n,m})_y \cap G^c_n \neq \phi \quad \text{for all } \ y \in SP(A^j_m) = X_{m,j}.$$

But now it is not difficult to prove that if $z \in SP(A^j_m)$, then:

$$SP(\Phi^j_{n,m})_z \cap G^c_n \neq \phi \Rightarrow z \in SP(A^j_m) \setminus G^j_m := (G^j_m)^c.$$

(Indeed, let $t \in SP(\Phi^j_{n,m})_z \cap G^c_n$. Then there is $g \in A_n \cap I$ such that $g(t) \neq 0$. Hence:

$$\|\Phi^j_{n,m}(g)(z)\| = \left\| g \mid_{SP(\Phi^j_{n,m})_z} \right\| > 0$$

since $t \in SP(\Phi^j_{n,m})_z$. But $g \in A_n \cap I$ and $\Phi_{n,m}(A_n \cap I) \subset A_m \cap I$. All these things imply that $z \in (G^j_m)^c$.)

Combining the last relations we get:

$$SP(A^j_m) = X_{m,j} \subset (G^j_m)^c \Leftrightarrow G^j_m = \phi \Leftrightarrow p^j_m = P_{m,j}.$$

Hence, we are in the case 1) when $\|d^j_m\| = 0$. This ends the proof of b) $\Rightarrow$ d).

d) $\Rightarrow$ e), e) $\Rightarrow$ a), d) $\Rightarrow$ f), c) $\Rightarrow$ b) and f) $\Rightarrow$ a) are obvious. The proof of the Theorem is completed. □

Remarks 3.2. a) Let us note that part c) in Theorem 3.1 is independent of the metric defining the topology of $X_{n,i}$.

b) Theorem 3.1 remains true if the spaces $X_{n,i}$ are arbitrary compact, connected, metrizable topological spaces (the proof is the same).

4. Stable rank one.

Our objective is the following result:

Theorem 4.1. *Let A be an AH algebra with slow dimension growth and such that any ideal of A is generated by its projections.*

Then A has stable rank one ($tsr\,(A) = 1$).

Note that if in this theorem we drop the slow dimension growth condition, then the result is not true in general, as follows from [**V**]. The above theorem generalizes results of Dadarlat-Nagy-Némethi-Pasnicu [**DNNP**], Blackadar-Dadarlat-Rørdam [**BDR**], when A is simple and of Elliott-Gong [**EG 2**] when A has real rank zero. For the proof of the above theorem we shall need Theorem 3.1, Lemma 2.11 and the following lemmas:

Lemma 4.2. *Let X be a compact Hausdorff space and let a be a non-invertible element of $A = C(X, M_n)$. Let $F = \{x \in X : a(x)$ is not invertible in $M_n\}$ and let $\epsilon > 0$.*

Then, there are $s, F_i = \overline{F}_i \subset U_i = \overset{\circ}{U}_i \subset X, b_i \in A$ and projections $p_i, q_i \in A$ $(1 \leq i \leq s)$ such that:

1) $F = \cup_{i=1}^{s} F_i$.
2) $p_i(x), q_i(x) \neq 0, x \in X,\ 1 \leq i \leq s$.
3) $b_i(x)p_i(x) = q_i(x)b_i(x) = 0, x \in U_i,\ 1 \leq i \leq s$.
4) $\|a - b_i\| < \epsilon,\ 1 \leq i \leq s$.

Proof. Let $x \in F$. Since $\det a(x) = 0$, it is not difficult to deduce that there are $x \in U_x = \overset{\circ}{U}_x \subset X, b_x \in A$ and projections $p_x, q_x \in C(X, M_n)$ such that:

$$p_x(t), q_x(t) \neq 0, t \in X$$

$$b_x(t)p_x(t) = q_x(t)b_x(t) = 0, \quad t \in U_x$$

$$\|a - b_x\| < \epsilon.$$

Choose $x \in W_x = \overset{\circ}{W}_x$ such that $\overline{W_x} \subset U_x$.

We make such a construction for any $x \in F$. Since $F \subset \cup_{x \in F} W_x$ and F is compact, it follows that there are $x_1, x_2, \dots, x_s \in F$ such that:

$$F \subset \cup_{i=1}^{s} W_{x_i} (\subset \cup_{i=1}^{s} \overline{W_{x_i}})$$

and hence

$$F = \cup_{i=1}^{s} (F \cap \overline{W_{x_i}}).$$

Denote $F_i = F \cap \overline{W_{x_i}},\ 1 \leq i \leq s$. Since $F_i = \overline{F_i}$ and $F_i \subset \overline{W_{x_i}} \subset U_{x_i}$ for each i, we may define $U_i = U_{x_i}, b_i = b_{x_i}, p_i = p_{x_i}, q_i = q_{x_i} (1 \leq i \leq s)$ and the proof is over. □

Lemma 4.3. *Let $\Phi : A = C(X, M_n) \to B = C(Y, M_m)$ be a homomorphism, where X, Y are compact Hausdorff spaces. Let $a \in A$ and let $\mathcal{W}$ be an open cover of X such that for every $W \in \mathcal{W}$ there are positive elements $p_W, q_W \in A$ such that for any $x \in W, p_W(x)$ and $q_W(x)$ are non-zero projections and:*

$$a(x)p_W(x) = q_W(x)a(x) = 0.$$

Then, there is an open cover $\mathcal{Z}$ of Y such that for every $Z \in \mathcal{Z}$ there are positive elements $P_Z, Q_Z \in \Phi(A)$ such that for any $t \in Z, P_Z(t)$ and $Q_Z(t)$ are projections with

$$\Phi(a)(t)P_Z(t) = Q_Z(t)\Phi(a)(t) = 0$$

and

$$\operatorname{rank} P_Z(t),\ \ \operatorname{rank} Q_Z(t) \geq |SP(\Phi)_t| = \frac{\operatorname{rank} \Phi(1_A)}{\operatorname{rank} 1_A}$$

(*count multiplicities*).

Proof. Let $t \in Y$. Suppose that $y_1, y_2, \dots, y_k$ are the distinct elements of $SP(\Phi)_t$. Choose $W_1, W_2, \dots, W_k \in \mathcal{W}$ such that $y_i \in W_i$ and choose also open sets U_i with $y_i \in U_i \subset W_i$ such that:

$$\overline{U_i} \cap (\overline{\cup_{j \neq i} U_j}) = \phi, \quad 1 \leq i \leq k.$$

Let Z be an open neighborhood of t in Y such that:

$$SP(\Phi)_t \subset \cup_{i=1}^k U_i, \quad t \in Z.$$

Let $\chi_i : X \to [0,1]$ be continuous maps such that $\chi_i(x) = 1$ for any $x \in U_i$ and $\chi_i(x) = 0$ for any $x \in \cup_{j \neq i} U_j$ $(1 \leq i \leq k)$.

Define $P_Z, Q_Z \in B$ by:

$$P_Z = \Phi\left(\sum_{i=1}^k \chi_i p_{W_i}\right)$$

$$Q_Z = \Phi\left(\sum_{i=1}^k \chi_i q_{W_i}\right).$$

Obviously, $0 \leq P_Z, Q_Z \in \Phi(A)$ and for any $t \in Z$ it is clear that $P_Z(t)$ and $Q_Z(t)$ are projections $\left(\text{since } \sum_{i=1}^k \chi_i p_i \text{ and } \sum_{i=1}^k \chi_i q_i \text{ are projections on } SP(\Phi)_t\right)$ and:

$$\Phi(a)(t) P_Z(t) = Q_Z(t) \Phi(a)(t) = 0$$

$\Bigg($since if $x \in SP(\Phi)_t \subset \cup_{i=1}^k U_i \Rightarrow x \in U_j$ for some j and then:

$$\begin{aligned} a(x) \cdot \left(\sum_{i=1}^k \chi_i p_{W_i}\right)(x) &= a(x)\chi_j(x) p_{W_j}(x) = a(x) p_{W_j}(x) = 0 \\ &= q_{W_j}(x) a(x) = \left(\sum_{i=1}^k \chi_i p_{W_i}\right)(x) \cdot a(x) \Bigg) \end{aligned}$$

and also:

$$\begin{aligned} \operatorname{rank} P_Z(t) &= \sum_{i=1}^k \left(\sum_{x \in SP(\Phi)_t \cap U_i} \operatorname{rank} p_{W_i}(x)\right) \\ &\geq \sum_{i=1}^k |SP(\Phi)_t \cap U_i| = |SP(\Phi)_t| = \frac{\operatorname{rank} \Phi(1_A)}{\operatorname{rank} 1_A} \end{aligned}$$

and similarly:

$$\operatorname{rank} Q_Z(t) \geq |SP(\Phi)_t|$$

(count multiplicities). $\square$

The next lemma is the Selection principle from [**DNNP**]:

Lemma 4.4 ([**DNNP**], Proposition 3.2)**.** *Let X be a Hausdorff compact space, let $k' \geq k \geq 1$ be integers, let $\mathcal{W}$ be an open cover of X and assume that for each $W \in \mathcal{W}$ there is given a continuous projection valued map $p_W : W \to M_n$ such that* rank $p_W(x) \geq k'$ *for $x \in W$. If $\dim(X) \leq 2(k'-k)-1$ then there is a continuous projection valued map $p : X \to M_n$ such that for $x \in X$:*

$$\operatorname{rank} p(x) \geq k$$

$$p(x) \leq \vee\{p_W(x) : W \in \mathcal{W}, x \in W\}.$$

Lemma 4.5 ([**DNNP**], Lemma 3.3)**.** *Let B be a unital C^*-algebra and let*

$$k \geq \max(tsr\,(B), csr\,(B)).$$

Then for any positive integer m and any $a \in M_m(B)$, the matrix $\begin{pmatrix} a & 0 \\ a & 0_k \end{pmatrix}$ belongs to the closure of $GL(m+k, B)$.

Proof of Theorem 4.1. We shall prove first the theorem in the case when $A = \varinjlim(A_n, \Phi_{n,m}), A_n = \oplus_{i=1}^{k_n} A_n^i, A_n^i = C(X_{n,i}, M_{[n,i]})$, where k_n, $[n,i]$ are positive integers and $X_{n,i}$ are connected, finite CW complexes.

