TRANSACTIONS OF THE
AMERICAN MATHEMATICAL SOCIETY
Volume 324, Number 1, March 1991

MAXIMAL REPRESENTATIONS OF SURFACE GROUPS IN BOUNDED SYMMETRIC DOMAINS

LUIS HERNÁNDEZ

ABSTRACT. Let Γ be the fundamental group of a hyperbolic surface of genus g; for $1 \le p \le q$, $PSU(p, q)$ is the group of isometries of a certain Hermitian symmetric space $D_{p,q}$ of rank p. There exists a characteristic number $c : \mathrm{Hom}(\Gamma, PSU(p, q)) \to \mathbb{R}$, which is constant on each connected component and such that $|c(\rho)| \le 4p\pi(g-1)$ for every representation ρ. We show that representations with maximal characteristic number (plus some nondegeneracy condition if $p > 2$) leave invariant a totally geodesic subspace of $D_{p,q}$ isometric to $D_{p,p}$.

1. INTRODUCTION

Let S be a closed Riemann surface of genus $g > 1$ and let Γ denote its fundamental group. Given a semisimple Lie group G, one wishes to understand both the topology of the space $\mathrm{Hom}(\Gamma, G)$ as well as that of the space of conjugacy classes $\mathrm{Hom}(\Gamma, G)/G$. For the case when G is compact several results are known; for example, when $G = U(n)$ (the unitary group) M. S. Narasimhan and C. S. Seshadri prove that the space of conjugacy classes $\mathrm{Hom}(\Gamma, U(n))/U(n)$ coincides with the set of rank n semistable holomorphic vector bundles of degree zero over S [NS]. Moreover, they show that this space carries the structure of a complex projective variety.

In this article we turn our attention to the case when G is not compact, and in particular, when it is the group of isometries of a certain bounded symmetric domain. See [Go2] for general facts about these spaces as well as for an exposition of the case when G is $PSL(2, \mathbb{R})$.

Let D denote the hyperbolic plane (i.e., the universal cover of S) and, for $p \le q$, let $D_{p,q} = \{q \times p$ complex matrices $Z |\ I_p - Z^*Z > 0\}$, equipped with its Bergmann metric and Kähler form ω. $D_{p,q}$ is a Hermitian symmetric space of rank p whose group of isometries may be identified with $PSU(p, q)$.

Any representation $\rho: \Gamma \to PSU(p, q)$ determines a flat bundle over S with fiber $D_{p,q}$, by taking the twisted product $D \times_\rho D_{p,q}$ over S. Since $D_{p,q}$ is contractible, this bundle has a section that is equivalent to an equivariant map

$$f: D \to D_{p,q}, \qquad f(\gamma \cdot x) = \rho(\gamma) \cdot f(x) \quad \forall\, \gamma \in \Gamma.$$

Received by the editors January 29, 1990.
1980 *Mathematics Subject Classification* (1985 *Revision*). Primary 32M15.

Let F be a fundamental domain for the action of Γ on D. For the representation ρ, define the characteristic number

$$c(\rho) = \int_F f^*\omega,$$

which is independent of the choice of f.

Example 1.1. Fix a faithful discrete representation ψ of Γ onto a co-compact subgroup of $PSU(1,1) \approx PSL(2,\mathbb{R})$; then, for $k = 1, \dots, p$ let

$$\rho_k : \Gamma \to \underbrace{PSU(1,1) \times \cdots \times PSU(1,1)}_{p \text{ times}} \subset PSU(p,q)$$

be defined by

$$\rho_k(\gamma) = \begin{pmatrix} aI_k & 0 & bI_k & 0 \\ 0 & I_{q-k} & 0 & 0 \\ cI_k & 0 & dI_k & 0 \\ 0 & 0 & 0 & I_{p-k} \end{pmatrix},$$

where $\psi(\gamma)$ is represented by $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$. Then

$$f : D \to \underbrace{D \times \cdots \times D}_{p \text{ times}} \subset D_{p,q}$$

may be chosen as

$$f(z) = \begin{pmatrix} zI_k & 0 \\ 0 & 0 \end{pmatrix}$$

and, so, $c(\rho_k) = k4\pi(g-1)$.