Fix a non-invertible element a in some A_n^i and fix also $\epsilon > 0$. Define $F = \{x \in SP(A_n^i) : a(x) \text{ is not invertible}\} = \{x \in SP(A_n^i) : \det a(x) = 0\}$. By Lemma 4.2 there are $s, F_\ell = \overline{F_\ell} \subset U_\ell = \overset{\circ}{U}_\ell \subset SP(A_n^i), b_\ell \in A_n^i$ and projections $p_\ell, q_\ell \in A_n^i$, $1 \leq \ell \leq s$ such that $F = \cup_{\ell=1}^s F_\ell$ and:

$$p_\ell(x), q_\ell(x) \neq 0,\, x \in SP(A_n^i), \quad 1 \leq \ell \leq s \tag{1}$$

$$b_\ell(x) \cdot p_\ell(x) = q_\ell(x) \cdot b_\ell(x) = 0, \quad x \in U_\ell, \quad 1 \leq \ell \leq s \tag{2}$$

$$\|a - b_\ell\| < \epsilon, \quad 1 \leq \ell \leq s. \tag{3}$$

By Theorem 3.1, there is $m_0 > n$ such that for any $m \geq m_0$ and any $1 \leq \ell \leq s$ any partial map $\Phi_{n,m}^{i,j} : A_n^i \to A_m^j$ $(1 \leq j \leq k_m)$ satisfies either:

$$SP(\Phi_{n,m}^{i,j})_y \cap F_\ell = \phi \quad \text{for all } \ y \in X_{m,j}$$

or

$$SP(\Phi_{n,m}^{i,j})_y \cap U_\ell \neq \phi \quad \text{for all } \ y \in X_{m,j}.$$

Using this and arguing by contradiction it is easy to see that for any $m \geq m_0$ and for any j there is $1 \leq k \leq s$ such that either:

$$SP(\Phi_{n,m}^{i,j})_y \cap F = \phi \quad \text{for any } \ y \in SP(A_m^j) \tag{4}$$

or

$$SP(\Phi_{n,m}^{i,j})_y \cap U_k \neq \phi \quad \text{for any } \ y \in SP(A_m^j). \tag{5}$$

Fix such m, j and k.

If (4) is true, then it is easy to see that $\Phi^{i,j}_{n,m}(a)$ is invertible in $\Phi^{i,j}_{n,m}(1_{A^i_n})A^j_m\Phi^{i,j}_{n,m}(1_{A^i_n})$.

Let us assume now that (5) is true. Then, using also (1) and (2) and working as in the proof of [**DNNP**], Lemma 2.2, we get an open cover $\mathcal{W}$ of $SP(A^j_m)$ such that for every $W \in \mathcal{W}$ there are positive elements p_W, q_W of A^j_m such that for any $x \in W, p_W(x)$ and $q_W(x)$ are non-zero projections with:

$$\Phi^{i,j}_{n,m}(b_k)(x)p_W(x) = q_W(x)\Phi^{i,j}_{n,m}(b_k)(x) = 0.$$

Denote $c_j = \Phi^{i,j}_{n,m}(b_k)$. By Lemma 2.11 we have that there is $r > m$ such that any partial map $\Phi^{j,s}_{m,r}$ satisfies either:

$$(6)\qquad \begin{cases} \|\Phi^{j,s}_{m,r}(c_j) - \Psi^{j,s}_{m,r}(c_j)\| < \epsilon \\ \text{where } \ \Psi^{j,s}_{m,r} : A^j_m \to \Phi^{j,s}_{m,r}(1_{A^j_m})A^s_r\Phi^{j,s}_{m,r}(1_{A^j_m}) \\ \text{is some homomorphism with finite dimensional range} \end{cases}$$

or:

$$(7)\qquad \operatorname{rank} \Phi^{j,s}_{m,r}(1_{A^j_m}) \geq \operatorname{rank} 1_{A^j_m} \cdot \left(\frac{3}{2}\dim\, X_{r,s} + 2\right).$$

Fix s and r. If (6) is true then, obviously, $\Phi^{j,s}_{m,r}(c_j)$ can be approximated within ϵ by an invertible element of $\Phi^{j,s}_{m,r}(1_{A^j_m})A^s_r\Phi^{j,s}_{m,r}(1_{A^j_m})$.

If (7) is true, then, by Lemma 4.3 it follows that there is an open cover $\mathcal{Z}$ of $SP(A^s_r)$ such that for any $Z \in \mathcal{Z}$ there are positive elements $P_Z, Q_Z \in \Phi^{j,s}_{m,r}(A^j_m)$ with the property that for any $t \in Z, P_Z(t), Q_Z(t)$ are projections with:

$$\Phi^{j,s}_{m,r}(c_j)(t)P_Z(t) = Q_Z(t)\Phi^{j,s}_{m,r}(c_j)(t) = 0$$

$$\operatorname{rank} P_Z(t),\ \ \operatorname{rank} Q_Z(t) \geq \frac{3}{2}\dim\, X_{r,s} + 2.$$

Using now Lemma 4.4 (the Selection principle) it follows that there are projections p and q in A^s_r such that:

$$(8)\qquad \operatorname{rank} p, \operatorname{rank} q \geq \dim\, X_{r,s} + \frac{3}{2}$$

$$(9)\qquad \Phi^{j,s}_{m,r}(c_j)p = q\Phi^{j,s}_{m,r}(c_j) = 0$$

$$(10)\qquad p, q \leq \Phi^{j,s}_{m,r}(1_{A^j_m}).$$

Using (8) and stability results for vector bundles (see [**H**]) and passing eventually to subprojections it follows that there are trivial projections $\overline{p}$ and $\overline{q}$ in A^s_r (i.e., unitarily equivalent to a constant projection in $C(X_{r,s}, M_{[r,s]}) = A^s_r$) such that $\overline{p} \leq p, \overline{q} \leq q$ and:

$$(11)\qquad \dim\, X_{r,s} + 1 \geq \operatorname{rank} \overline{p} = \operatorname{rank} \overline{q} \geq \left[\frac{\dim\, X_{r,s} + 1}{2}\right] + 1.$$

Since by [**N**] and [**R**]:

$$\left[\frac{\dim\, X_{r,s}+1}{2}\right]+1 \geq \max\{tsr\, C(X_{r,s}), csr\, C(X_{r,s})\}$$

it follows that:

$$\operatorname{rank} \overline{p} = \operatorname{rank} \overline{q} \geq \max\{tsr\, C(X_{r,s}), csr\, C(X_{r,s})\}.$$

Since $\overline{p}$ and $\overline{q}$ are trivial projections, (10) and (11) imply that:

$$[\overline{p}] = [\overline{q}] \in K_0(A_r^s) = K_0(\Phi_{m,r}^{j,s}(1_{A_m^j})A_r^s\Phi_{m,r}^{j,s}(1_{A_m^j})). \tag{12}$$

We have by (11):

$$\operatorname{rank} \overline{p} = \operatorname{rank} \overline{q} \geq \left\langle \frac{\dim\, X_{r,s}}{2} \right\rangle$$

($\langle\cdot\rangle$ denotes the least integer $\geq$) and by (7) and (11):

$$\begin{aligned}
&\operatorname{rank}(\Phi_{m,r}^{j,s}(1_{A_m^j}) - \overline{p}) = \operatorname{rank} \Phi_{m,r}^{j,s}(1_{A_m^j}) - \operatorname{rank} \overline{p} \\
&\geq \frac{3}{2}\dim\, X_{r,s} + 2 - (\dim\, X_{r,s}+1) = \frac{\dim\, X_{r,s}}{2} + 1 \geq \left\langle \frac{\dim\, X_{r,s}}{2} \right\rangle.
\end{aligned}$$

Using also (12), (10) and stability results for vector bundles in [**H**] it follows that:

$$\overline{p} = u\overline{q}u^* \ \text{ for some unitary } u \text{ in } \ \Phi_{m,r}^{j,s}(1_{A_m^j})A_r^s\Phi_{m,r}^{j,s}(1_{A_m^j}). \tag{13}$$

Since $\Phi_{m,r}^{j,s}(c_j)\overline{p} = \overline{q}\Phi_{m,r}^{j,s}(c_j) = 0$ (see (9)), using (13) we get that:

$$\left(u^*\Phi_{m,r}^{j,s}(c_j)\right)\overline{p} = \overline{p}\left(u^*\Phi_{m,r}^{j,s}(c_j)\right) = 0.$$

Since $\overline{p}, u^*\Phi_{m,r}^{j,s}(c_j) \in \Phi_{m,r}^{j,s}(1_{A_m^j})A_r^s\Phi_{m,r}^{j,s}(1_{A_m^j})$, using Lemma 4.5 as in the past part of the proof of [**DNNP**], Theorem 3.6, we can conclude that:

$$\begin{aligned}
u^*\Phi_{m,r}^{j,s}(c_j) \in \overline{GL(\Phi_{m,r}^{j,s}(1_{A_m^j})A_r^s\Phi_{m,r}^{j,s}(1_{A_m^j}))} \Leftrightarrow \\
\Phi_{m,r}^{j,s}(c_j) \in \overline{GL(\Phi_{m,r}^{j,s}(1_{A_m^j})A_r^s\Phi_{m,r}^{j,s}(1_{A_m^j}))},
\end{aligned}$$

(see (13)).

In conclusion, A has stable rank one (see (3)).

Let us consider now the general case, when:

$$A = \varinjlim(A_n, \Phi_{n,m})$$

where $A_n = \oplus_{i=1}^{k_n} P_{n,i}C(X_{n,i}, M_{[n,i]})P_{n,i}$ with $X_{n,i}$ finite, connected CW complexes and $P_{n,i} \in C(X_{n,i}M_{[n,i]})$ projections. Then, as in [**EG 2**], 4.24, we use the proof of [**EG 2**], Lemma 2.13 to construct an AH algebra

$\tilde{A} = \varinjlim(\tilde{A}_n, , \ \tilde{\Phi}_{n,m})$, where $\tilde{A}_n$ is a direct sum of matrix algebras over the same CW complexes $X_{n,i}, A \subset \tilde{A}$ and:

$$A = \overline{\cup_{n=1}^{\infty} P_n \tilde{A} P_n}$$

for some increasing sequence of projections $P_1 \leq P_2 \leq \ldots \leq P_n \leq \ldots$. In fact, we may suppose that A_n is a corner subalgebra of $\tilde{A}_n$ and that $\Phi_{n,m} = \tilde{\Phi}_{n,m|A_n}$. It is clear now that $\tilde{A} = \varinjlim(\tilde{A}_n, \tilde{\Phi}_{n,m})$ has slow dimension growth and that by Theorem 3.1 it follows that any ideal of $\tilde{A}$ is generated by its projections (since any ideal of A has the same property). On the other hand, due to the special form of the connecting homomorphisms $\tilde{\Phi}_{n,m}$, one can easily construct an isomorphsim between $A \otimes \mathcal{K}$ and $\tilde{A} \otimes \mathcal{K}$ ($\mathcal{K}$ = the compact operators acting on a separable, infinite dimensional Hilbert space). Since $tsr\,(\tilde{A}) = 1$ (by the first part of the proof), a theorem of Rieffel [**R**], Theorem 3.6 implies that $tsr\,(A) = 1$. □

5. Other nonstable K-theoretical results.

In this section we shall prove other nonstable K-theory results for the AH algebras considered in Theorem 4.1.

Theorem 5.1. *Let A be an AH algebra with slow dimension growth and such that any ideal is generated by its projections.*
Then:

a) $K_0(A)$ *is weakly unperforated in the sense of Elliott* ([**Ell 2**]).
b) *If furthermore, all the connecting homomorphisms in the inductive system with slow dimension growth whose limit is A are unital then A has* (SC) (*see* [**P 2**]), *i.e., if p and q are projections in A such that $\tau(p) < \tau(q)$ for any tracial state τ of A, then p is Murray-von Neumann equivalent to a proper subprojection of q.*

The above theorem generalizes results of Dadarlat-Némethi [**DN**], Martin-Pasnicu [**MP**] and Blackadar [**Bl 2**].

To prove the above theorem we need, among other things, the following result:

Proposition 5.2 (see [**P 2**] and also [**MP**]). *Let $A = \varinjlim(A_n, \Phi_n)$ and $B = \varinjlim(B_n, \Psi_n)$, where A_n, B_n are arbitrary unital C^*-algebras and the connecting homomorphisms Φ_n, Ψ_n are unital.*

Suppose that there is an EP-commutative diagram (*see* [**MP**, 2.3]) *with unital homomorphisms α_n and $\beta_n (n \geq 1)$:*

$$\begin{array}{ccccccccc} A_1 & \xrightarrow{\Phi_1} & A_2 & \xrightarrow{\Phi_2} & A_3 & \xrightarrow{\Phi_3} & A_4 & \xrightarrow{\Phi_4} & \cdots \\ {\scriptstyle\alpha_1}\searrow & {\scriptstyle\beta_1}\nearrow & {\scriptstyle\alpha_2}\searrow & {\scriptstyle\beta_2}\nearrow & {\scriptstyle\alpha_3}\searrow & {\scriptstyle\beta_3}\nearrow & & & \\ & B_1 & \xrightarrow{\Psi_1} & B_2 & \xrightarrow{\Psi_2} & B_3 & \xrightarrow{\Psi_3} & \cdots & \end{array}$$

(*note that this happens if e.g., the above diagram is commutative at the level of homotopy*).

Then:

a) $T(A) \neq \phi \Leftrightarrow T(B) \neq \phi$.

b) *A has* $(SC) \Leftrightarrow B$ *has* (SC).

If furthermore the above diagram is a stably EP-commutative diagram (*that is, after taking the tensor product with* M_n *for any* n*, it is still an EP-commutative diagram*) *then:*

c) *A has cancellation* $\Leftrightarrow$ *B has cancellation.*

d) $K_0(A)$ *is weakly unperforated in the sense of Elliott* $\Leftrightarrow K_0(B)$ *is weakly unperforated in the sense of Elliott.*

a), b) and c) appeared in [**P 2**], Proposition 2.5. d) follows from the fact that the preordered groups $(K_0(A), K_0(A)_+)$ and $(K_0(B), K_0(B)_+)$ are isomorphic. Obviously, we need in fact the homomorphisms to be unital in the diagram only for a) and b).

Proof of Theorem 5.1. The proof is inspired by [**P 2**] and [**MP**]. Assume that A is the C^*-inductive limit of an inductive system:

$$(A_n = \oplus_{i=1}^{k_n} A_n^i, \Phi_{n,m})$$

$A_n^i = P_{n,i}C(X_{n,i}, M_{[n,i]})P_{n,i}$, with slow dimension growth. Using Lemma 2.11 it follows that there is a sequence of positive integers $\ell_1 < \ell_2 < ... < \ell_n < ...$ and homomorphisms $\Psi_n : A_{\ell_n} \to A_{\ell_{n+1}}$ such that $\Phi_{\ell_n,\ell_{n+1}}$ is homotopic to Ψ_n and each partial map

$$(\Psi_n)^{i,j} : A_{\ell_n}^i \to (\Psi_n)^{i,j}(1_{A_{\ell_n}^i})A_{\ell_{n+1}}^j(\Psi_n)^{i,j}(1_{A_{\ell_n}^i})$$

induced by Ψ_n satisfies either:

$$\mathrm{rank}\,((\Psi_n)^{i,j}(P_{\ell_n,i})) \geq (\dim \ X_{\ell_{n+1},j} + 1)\mathrm{rank}\,(P_{\ell_n,i})$$

or

$$(\Psi_n)^{i,j} \text{ has finite dimensional range.}$$

If we denote $B = \varinjlim(A_{\ell_n}, \Psi_n)$, then obviously A and B are shape equivalent (and, moreover, if the $\Phi_{n,m}$'s are unital then all the Ψ_n's above can be chosen also unital).

Now, by Proposition 5.2, it is enough to prove that $K_0(B)$ is weakly unperforated in the sense of Elliott and that (in the unital case) B has (SC). But this can be shown using stability results for vector bundles ([**H**]) as in [**P 2**] (see also [**MP**]). $\square$

Remarks 5.3. We know that any AH algebra with slow dimension growth and real rank zero is isomorphic to an AH algebra over spaces of dimension ≤ 3 and with real rank zero ([**Da 2**], [**Go 1**]). Therefore, using also Theorem 2.6 and Proposition 5.2, the proof of the above Theorem reduces to the proof of the fact that any AH algebra over spaces of dimension ≤ 3 and with real rank zero has the required properties. But this is a standard fact. We prefered the other proof of Theorem 5.1 beause it is more "elementary".

References

[Bl 1] B. Blackadar, *Shape theory for C^*-algebras*, Math. Sand., **56** (1985), 249-275.

[Bl 2] ———, *Matricial and ultramatricial topology*, Operator Algebras, Mathematical Physics and Low Dimensional Topology (editors R.H. Herman, B. Tambay), A.K. Peters, Wellesley, Massachusetts, (1993), 11-38.

[BDR] B. Blackadar, M. Dadarlat and M. Rørdam, *The real rank of inductive limit C^*-algebras*, Math. Scand., **69** (1991), 211-216.

[Br] O. Bratteli, *Inductive limits of finite-dimensional C^*-algebras*, Trans. Amer. Math. Soc., **171** (1972), 195-234.

[BP] L.G. Brown and G.K. Pedersen, *C^*-algebras of real rank zero*, J. Funct. Anal., **99** (1991), 131-149.

[Da 1] M. Dadarlat, *On homomorphisms of certain C^*-algebras*, preprint, 1986.

[Da 2] ———, *Reduction to dimension three of local spectra of real rank zero C^*-algebras*, J. Reine Angew. Math., **460** (1995), 189-212.

[DG] M. Dadarlat and G. Gong, *A classification result for approximately homogeneous C^*-algebras of real rank zero*, GAFA, Geom. Funct. Anal., **7** (1997), 646-711.

[DL 1] M. Dadarlat and T. Loring, *Classifying C^*-algebras via ordered,* mod-p *K-theory*, Math. Ann., **305** (1996), 601-616.

[DL 2] ———, *A universal multi-coefficient theorem for the Kasparov groups*, Duke Math. J., **84** (1996), 355-377.

[DNNP] M. Dadarlat, G. Nagy, A. Némethi and C. Pasnicu, *Reduction of topological stable rank in inductive limits of C^*-algebras*, Pacific J. Math., **153**(**2**) (1992), 267-276.

[DN] M. Dadarlat and A. Némethi, *Shape theory and* (*connective*) *K-theory*, J. Operator Theory, **23** (1990), 207-291.

[Di] J. Dixmier, *On some C^*-algebras considered by Glimm*, J. Funct. Anal., **1** (1967), 182-203.

[Eff] E.G. Effros, *On the structure of C^*-algebras: Some old and some new problems*, in 'Operator Algebras and Applications', Prop. Symp. Pure Math., **38**, part 1, A.M.S., Providence, R.I. (1982), 19-34.

[EK] E.G. Effros and J. Kaminker, *Homotopy continuity and shape theory for C^*-algebras*, Geometric methods in operator algebras, U.S.-Japan seminar at Kyoto 1983, Pitman, 1985.

[Ei] S. Eilers, *A complete invariant for AD algebras with bounded dimension drop in K_1*, J. Funct. Anal., **139** (1996), 325-348.

[Ell 1] G.A. Elliott, *On the classification of inductive limits of semi-simple finite dimensional algebras*, J. Algebra, **38** (1976), 29-44.

[Ell 2] ______, *Dimension groups with torsion*, Internat. J. of Math., **1** (1990), 361-380.

[Ell 3] ______, *On the classification of C^*-algebras of real rank zero*, J. Reine Angew. Math., **443** (1993), 179-219.

[Ell 4] ______, *A classification of certain simple C^*-algebras*, Quantum Non-Commutative Analysis (editors H. Araki et al.), Kluwer, Dordrecht, (1993), 373-385.

[Ell 5] ______, *A classification of certain simple C^*-algebras*, II, J. Ramanujan Math. Soc., **12**(**1**) (1997), 97-134.

[Ell 6] ______, *The classfication problem for amenable C^*-algebras*, Proceedings of the International Congress of Mathematicians, Zürich, Switzerland 1994, Birkhäuser Verlag, Basel, Switzerland (1995), 922-932.

[EG 1] G.A. Elliott and G. Gong, *On inductive limits of matrix algebras over the two-torus*, Amer. J. of Math., **118** (1996), 263-290.

[EG 2] ______, *On the classification of C^*-algebras of real rank zero*, II, Annals of Math., **144** (1996), 497-610.

[EGL] G.A. Elliott, G. Gong and L. Li, *On the classification of simple inductive limit C^*-algebras*, II, preprint.

[EGLP 1] G.A. Elliott, G. Gong, H. Lin and C. Pasnicu, *Abelian C^*-subalgebras of C^*-algebras of real rank zero and inductive limit C^*-algebras*, Duke Math. J., **85** (1996), 511-554.

[EGLP 2] ______, *Homomorphisms, homotopies and approximations by circle algebras*, C.R. Math. Rep. Acad. Sci. Canada, **XVI** (1994), 45-50.