That one cannot construct a representation with larger characteristic number than $c(\rho_p)$ follows from the inequality

$$|c(\rho)| \le 4p\pi(g-1) \tag{1.1}$$

proved by Domic and Toledo in [DT].

Definition 1.2. We say that a representation ρ is maximal if $|c(\rho)| = 4p\pi(g-1)$.

For $p = q = 1$ inequality (1.1) was proved by J. W. Wood [Wd], based on a similar inequality by J. W. Milnor [M] for the group $SL(2,\mathbb{R})$. Milnor-Wood's inequality implies a classical result of H. Kneser [K]: *Let $\phi: S \to S'$ be a map between two hyperbolic surfaces; then*

$$|\deg\phi| \operatorname{vol}(S') = \left|\int_S \phi^*\omega\right| \le \operatorname{vol}(S)$$

by choosing the representation induced by the map ϕ; moreover, equality implies that ϕ is homotopic to a covering map.

For representations, the case of equality in Milnor-Wood's inequality was settled by W. Goldman [Go1]. He proves that *ρ is maximal if and only if ρ*

is an isomorphism of Γ *onto a discrete co-compact subgroup of* $PSL(2, \mathbb{R})$. In particular, the Teichmüller space of S is equal to the set of equivalence classes of representations ρ with $c(\rho) = 4\pi(g-1)$.

In a different direction, M. Gromov has generalized Kneser's inequality for maps between hyperbolic manifolds of higher dimension. Also, the case of equality was settled: *Let* $\phi\colon M \to N$ *be a map between closed hyperbolic manifolds of dimension* $n \geq 3$. *Suppose that* $|\deg\phi|\ \operatorname{vol}(N) = \operatorname{vol}(M)$; *then* ϕ *is homotopic to a local isometry.*

It was the proof of this theorem given by W. Thurston [Th, Theorem 6.4] that first suggested to D. Toledo that maximal representations into $PSU(1, q)$ should also satisfy some rigidity condition. He shows in [To] that *if* $\rho\colon \Gamma \to PSU(1, q)$ *is maximal then* $\rho(\Gamma)$ *leaves invariant a complex line in the ball* $B^q = D_{1,q}$.

Putting together Goldman's and Toledo's results it follows that the set of isomorphism classes of flat $PSU(1, q)$-bundles over S with $c(\rho) = 4\pi(g-1)$ forms a component of the space of all flat $PSU(1, q)$-bundles, and that this component is equal to the Cartesian product of the Teichmüller space of S with the space of isomorphism classes of flat $U(q-1)$-bundles over S.

Our work deals with possible generalizations of the above results to the case of rank bigger than one. The main result is the following.

Theorem 1.3. *Let* $\rho\colon \Gamma \to PSU(2, q)$ *be a maximal representation. Then* $\rho(\Gamma)$ *leaves invariant a totally geodesic subspace isometric to* $D_{2,2}$.

In view of this result, the study of maximal representations in $PSU(2, q)$ is reduced to that of maximal ones in $PSU(2, 2)$.

A key step in the proof of this theorem is the existence of an extension of the map f to the boundary of D; it was first constructed in [To] for $p = 1$. The precise statement is

Theorem 1.4. *For any* p, q *if* ρ *is maximal then the equivariant map* $f\colon D \to D_{p,q}$ *has an equivariant measurable extension* $\overline{f}\colon \overline{D} \to \overline{D}_{p,q}$, *given by radial limits. Moreover, the image of* ∂D *is contained in the Šilov boundary of* $D_{p,q}$. *(Here* $\overline{D}$ *and* $\overline{D}_{p,q}$ *denote the closures of* D, $D_{p,q}$ *in* $\mathbb{C}$ *and* $\mathbb{C}^{pq}$ *respectively.)*

It is worth mentioning that although a theorem of R. Zimmer [Z, Theorem 4.3.9] guarantees the existence of a Γ-equivariant measurable function between ∂D and some boundary of $D_{p,q}$, one cannot identify the boundary or how this function is related to our original map f.