[Gl] J. Glimm, *On a certain class of operator algebras*, Trans. Amer. Math. Soc., **95** (1960), 318-340.

[Go 1] G. Gong, *On inductive limits of matrix algebras over higher dimensional spaces*, Part II, Math. Scand., **80** (1997), 56-100.

[Go 2] ______, *On the classification of C^*-algebras of real rank zero and unsuspended E-equivalence types*, J. Funct. Anal., to appear.

[Go 3] ______, *On the classification of simple inductive limit C^*-algebras,* I: *The reduction theorem*, preprint.

[GL] G. Gong and H. Lin, *Exponential rank of inductive limit C^*-algebras*, Math. Scand., **71** (1992), 301-319.

[Goo] K.R. Goodearl, *Riesz decomposition in inductive limit C^*-algebras*, Rocky Mountain J. Math., **24**(**4**) (1994), 1405-1430.

[H] D. Husemoller, *Fibre Bundles*, 2nd ed., Springer Verlag, 1966.

[L] L. Li, *Classification of simple C^*-algebras: Inductive limits of matrix algebras over trees*, Memoirs of Amer. Math. Soc., **605** (1997).

[MP] M. Martin and C. Pasnicu, *Some comparability results in inductive limit C^*-algebras*, J. Operator Theory, **30** (1993), 137-147.

[N] V. Nistor, *Stable range for tensor products of extensions of $\mathcal{K}$ by $C(X)$*, J. Operator Theory, **16** (1986), 387-396.

[P 1] C. Pasnicu, *On inductive limits of certain C^*-algebras of the form $C(X) \otimes F$*, Trans. Amer. Math. Soc., **310** (1988), 703-714.

[P 2] ———, *Nonstable K-theory results for some AH-algebras*, Math. Scand., **78** (1996), 112-120.

[R] M.A. Rieffel, *Dimension and stable rank in the K-theory of C^*-algebras*, Proc. London Math. Soc., **46** (1983), 301-333.

[St] K.H. Stevens, *The classification of certain non-simple approximate interval algebras*, Ph.D. thesis, University of Toronto, 1994.

[Su] H. Su, *On the classification of real rank zero C^*-algebras: Inductive limits of matrix algebras over non-Hausdorff graphs*, Memoirs of Amer. Math. Soc., **114** (1995).

[V] J. Villadsen, *On the stable rank of simple C^*-algebras*, preprint.

Received April 10, 1998 and revised November 10, 1998. The results contained in this paper have been presented at the Great Plains Operator Theory Seminar (GPOTS), May 1995, Cincinnati, Ohio. The author was partially supported by NSF grant DMS-9401515.

University of Puerto Rico
San Juan, PR 00931-3355
E-mail address: cpasnic@upracd.upr.clu.edu

PACIFIC JOURNAL OF MATHEMATICS
Vol. 192, No. 1, 2000

HARMONIC FUNCTIONS ON MANIFOLDS WITH NONNEGATIVE RICCI CURVATURE AND LINEAR VOLUME GROWTH

Christina Sormani

In this paper we prove that if a complete noncompact manifold with nonnegative Ricci curvature and linear volume growth has a nonconstant harmonic function of at most polynomial growth, then the manifold splits isometrically.

Lower bounds on Ricci curvature limit the volumes of sets and the existence of harmonic functions on Riemannian manifolds. In 1975, Shing Tung Yau proved that a complete noncompact manifold with nonnegative Ricci curvature has no nonconstant harmonic functions of sublinear growth [**Yau2**]. That is, if

$$\limsup_{R\to\infty} \frac{\max_{B_p(R)} |f|}{R} = 0 \tag{1}$$

and if f is harmonic, then f is a constant. In the same paper, Yau used this result to prove that a complete noncompact manifold with nonnegative Ricci curvature has at least linear volume growth,

$$\liminf_{R\to\infty} \frac{\mathrm{Vol}\,(B_p(R))}{R} = C \in (0,\infty]. \tag{2}$$

There are many manifolds with nonnegative Ricci curvature that actually have linear volume growth

$$\limsup_{R\to\infty} \frac{\mathrm{Vol}\,(B_p(R))}{R} = V_0 < \infty. \tag{3}$$

Some interesting examples of such manifolds can be found in [**So1**].

In this paper, we prove the following theorem concerning harmonic functions on these manifolds.

Theorem 1. *Let M^n be a complete noncompact manifold with nonnegative Ricci curvature and at most linear volume growth,*

$$\limsup_{R\to\infty} \frac{\mathrm{Vol}\,(B_p(R))}{R} = V_0 < \infty. \tag{4}$$

If there exists a nonconstant harmonic function, f, of polynomial growth of any given degree q,

$$\Delta f = 0 \quad \textit{and} \quad |f(x)| \le C(d(x,p)^q + 1), \tag{5}$$

then the manifold splits isometrically, $M^n = N^{n-1} \times \mathbb{R}$.

Harmonic functions of polynomial growth have been an object of study for some time. Until recently it was not known whether the space of harmonic functions of polynomial growth of a given degree on a manifold with nonnegative Ricci curvature was finite dimensional. Atsushi Kasue proved this result with various additional assumptions in [**Kas1**, **Kas2**]. Tobias Colding and Bill Minnicozzi have recently proven that this space is indeed finite dimensional with no additional assumptions [**CoMin**]. With our stronger condition of linear volume growth, we are able to prove that this space is only one dimensional directly using a gradient estimate of Cheng and Yau [**ChgYau**] and results from [**So1**, **So2**].

For background material consult the textbooks [**SchYau**] and [**Li**].

The author would like to thank Professor Shing-Tung Yau for conjecturing this theorem and for his encouragement during her year at Harvard University. She would also thank to thank Professors Tobias Colding, Jeff Cheeger, William Minicozzi and Gang Tian for their continued interest in her work.

1. Background.

A *ray*, $\gamma : [0,\infty) \mapsto M^n$, is a geodesic which is minimal on any subsegment, $d(\gamma(t),\gamma(s)) = |t-s|$. Every complete noncompact Riemannian manifold contains a ray. Given a ray, one can define its associated *Busemann function*, $b : M^n \to \mathbb{R}$, as follows:

$$b(x) = \lim_{R\to\infty} \big(R - d(x,\gamma(R))\big). \tag{6}$$

The Busemann function is a Lipschitz function whose gradient has unit length almost everywhere, [**Bu**].

In Euclidean space, the level sets of the Busemann function associated with a given ray are the planes perpendicular to the given ray. In contrast, the Busemann function defined on a manifold with nonnegative Ricci curvature and linear volume growth has compact level sets with bounded diameter growth [**So1**, Thm. 15]. In that paper, the author also proved that if such a manifold is not an isometrically split manifold, then the Busemann function is bounded below and $b^{-1}((-\infty,r])$ is a compact set for all r [**So1**, Cor. 19].

Lemma 2. *Let M^n be a complete manifold with nonnegative Ricci curvature. Suppose that there is a Busemann function, b, which is bounded below* *by b_{min} and that diameter of the level sets grows at most linearly,*

$$\operatorname{diam}(b^{-1}(b_{\min}+r)) \leq C_D(r+1). \tag{7}$$

Then there exists a universal constant, C_n, depending only on the dimension, n, such that any harmonic function, f, satisfies the following gradient estimate:

$$\sup_{b^{-1}([b_{\min},b_{\min}+r))} |\nabla f| \;\leq\; \frac{C_n}{2(r+D)} \sup_{b^{-1}(b_{\min}+2(r+D))} |f| \tag{8}$$

for all $D \geq C_D(r+1)$.

Proof. First note that the boundary of the compact set, $b^{-1}([b_{\min}, r))$, is just $b^{-1}(r)$. So, by the maximum principal, we know that for any harmonic function, f,

$$\max_{b^{-1}([b_{\min},r))} f \leq \max_{b^{-1}(r)} f \quad \text{and} \quad \min_{b^{-1}([b_{\min},r))} f \geq \min_{b^{-1}(r)} f. \tag{9}$$

Furthermore, Cheng and Yau have proven the following gradient estimate for harmonic functions on balls in manifolds with nonnegative Ricci curvature,

$$\sup_{B_p(a/2)} |\nabla f| \leq \frac{C_n}{a} \sup_{B_p(a)} |f| \tag{10}$$

where C_n is a universal constant depending only on the dimension, n. [**ChgYau**, Thm. 6], see also [**ChgYau**, p. 21, Cor. 2.2]. This will be the constant in (8). Thus, we need only relate balls to regions defined by the Busemann function to prove the theorem.

Let x_0 be a point in $b^{-1}(b_{\min})$. Note that

$$B_{x_0}(a) \subset b^{-1}([b_{\min}, b_{\min}+a)) \tag{11}$$

because the triangle inequality implies that

$$\begin{aligned} b(x) &= \lim_{R\to\infty} R - d(x, \gamma(R)) \\ &\leq \lim_{R\to\infty} R - d(x_0, \gamma(R)) + d(x_0, x) \\ &= b(x_0) + d(x_0, x). \end{aligned}$$

On the other hand, using our diameter bound in (7), we claim that

$$b^{-1}([b_{\min}, b_{\min}+r)) \subset B_{x_0}(r+D) \qquad \forall D \geq C_D(r+1). \tag{12}$$

To see this we will construct a ray, σ, emanating from x_0 such that for all $t \geq 0$, $\sigma(t) \in b^{-1}(r_{\min}+t)$. Then, for any $y \in b^{-1}([b_{\min}, b_{\min}+r))$, we let $t = b(y)$ and we have

$$d(x_0, y) \leq d(x_0, \sigma(t)) + d(\sigma(t), y) \leq t + \operatorname{diam}(b^{-1}(r_{\min}+t)) \leq r + D$$

which implies (12). The ray, σ, is constructed by taking a limit of minimal geodesics, σ_i, from x_0 to $\gamma(R_i)$. A subsequence of such a sequence of minimal geodesics always converges. The limit ray satisfies the required property,

$$\begin{aligned} b(\sigma(t)) &= \lim_{i\to\infty} b(\sigma_i(t)) = \lim_{i\to\infty}\lim_{R\to\infty} R - d(\sigma_i(t),\gamma(R)) \\ &= \lim_{R\to\infty} R - (d(\sigma_i(0),\gamma(R)) - t) \;=\; b(x_0)+t. \end{aligned}$$

We can now combine the relationships between Busemann regions and balls, (12) and (11), with the gradient estimate, (10), and the maximum principal, (9), to prove the lemma. That is, for all $D \geq C_D(r+1)$, we have

$$\begin{aligned} \sup_{b^{-1}([b_{\min}, b_{\min}+r))} |\bigtriangledown f| &\leq \sup_{B_{x_0}(r+D)} |\bigtriangledown f| \text{ by (12)}, \\ &\leq \frac{C_n}{2(r+D)} \sup_{B_{x_0}(2(r+D))} |f| \\ &\leq \frac{C_n}{2(r+D)} \sup_{b^{-1}([b_{\min}, b_{\min}+2(r+D)))} |f| \\ &\leq \frac{C_n}{2(r+D)} \sup_{b^{-1}(b_{\min}+2(r+D))} |f|. \end{aligned}$$

□

We employ this lemma and elements of the proof to prove our theorem.

2. Proof of the Theorem.

The given manifold, M^n, has nonnegative Ricci curvature and linear volume growth. We will assume that M^n doesn't split isometrically and demonstrate that the harmonic functions of polynomial growth must be constant. Since the manifold doesn't split isometrically and has linear volume growth, any Busemann function, b, has a minimum value by [**So1**, Cor. 23]. Furthermore, by [**So2**, Thm. 1], the diameters of the level sets of the Busemann function grow sublinearly. Thus we satisfy the hypothesis of Lemma 2 with $C_D = 1$ in (7).

Let $M(r) = \max_{b^{-1}(b_{\min}+r)} |f|$, where f is a harmonic function of polynomial growth. Note that M is an nondecreasing function by the maximum principal, (9). By the lemma, we know that for all $r \geq b_{\min}$ and for all $D \geq (r+1)$, we can bound the gradient of f in terms of M,

$$\sup_{b^{-1}([b_{\min}, b_{\min}+r))} |\bigtriangledown f| \;\leq\; \frac{C_n M(2(r+D))}{2(r+D)}. \tag{13}$$

Since $b^{-1}(r)$ is compact, there exists $x_r, y_r \in b^{-1}(b_{\min}+r)$ such that

$$f(x_r) = \min_{b^{-1}(b_{\min}+r)} f \qquad \text{and} \qquad f(y_r) = \max_{b^{-1}(b_{\min}+r)} f. \tag{14}$$

We claim that, for r sufficiently large, $M(r) \leq f(y_r) - f(x_r)$.

First recall that if f is a positive or negative harmonic function on a manifold with nonnegative Ricci curvature, then f must be constant [**Yau1**, Cor. 1]. So there exists a point $z \in M^n$ such that $f(z) = 0$. Thus, by the maximum principal, if $r \geq b(z)$ we know that $f(y_r) \geq 0$ and $f(x_r) \leq 0$. So $M(r) = \max(f(y_r), -f(x_r)) \leq f(y_r) - f(x_r)$.

We can now estimate $M(r)$ from above in terms of the gradient of f and the diameter of the level set, $b^{-1}(r)$. First we join x_r to y_r by a smooth minimal geodesic, γ_r. Note that the length of γ_r, is less than or equal to $\operatorname{diam}(b^{-1}(r))$ by the definition of diameter. So $\gamma_r \subset b^{-1}(b_{\min}, r + \operatorname{diam}(b^{-1}(r)))$. Thus for all $r \geq b(z)$, for all $D \geq (r+1)$, we have

$$
\begin{aligned}
M(r) &\leq f(y_r) - f(x_r) \\
&\leq \int_0^{L(\gamma_r)} \frac{d}{dt} f(\gamma(t))\, dt \\
&\leq \int_0^{L(\gamma_r)} |\bigtriangledown f|\, |\gamma'(t)|\, dt \\
&\leq \sup_{b^{-1}([b_{min}, r + \operatorname{diam}(b^{-1}(r))))} |\bigtriangledown f| \int_0^{L(\gamma_r)} |\gamma'(t)|\, dt \\
&\leq \frac{C_n M(2(r + \operatorname{diam}(b^{-1}(r)) + D))}{2(r + \operatorname{diam}(b^{-1}(r)) + D)} \operatorname{diam}(b^{-1}(r)) \\
&\leq C_n M(2(r + \operatorname{diam}(b^{-1}(r)) + D)) \, \frac{\operatorname{diam}(b^{-1}(r))}{2r} \\
&\leq C_n M(2(r + (r+1) + D)) \, \frac{\operatorname{diam}(b^{-1}(r))}{2r}.
\end{aligned}
$$

Setting $D = r + 1$ and taking $r \geq 1$, we have

$$M(r) \leq C_n\, M(6r)\, \frac{\operatorname{diam}(b^{-1}(r))}{2r}. \tag{15}$$

Recall that our manifold has sublinear diameter growth by [**So2**, Thm. 1]. So, given any $\delta > 0$, we can find $R_\delta \geq 1$ such that

$$\frac{\operatorname{diam}(b^{-1}(r))}{2r} < \delta \qquad \forall r \geq R_\delta. \tag{16}$$

Then, for all $k \in \mathbf{N}$ and for all $R \geq R_\delta$, we have

$$M(R) \leq C_n\, M(6R)\, \delta \leq \cdots \leq C_n^k\, M(6^k R)\, \delta^k. \tag{17}$$

Now f has polynomial growth of order q, (5), so

$$M(r) = \max_{x \in b^{-1}(b_{\min} + r)} |f(x)| \leq \max_{x \in b^{-1}(b_{\min} + r)} C(d(x, x_0)^q + 1). \tag{18}$$

Applying (12) with $C_D = 1$ and $D = C_D(r+1)$, we get

$$M(r) \le C((r+(r+1))^q+1) \le C(6r^2)^q \qquad \forall\, r \ge 1. \tag{19}$$

Substituting this information into (17), we get

$$\begin{aligned} M(R) &\le C_n^k\, C\left(6(6^k R)^2\right)^q \delta^k \\ &\le C\, 6^q\, R^{2q}\, (C_n 6^{2q}\delta)^k \qquad \forall R \ge R_\delta. \end{aligned}$$

Fix $\delta = 1/(2C_n 6^{2q})$, so R_δ is fixed by (16). Then, for all $R \ge R_\delta$, we have

$$M(R) \le \lim_{k\to\infty} C\, 6^q\, R^{2q}\, (1/2)^k = 0. \tag{20}$$

Since $M(r)$ is nondecreasing and nonnegative, $M(r) = 0$ everywhere. Thus, f is a constant. □

References

[Bu] H. Busemann, *The Geometry of Geodesics*, Academic Press Inc., New York, N. Y., 1955.

[ChGl] J. Cheeger and D. Gromoll, *The splitting theorem for manifolds of nonnegative Ricci curvature*, J. Differential Geometry, **6** (1971/72), 119-128.

[ChgYau] S.Y. Cheng and S.T. Yau, *Differential equations on Riemannian manifolds and their geometric applications*, Communications on Pure and Applied Mathematics, Vol. XXVII, (1975), 333-354.

[CoMin] T. Colding and W. Minicozzi, *Harmonic functions on manifolds*, Ann. of Math., **146**(**3**) (1997), 725-747.

[Kas1] A. Kasue, *Harmonic functions of polynomial growth on complete Riemannian manifolds* I, Proc. Symp. Pure Math., Vol. 54, part II, (1993), 281-290.

[Kas2] ______, *Harmonic functions of polynomial growth on complete Riemannian manifolds* II, J. Math. Soc. Japan, Vol. 47, (1995), 37-65.

[Li] P. Li, *Lecture Notes on Geometric Analysis*, Lecture Notes Series, **6**, Seoul National University, Research Institute of Mathematics, Global Analysis Research Center, Seoul, 1993.

[SchYau] R. Schoen and S.-T. Yau, *Lectures on Differential Geometry*, Conference Proceedings and Lecture Notes in Geometry and Topology, Vol. 1, International Press Inc., Cambridge, MA, 1994.

[So1] C. Sormani, *Busemann functions on manifolds with lower Ricci curvature bounds and minimal volume growth*, Journal of Differential Geometry, **48** (1998), 557-585.

[So2] ______, *The rigidity and almost rigidity of manifolds with minimal volume growth*, Communications in Analysis and Geometry, to appear.

[So3] ______, *Complete Noncompact Manifolds with Lower Bounds on Ricci Curvature and Minimal Volume Growth*, PhD Thesis, CIMS, 1996.

[Yau1] S.T. Yau. *Harmonic functions on complete Riemannian manifolds*, Comm. Pure Appl. Math., **28** (1975), 201-228.

[Yau2] ———, *Some function theoretic properties of complete Riemannian manifold and their applications to geometry*, Indiana University Mathematics Journal, **25**(**7**) (1976), 659-670.

Received April 3, 1998.

Johns Hopkins University
Baltimore, MD 21218
E-mail address: sormanic@member.ams.org

PACIFIC JOURNAL OF MATHEMATICS
Vol. 192, No. 1, 2000

ON UNIVALENT HARMONIC MAPPINGS AND MINIMAL SURFACES

Allen Weitsman

If S is the graph of a minimal surface, then when given parametrically by the Weierstrass representation, the first two coordinate functions give a univalent harmonic mapping. In this paper, the starting point is a univalent harmonic mapping f of the unit disk U. A height function is defined on an appropriate Riemann surface over the range of f which satisfies the minimal surface equation away from the branch points. This height function is then used to obtain function theoretic information about f.

1. Introduction.

Let f be a univalent harmonic mapping of the unit disk U. By this it is meant not only that f is $1-1$ and harmonic, but also that f is sense preserving.

Harmonic univalent mappings were first studied in connection with minimal surfaces by E. Heinz [**H**]. However, considerable interest in their function theoretic properties, quite apart from this connection, was generated by Clunie and Sheil-Small [**CS-S**].

Now, the Jacobian of $f(\zeta)$ is $J = |f_\zeta|^2 - |f_{\overline{\zeta}}|^2$, and f can be written

$$f = h + \overline{g} \tag{1.1}$$

where h and g are analytic in U. If $a(\zeta)$ is defined by

$$a(\zeta) = \overline{f_{\overline{\zeta}}(\zeta)}/f_\zeta(\zeta) = g'(\zeta)/h'(\zeta), \tag{1.2}$$

then $a(\zeta)$ is analytic and $|a(\zeta)| < 1$ in U. We shall refer to $a(\zeta)$ as the *analytic dilatation* as opposed to the usual dilatation $f_{\overline{\zeta}}/f_\zeta$ in the theory of quasiconformal mappings.

The case where $a(\zeta)$ is a finite Blaschke product is of special interest since this case arises in taking Fourier series of step functions [**S-S**]. Their function theoretic properties have been studied in [**HS2**] as well as in [**S-S**], and infinite Blaschke products have been considered in [**L**].

In the present paper we shall study a connection between harmonic mappings and the theory of minimal surfaces, and in §4 we use this to prove a special case of uniqueness for the Riemann mapping theorem of Hengartner and Schober [**HS1**]. As we have shown elsewhere, uniqueness fails in general [**W**].

2. Definition of the height function and conjugate height function.

Using the Weierstrass representation [**O**, p. 63] we shall associate with f, a minimal surface given parametrically in a simply connected subdomain $N \subseteq U$ where $a(\zeta)$ does not have a zero of odd order.

With g and h as in (1.1) we define up to an additive constant, a branch of

(2.1)
$$F(\zeta) = 2i \int \sqrt{h'(\zeta)g'(\zeta)}\; d\zeta = 2i \int h'(\zeta)\sqrt{a(\zeta)}\; d\zeta = 2i \int f_\zeta(\zeta)\sqrt{a(\zeta)}\; d\zeta.$$

Then, by (1.2) it follows that a branch of F can be defined in N, and for $\zeta \in N$,

$$\zeta \to (f(\zeta),\ \operatorname{Re} F(\zeta)) \tag{2.2}$$

gives a parametric representation of a minimal surface. Here we have identified $\mathbb{R}^2$ with $\mathbb{C}$ by $(x, y) \leftrightarrow (\operatorname{Re} f, \operatorname{Im} f)$.

Let $\hat{U}$ be the Riemann surface of the function $\sqrt{a(\zeta)}$. Then $\hat{U}$ has algebraic branch points corresponding to those points $\zeta \in U$ for which $a(\zeta)$ has a zero of odd order. Specifically, $\hat{U}$ can be concretely described (the *analytic configuration* [**Sp**, 69-74]) in terms of function elements (α, F_α) where $\alpha \in U$, and F_α is a power series expansion of a branch of F in a neighborhood of α if $a(\zeta)$ does not have a zero of odd order at $\zeta = \alpha$, and F_α a power series in $\sqrt{\zeta - \alpha}$ otherwise. The mapping $p\colon (\alpha, F_\alpha) \to \alpha$ is the *projection* of the surface so realized. The mapping F may now be lifted to a mapping $\hat{F}$ on $\hat{U}$.

By continuation, we may induce a mapping $\hat{U} \to \tilde{U}$ to a surface $\tilde{U}$ with a real analytic structure defined in terms of elements $(\beta, \tilde{F}_\beta)$ with $\beta \in f(U)$ by $\alpha = f^{-1}(\beta)$ and $\tilde{F}_\beta = F_\alpha \circ f^{-1}$. We again define a projection by $\pi\colon (\beta, \tilde{F}_\beta) \to \beta$.

We shall refer to a point $\hat{\zeta} \in \hat{U}$ to be over ζ, if $p(\hat{\zeta}) = \zeta$, and $\tilde{z} \in \tilde{U}$ to be over z if $\pi(\tilde{z}) = z$.

The harmonic mapping $f\colon U \to f(U)$ lifts to a mapping $\hat{f}\colon \hat{U} \to \tilde{U}$ which is $1-1$, onto, and satisfies the condition $\pi(\hat{f}(\hat{\zeta})) = f(p(\hat{\zeta}))$ for all $\zeta \in \hat{U}$. With these notations, we shall extend the meaning of (2.2). Thus

$$\hat{\zeta} \to (\hat{f}(\hat{\zeta}), \operatorname{Re} \hat{F}(\hat{\zeta})) \tag{2.3}$$

gives a parametric representation of a minimal surface in the sense that in a neighborhood of $\hat{\zeta} \in \hat{U} \backslash \mathcal{B}$ where $\mathcal{B}$ is the branch set, that is, the points above the zeros of a of odd order, then (2.2) is the same as (2.3) computed in terms of local coordinates given by projection.

We may also define the surface nonparametrically on $\tilde{U}\backslash\tilde{\mathcal{B}}$, where $\tilde{\mathcal{B}} = \hat{f}(\mathcal{B})$, as follows. Let D be an open disk in $f(U)$ such that $f^{-1}(D)$ contains no zeros of a of odd multiplicity. Let $w = \varphi(x, y)$ be the nonparametric description of the minimal surface corresponding to (2.2), that is, for $\zeta \in f^{-1}(0)$ (cf. [**HS3**, p. 87]),

$$x = \operatorname{Re} f(\zeta) \qquad y = \operatorname{Im} f(\zeta), \tag{2.4}$$
$$\varphi(x, y) = \operatorname{Re} F(\zeta).$$

Then, by continuation φ lifts to a function $\tilde{\varphi}$ on $\tilde{U}$ which satisfies the minimal surface equation when computed in local coordinates given by projection off the branch set $\tilde{\mathcal{B}}$. We shall call $\tilde{\varphi}(\tilde{z})$ a *height function* corresponding to f. Finally, we define a *conjugate height function* $\tilde{\psi}(z)$ by solving locally

$$\psi_y = \varphi_x/W, \ \psi_x = -\varphi_y/W \qquad \left(W = \sqrt{1 + \varphi_x^2 + \varphi_y^2}\right) \tag{2.5}$$

(cf. [**F1**, p. 344]) and lifting to $\tilde{U}\backslash\tilde{\mathcal{B}}$ as was done for φ. Let $\tilde{F} = \tilde{\varphi} + i\tilde{\psi}$. Then $\tilde{F}$ is real analytic and locally quasiconformal on $\tilde{U}\backslash\tilde{\mathcal{B}}$, with dilatation whose magnitude is $(W-1)/(W+1)$. The fact that $\tilde{\psi}$ and $\tilde{F}$ are well defined on $\tilde{U}\backslash\tilde{B}$ follows from Theorem 1.

A glossary of terminology is given schematically in Figure 1.

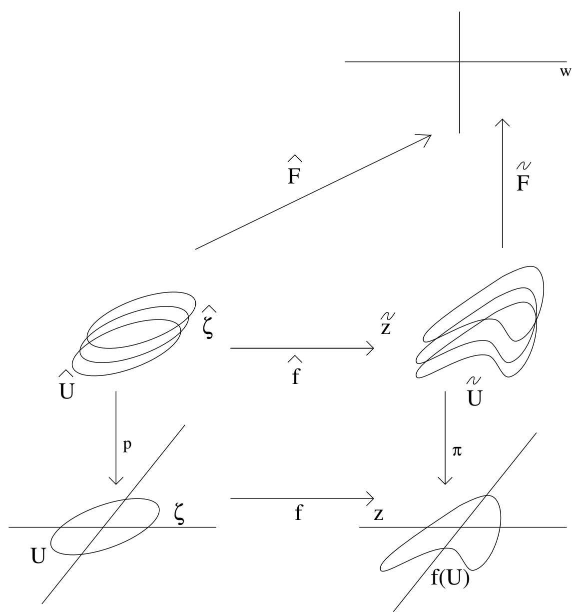

Figure 1.

Theorem 1. *With the above notations, $\hat{F} = \tilde{F} \circ \hat{f} + C$ for some constant C.*

Proof. Let D be an open disk in $f(U)$ such that $f^{-1}(D)$ contain no zeros of odd multiplicities of a. We fix a branch of $\sqrt{a}$ in $f^{-1}(D)$, and consider $\hat{\varphi}(\hat{\zeta}) + i\hat{\psi}(\hat{\zeta}) = \hat{F}(\hat{\zeta})$ for points in a component of $\hat{U}$ over $f^{-1}(D)$, and $\tilde{\varphi}(\tilde{z}) + i\tilde{\psi}(\tilde{z}) = \tilde{F}(\tilde{z})$ for points in a component of $\tilde{U}$ over D. Since we shall compute in local coordinates given by projection, to reduce notation in this proof, we shall subsequently write $\hat{F}$, $\hat{\varphi}$, $\hat{\psi}$ in place of $\hat{F}\circ p^{-1}$, $\hat{\varphi}\circ p^{-1}$, $\hat{\psi}\circ p^{-1}$, and $\tilde{F}$, $\tilde{\varphi}$, $\tilde{\psi}$ in place of $\tilde{F} \circ \pi^{-1}$, $\tilde{\varphi} \circ \pi^{-1}$, $\tilde{\psi} \circ \pi^{-1}$ respectively. With this notation, by (2.4) we have that

$$\hat{\varphi} = \tilde{\varphi} \circ f, \tag{2.6}$$

so it suffices to show that

$$\hat{\psi} = \tilde{\psi} \circ f + C. \tag{2.7}$$

The result then follows from continuation. □

In fact, since $\hat{\varphi} + i\hat{\psi}$ is analytic in $f^{-1}(D)$, it follows from (2.6) that to prove (2.7) it suffices to show that $\tilde{F} \circ f$ is analytic in $f^{-1}(D)$.

We first record the relationship between $a(\zeta)$ of (1.2) and $W(z)$ $(z = f(\zeta))$ of (2.5). This is given by [**O**, p. 105], [**HS3**, pp. 87-88] as

$$|a| = \frac{W-1}{W+1}. \tag{2.8}$$

Now,

$$(\tilde{F} \circ f)_{\overline{\zeta}} = \tilde{F}_z f_{\overline{\zeta}} + \tilde{F}_{\overline{z}} \overline{f}_{\overline{\zeta}} = \tilde{F}_z f_{\overline{\zeta}} + \tilde{F}_{\overline{z}} \overline{(f_\zeta)}. \tag{2.9}$$

A simple computation using (2.5) gives

$$F_z = \frac{W+1}{W}\varphi_z, \qquad F_{\overline{z}} = \frac{W-1}{W}\varphi_{\overline{z}}.$$

When used in (2.9) these give

$$(\tilde{F} \circ f)_{\overline{\zeta}} = \frac{W+1}{W}\tilde{\varphi}_z f_{\overline{\zeta}} + \frac{W-1}{W}\tilde{\varphi}_{\overline{z}}\overline{(f_\zeta)}. \tag{2.10}$$

Again, a direct computation gives

$$\tilde{\varphi}_z = \frac{\hat{\varphi}_\zeta \overline{(f_\zeta)} - \hat{\varphi}_{\overline{\zeta}}\overline{(f_{\overline{\zeta}})}}{|f_\zeta|^2 - |f_{\overline{\zeta}}|^2}\ , \quad \tilde{\varphi}_{\overline{z}} = \frac{\hat{\varphi}_{\overline{\zeta}} f_\zeta - \hat{\varphi}_\zeta f_{\overline{\zeta}}}{|f_\zeta|^2 - |f_{\overline{\zeta}}|^2}.$$

When used in (2.10) this gives

$$(\tilde{F}\circ f)_{\overline{\zeta}} = \frac{1}{W(|f_\zeta|^2 - |f_{\overline{\zeta}}|^2)}\left(2\hat{\varphi}_\zeta f_{\overline{\zeta}}\overline{(f_\zeta)} + \hat{\varphi}_{\overline{\zeta}}|f_\zeta|^2\left(W - 1 - \frac{|f_{\overline{\zeta}}|^2}{|f_\zeta|^2}(W+1)\right)\right). \tag{2.11}$$

Now, by (1.2), (2.1), and (2.8) we have,

$$\hat{\varphi}_\zeta = ig'/\sqrt{a},\ \hat{\varphi}_{\overline{\zeta}} = -i\overline{g'}/\overline{\sqrt{a}},\ f_\zeta = g'/a,\ f_{\overline{\zeta}} = \overline{g'},$$

and

$$W - 1 - \frac{|f_{\overline{\zeta}}|^2}{|f_\zeta|^2}(W+1) = W - 1 - |a|^2(W+1) = 2(W-1)/(W+1).$$

Substituting into (2.11) we obtain

$$\begin{aligned}(\tilde{F} \circ f)_{\overline{\zeta}} &= \frac{1}{W(|f_\zeta|^2 - |f_{\overline{\zeta}}|^2)}\left(\frac{2ig'(\overline{g'})^2}{\sqrt{a}\overline{a}} - \frac{2i\overline{g'}|g'|^2}{\overline{\sqrt{a}}|a|^2}\left(\frac{W-1}{W+1}\right)\right) \\ &= 0.\end{aligned}$$

Thus, $\tilde{F} \circ f$ is analytic and (2.7) follows.

3. The height function corresponding to Poisson integrals of step functions.

Let $\mathcal{P}$ be a polygon with vertices $c_1, \dots, c_n$ given cyclically, and in order induced by a positive orientation of $\partial\mathcal{P}$. Let f be the Poisson integral of a step function on ∂U having values $c_1, \dots, c_n$ and suppose that f is then a univalent harmonic mapping, $f\colon U \to \mathcal{P}$. If $\mathcal{P}$ is convex, for example, this will always be the case [**C**], [**K**]. The analytic dilatation $a(\zeta)$ for such mappings were studied in [**HS2**] and [**S-S**]. In general, $a(\zeta)$ is a Blaschke product of order at most $n-2$, and of order precisely $n-2$ if $\mathcal{P}$ is convex [**S-S**, pp. 469, 473].

We shall now explore the boundary behavior of height functions corresponding to such mappings. The prototype for this is Scherk's minimal surface over the square $-\pi/2 < x < \pi/2,\ -\pi/2 < y < \pi/2$, given by

$$\psi(x, y) = \log(\cos x/\cos y) \tag{3.1}$$

which tends to $+\infty$ and $-\infty$ over alternate sides. It seems remarkable that this type of behavior persists in general for height functions corresponding to all such f described above.

Theorem 2. *Let $\mathcal{P}$ be a polygon having vertices $c_1, \dots, c_n$ given cyclically, and ordered by a positive orientation on $\partial\mathcal{P}$. Let f be a univalent harmonic mapping of U such that f is the Poisson integral of a step function having the ordered sequence $c_1, \dots, c_n$ as its values. Then the analytic dilatation $a(\zeta)$ of f is a finite Blaschke product of order at most $n-2$, $f(U) = \mathcal{P}$, and if φ is a height function for f, then φ tends to $+\infty$ or $-\infty$ at points over the open segments making up the sides of $\mathcal{P}$. If $\mathcal{P}$ is convex, then $+\infty$ and $-\infty$ alternate on adjacent sides.*

Proof. That $a(\zeta)$ is a Blaschke product of order at most $n-2$ and $f(U) = \mathcal{P}$ follow from general properties of Poisson integrals [**S-S**, p. 469], [**HS2**, p. 203]. $\square$

Let $f = h + \overline{g}$ as in (1.1). Then we may write h' and g' in the form [**S-S**, pp. 460-461]

$$h'(\zeta) = \sum_{k=1}^{n} \frac{\alpha_k}{\zeta - \zeta_k}, \ g'(\zeta) = -\sum_{k=1}^{n} \frac{\overline{\alpha_k}}{\zeta - \zeta_k},$$

where $\alpha_k \neq 0, \quad k = 1, \dots, n$.

With F as in (2.1), we are then interested in the branches of

$$F(\zeta) = -2 \int \sqrt{\sum_{k=1}^{n} \frac{\alpha_k}{\zeta - \zeta_k} \sum_{k=1}^{n} \frac{-\overline{\alpha_k}}{\zeta - \zeta_k}} \, d\zeta \tag{3.2}$$

as $\zeta \to \zeta_k, \quad k = 1, \dots, n$. The cluster sets for the nontangential approaches to points over the ζ_k give the points lying over the open segments making up the sides of $\mathcal{P}$.

Thus, take a vertex ζ_j, and an open segment l_j of $\partial\mathcal{P}$ corresponding to it. Then, as $\zeta \to \zeta_j$,

$$\sum_{k=1}^{n} \frac{\alpha_k}{\zeta - \zeta_k} \sum \frac{-\overline{\alpha_k}}{\zeta - \zeta_k} = \frac{|\alpha_j|^2}{(\zeta - \zeta_j)^2}(1 + o(1)),$$

and hence, by (3.2), a branch of F satisfies

$$F(\zeta) = \pm 2|\alpha_j| \log(\zeta - \zeta_j) + o(1) \tag{3.3}$$

as $\zeta \to \zeta_j$, for a fixed branch of the log. Suppose the fixed branch of (3.3) has minus sign, and let $\phi(z) = \operatorname{Re} F \circ f^{-1}(z)$ be a corresponding branch in $\mathcal{P}$ for points near the corresponding side l_j. Now suppose $\mathcal{P}$ is convex and $F(\zeta)$ is analytically continued to an adjacent point, say ζ_{j+1}, so that ϕ is then continued to a corresponding side l_{j+1} having common endpoint c_j with l_j. Since $\phi \to -\infty$ as $z \to l_j$, it remains to show that $\phi \to +\infty$ as $z \to l_{j+1}$. This effect has been noted for minimal surfaces [**JS**], and can be accomplished by a simple barrier argument. I thank Professor Finn for pointing this out.

Let $0 < \beta < \pi$ be the angle in $\mathcal{P}$ between l_j and l_{j+1}. Suppose that $\phi \to -\infty$ on both open segments l_j and l_{j+1}. Since ϕ satisfies the minimal surface equation, ϕ can only tend to $-\infty$ over line segments [**O**, p. 102]. Since we make no assumption at the common endpoint c_j, in order to get a contradiction we must show that $\phi \to -\infty$ at c_j as well. We may assume that $c_j = (\pi/2, 0)$, and l_j, l_{j+1} make the angle β symmetrically with respect to the x axis, opening toward the origin. Let $0 < \varepsilon < (\pi/2)\cot(\beta/2)$ be small enough so that the isosceles triangle N formed by the sector and the line $x = \pi/2 - \varepsilon$ has the given branch of F single valued. Then, two of the sides of N are contained in the segments l_j and l_{j+1}, and the third is $x = \pi/2 - \varepsilon$, $-\delta < y < \delta$, where $\delta = \varepsilon \tan(\beta/2)$. If ψ is the height function for Scherk's surface given by (3.1), then for any $M > 0$, clearly

$$\phi(x, y) < -\psi(x - \pi + \varepsilon, y) - M \tag{3.4}$$

on $\partial N \backslash \{c_j\}$. By the extended maximum principle [**F1**, pp. 342-343], it follows that (3.4) holds thoughout N. Since $M > 0$ was arbitrary, it follows that $\phi \equiv -\infty$ on N, a contradiction. Thus $\phi = +\infty$ on l_{j+1}.

4. An application to the Riemann mapping theorem.

One of the most basic results in the theory of univalent harmonic mappings is the Riemann mapping theorem of Hengartner and Schober [**HS1**].

Theorem A. *Let D be a bounded simply connected domain whose boundary is locally connected. Fix $w_0 \in D$, and let $a(\zeta)$ be analytic in U, with $a(U) \subseteq U$. Then there exists a univalent harmonic mapping f with the following properties.*

a) *f maps U into D and $f(0) = w_0$, $f_z(0) > 0$.*
b) *f satisfies the equation $\overline{(f_{\overline{\zeta}})} = a f_\zeta$.*
c) *Except for a countable set $E \subseteq \partial U$, the unrestricted limit $f^*(e^{it}) = \lim_{\zeta \to e^{it}} f(\zeta)$ exists and belongs to ∂D.*
d) *The one sided limits $\lim_{\tau \to t^+} f*(e^{i\tau})$, $\lim_{\tau \to t^-} f^*(e^{i\tau})$ through values of $e^{i\tau} \notin E$ exist and belong to ∂D; for $e^{it} \notin E$ they are equal and for $e^{it} \in E$ they are different.*
e) *The cluster set of f at $e^{it} \in E$ is the straight line segment joining the left and right limits in* d).

If in Theorem A, the set D is convex, and $a(\zeta)$ is a finite Blaschke product, one can say more [**HS2**, p. 203], [**S-S**, p. 473].

Theorem B. *Let f be as in Theorem A with D bounded and convex, and $a(\zeta)$ a Blaschke product of order $n - 2$. Then $f(U)$ is a polygon with n vertices all of which lie on ∂D.*

We shall prove uniqueness in the case $a(\zeta) = \zeta^n$ and D convex. The case of uniqueness when $D = U$ and $a(\zeta) = \zeta$ was done in [**HS2**, p. 204].

The proof involves a combinatorial argument with the level sets of the height function. Such arguments are often useful in the theory of partial differential equation, and in particular the minimal surface equation [**F1**], [**FO**], [**JS**], [**Se**].

Theorem 3. *The solution $f(\zeta)$ to the Riemann mapping theorem above with D convex and*

$$a(\zeta) = \zeta^{n-2} \tag{4.1}$$

is unique for each $n = 3, 4, \ldots$

Proof. Let f_1 and f_2 be Riemann mappings corresponding to D. We may assume $f_1(0) = f_2(0) = 0$. Let Δ be a disk centered at 0, and contained in $f_1(U) \cap f_2(U)$.

If n is even, then $\hat{U} = U$ and if n is odd $\hat{U}$ is a two sheeted cover of U with branch point over 0. Similarly, if $\tilde{U}_1$ corresponds to $f_1(U)$ and $\tilde{U}_2$ to $f(U_2)$, then $\tilde{U}_1$ and $\tilde{U}_2$ are one or two sheeted according as n is even or odd. We consider the case where n is odd. The even case goes the same way, but is simpler since one can bypass discussion of Riemann surfaces.

Let $\varphi_j, \psi_j, \tilde{\varphi}_j, \tilde{\psi}_j, \tilde{F}_j, \tilde{U}_j, \pi_j, \quad j = 1, 2$ be the quantities of §2 defined for f_1 and f_2 respectively. We may assume that $\tilde{F}_1(\tilde{0}) = \tilde{F}_2(\tilde{0}) = 0$. If $\tilde{\Delta}$ represents the Riemann surface of $\sqrt{z}$ over Δ, then we may consider $\tilde{\Delta} \subseteq \tilde{U}_1$ and $\tilde{\Delta} \subseteq \tilde{U}_2$, so that $\tilde{F}_1$ and $\tilde{F}_2$ may both be considered as defined for all $\tilde{z} \in \tilde{\Delta}$. For brevity of notation, we shall write $\tilde{F}$ for $\tilde{F} \circ \pi^{-1}$.

Since the analytic dilatation for $f_1(\zeta)$ and $f_2(\zeta)$ is 0 when $\zeta = 0$, it follows from (1.2), (4.1), and a) of Theorem A, that

$$f_j(\zeta) = c_j \zeta(1 + o(1)) \quad (\zeta \to 0,\ c_j > 0,\ j = 1, 2). \tag{4.2}$$

Then, from (2.1), (4.1), (4.2), and Theorem 1 we may take determinations of $\tilde{F}_1$ and $\tilde{F}_2$ in $\tilde{\Delta}$ so that

$$\tilde{\varphi}_j(z) + i\tilde{\psi}_j(z) = \tilde{F}_j(z) = d_j z^{n/2}(1 + o(1)) \quad (j = 1, 2 \quad z \to 0) \tag{4.3}$$

with $d_1, d_2 > 0$ and $z^{n/2}$ is some fixed branch.

Having thus fixed branches in (4.3) we may then take a constant $\lambda > 0$ such that

$$\tilde{F}_1(z) - \lambda \tilde{F}_2(z/\lambda) = C z^{\frac{p+2}{2}}(1 + o(1)) \quad (z \to 0) \tag{4.4}$$

for some constant C and integer $p \geq n$. We suppose $\lambda \geq 1$; otherwise we interchange $\tilde{F}_1$ and $\tilde{F}_2$. Now, the change from $F(z)$ to $\lambda F(z/\lambda)$ corresponds to replacing f by λf. Then the analytic dilatation is unchanged, and following the change in (2.1) it gives the parametrization $\zeta \to (\lambda f(\zeta), \operatorname{Re} \lambda F(\zeta))$.

Let $\varphi_3, \psi_3, \tilde{\varphi}_3, \tilde{\psi}_3$ correspond to $f_3 = \lambda f_2$ so that $f_3(U)$, is nothing more than $f_1(U)$ dilated by the constant $\lambda \geq 1$, and (4.5) becomes

$$\tilde{F}_1(\tilde{z}) - \tilde{F}_3(\tilde{z}) = C z^{\frac{p+2}{2}}(1 + o(1)) \quad (z \to 0). \tag{4.5}$$

Case 1. $C = 0$ for every p. Since $\tilde{F}_1(z^2) - \tilde{F}_3(z^2)$ is real analytic, then $\tilde{F}_1 \equiv \tilde{F}_3$. Thus, in particular $\lambda = 1$ and $f_1(U) = f_3(U) = \mathcal{P}$. In order to show that $f_1 \equiv f_3$ we use the subordination principle of [**BHH**, p. 170]. Briefly, since $\mathcal{P}$ is a convex polygon by Theorem B, and $(f_1)_z(0),\ (f_3)_z(0) > 0$, we may apply the argument principle in [**BHH**, p. 170] to

$$G(z) = (f_3)_z(0) f_1(z) - (f_1)_z(0) f_3(z)$$

to deduce that $(f_1)_z(0) = (f_3)_z(0)$. Then, another application of the argument principle as in [**BHH**] to $G_\varepsilon(z) = (1+\varepsilon)f_1(z) - f_3(z)$ $(\varepsilon \to 0)$ shows that $f_1 \equiv f_3$.

Case 2. $C \neq 0$ for some $p \geq n$. In this case, near the origin on $\tilde{\Delta}$, by (4.5) there are $2p+4$ level curves $\tilde{\varphi}_1 - \tilde{\varphi}_3 = 0$ emanating from $\tilde{0}$. Between the level curves, $\tilde{\varphi}_1 - \tilde{\varphi}_3$ alternates in sign. In order to analyze the component sets between the level sets, we must modify f_3.

Let $\eta_1, \eta_2, \dots$ be homeomorphisms of $|\zeta| = 1$ onto the boundary of λD, which converge to the (step function) boundary values of f_3, and let $f_3^{(n)}$, $n = 1, 2, \dots$ their corresponding Poisson integrals so that $f_3^{(n)} \to f_3$ uniformly on compact subsets of U.

The level sets of $\tilde{\varphi}_1 - \tilde{\varphi}_3 = 0$ create $2p+4$ disjoint component open sets $O_1,\ O_2, \dots, O_{2p+4}$ where $\tilde{\varphi}_1 - \tilde{\varphi}_3 > 0$ in O_{2j-1} and $\tilde{\varphi}_1 - \tilde{\varphi}_3 < 0$ in O_{2j} for $j = 1, \dots, p+2$. These components alternate in position around the origin.

For $\varepsilon > 0$ we can find nonempty components at $O_1(\varepsilon),\ O_2(\varepsilon), \dots, O_{2p+4}(\varepsilon)$ where $\tilde{\varphi}_1 - \tilde{\varphi}_3^{(n)} > \varepsilon$ in $O_{2j-1}(\varepsilon)$, $\tilde{\varphi} - \tilde{\varphi}_3^{(n)} = \varepsilon$ on $\partial O_{2j-1}(\varepsilon)$, $\tilde{\Delta} \cap O_{2j-1}(\varepsilon) \subseteq O_{2j-1}$, $j = 1, \dots, 2p$, and analogous statements hold for $O_{2j}(\varepsilon)$, $j = 1, \dots, p+2$.

Now, $f_3^{(j)}(U) = \lambda D$, so by the maximum principle for solutions to the minimal surface equation, the level sets forming the boundaries of the $O_j(\varepsilon)$'s must extend to points over the boundary of $\mathcal{P} = f_1(U)$. As in [**FO**, pp. 357-358], we observe that since $\tilde{F}_1$ is $\pm\infty$ over the sides of $\mathcal{P}$ by Theorem 2, if a component $O_j(\varepsilon)$ has a boundary point over an interior point of a side of $\mathcal{P}$, then the boundary must contain that side. Since, by Theorem B, $\mathcal{P}$ has n sides, then $\tilde{\mathcal{P}} = \pi_1^{-1}(\mathcal{P})$ has $2n$ sides. This implies that there are at most $2n$ sets $O_j(\varepsilon)$ whose boundaries have interior points over $\partial\mathcal{P}$. If $O_j(\varepsilon)$ were a component whose boundary contained no points over $\partial\mathcal{P}$, then its boundary could only be interior points over $\mathcal{P}$, or vertices. As pointed out in [**FO**, p. 358], this is impossible by a theorem of Finn [**F1**, pp. 342-343]. Thus, $2p+4 \leq 2n$. Since $p \geq n$, we obtain a contradiction and the theorem is proved. $\square$

References

[BHH] D. Bshouty, N. Hengartner and W. Hengartner, *A constructive method for starlike harmonic mappings*, Numer. Math., **54** (1988), 167-178.

[C] G. Choquet, *Sur un type de transformation analytique généralisant la représentation conforme et définie an moyen de fonctions harmoniques*, Bull. Sci. Math., **69** (1945), 156-165.

[CS-S] J. Clunie and T. Sheil-Small, *Harmonic univalent functions*, Ann. Acad. Sci. Fenn. Ser. AI Math., **9** (1984), 3-25.

[F1] R. Finn, *New estimates for equations of minimal surface type*, Arch. Rat. Mech. Anal., **14** (1963), 337-375.

[F2] ______, *Remarks relevant to minimal surfaces and to surfaces of prescribed mean curvature*, J. d'Analyse Math., **14** (1965), 139-160.

[FO] R. Finn and R. Osserman, *On the Gauss curvature of non-parametric minimal surfaces*, J. d'Analyse Math., **12** (1964), 351-364.

[H] E. Heinz, *Über die Losungen der Minimalflachengleichung*, Nachr. Akad. Wiss. Gottingen Math. Phys., **K1** (1952), 51-56.

[HS1] W. Hengartner and G. Schober, *Harmonic mappings with given dilatation*, J. Lond. Math. Soc., **33**(**2**) (1986), 473-483.

[HS2] ______, *On the boundary behavior of orientation-preserving harmonic mappings*, Complex Variables, **5** (1986), 197-208.

[HS3] ______, *Curvature estimates for some minimal surfaces*, in 'Complex Analysis', Birkhaüser Verlag, 1988, 87-100.

[JS] H. Jenkins and J. Serrin, *Variational problems of minimal surface type* II. *Boundary value problems for the minimal surface equation*, Arch. Rat. Mech. Anal., **21** (1965/66), 321-342.

[K] H. Kneser, *Lösung der Aufgabe* 41, Jahresber. Deutsch. Math. Verein., **35** (1925), 123-124.

[L] R. Laugesen, *Planar harmonic maps with inner and Blaschke dilatations*, J. Lond. Math. Soc., **56**(**2**) (1997), 37-48.

[O] R. Osserman, *A Survey of Minimal Surfaces*, Dover, 1986.

[Se] J. Serrin, *A priori estimates for solutions of the minimal surface equation*, Arch. Rat. Mech. Anal., **14** (1963), 376-383.

[S-S] T. Sheil-Small, *On the Fourier series of a step function*, Mich. Math. J., **36** (1989), 459-475.

[Sp] G. Springer, *Introduction to Riemann Surfaces*, Addison-Wesley, 1957.

[W] A. Weitsman, *A counterexample to uniqueness in the Riemann mapping theorem for univalent harmonic mappings*, Bull. Lond. Math. Soc., **31** (1999), 87-89.

Received March 9, 1998.

Purdue University
West Lafayette, IN 47907-1395
E-mail address: weits@math.purdue.edu

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 192 No. 1 January 2000

0030-8730(200001)192:1;1-A