The estimates needed for the proof of the extension theorem come from two sources. On one hand (§§2 and 3), those coming from the study of "large" triangles in D and $D_{p,q}$; a large triangle being one where the integral of the Kähler form ω is almost maximal (i.e., close to π and $p\pi$ respectively [DT]). This argument has its roots in [Th, Theorem 6.4], where almost ideal 3-simplices are studied to construct a map extension. The second type of estimates—Lemma

4.1—are obtained by the interplay between the geometry of $D_{p,q}$ and its Euclidean geometry as a subset of $\mathbb{C}^{pq}$.

In §5 the extension map is used to construct the limit set $\mathscr{R}$ of values of $\overline{f}$ at those points of ∂D where it is defined. We say that $\mathscr{R}$ is nondegenerate if it contains two limit points U and V with $\det(I - U^*V) \neq 0$. It is then showed that

Theorem 1.5. *If ρ is maximal and $\mathscr{R}$ is nondegenerate then $\rho(\Gamma)$ leaves invariant a totally geodesic subspace isometric to $D_{p,p}$.*

Up to this point, everything is valid for arbitrary $p \leq q$. However it is only in the case $p = 2$ that we have been able, so far, to remove the nondegeneracy of $\mathscr{R}$ as hypothesis in the above result. In fact, it can be shown that, when $p = 2$, $\mathscr{R}$ degenerate implies that $\rho(\Gamma)$ stabilizes a boundary component; whether this remains true for other values of p is still unclear. Finally, it is not hard to see that if $\rho(\Gamma)$ stabilizes a boundary component, ρ cannot be maximal.

2. Geometry of $D_{p,q}$

2.1. Basic properties. As a Hermitian symmetric space

$$D_{p,q} = SU(p,q)/S(U(p) \times U(q)),$$

where

$$SU(p,q) = \{M \in SL(p+q, \mathbb{C}) | M^*HM = H\}$$

and

$$H = \begin{pmatrix} -I_q & 0 \\ 0 & I_p \end{pmatrix}.$$

A model for $D_{p,q}$ is the bounded domain, in $\mathbb{C}^{pq}$,

$$D_{p,q} = \{Z = q \times p \text{complex matrix} | I_p - Z^*Z > 0\}.$$

An element $\left(\begin{smallmatrix} A & B \\ C & D \end{smallmatrix}\right) \in SU(p,q)$ acts as a generalized Möbius transformation, that is

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix} \cdot Z = (AZ+B)(CZ+D)^{-1}.$$

$D_{p,q}$ admits a natural compactification by taking its closure in $\mathbb{C}^{pq}$, and so the boundary of $D_{p,q}$, $\partial D_{p,q}$, consists of those matrices W with $I_p - W^*W \geq 0$ and $\det(I_p - W^*W) = 0$. For all $0 \leq r \leq p$ define

$$\mathscr{F}_r = \left\{ W \in \partial D_{p,q} | W = \left(\begin{smallmatrix} I_r & 0 \\ 0 & Z \end{smallmatrix}\right) \text{ with } I_{p-r} - Z^*Z > 0 \right\}.$$

Definition 2.1. The sets of the form $g\mathscr{F}_r$, where g is an isometry, are called the boundary components of $D_{p,q}$ of rank $p-r$. The Šilov boundary of $D_{p,q}$, denoted $\check{S}$, is the set of $q \times p$ complex matrices U satisfying $U^*U = I_p$ (i.e., is the union of all the rank zero boundary components).

Another useful description of this space is to view it as an open subspace of its compact dual, the Grassmannian of complex p-planes in $\mathbb{C}^{p+q}$. If h denotes the Hermitian form with matrix H, then the set in question is

$$\mathscr{D}_{p,q} = \{p\text{-planes } \Pi \subset \mathbb{C}^{p+q} | h|_{\Pi} > 0\}.$$

The two models are related by Borel's embedding $Z \to [{Z \atop I_p}]$, where $[{Z \atop I_p}]$ represents the p-plane spanned by the columns of the matrix $({Z \atop I_p})$. We will use these two models interchangeably as convenience will dictate.

The following facts about the geometry of $D_{p,q}$ will be assumed; a good reference is [P]:

1. The set

$$F_0 = \left\{ X \in D_{p,q} | X = \begin{pmatrix} R \\ 0 \end{pmatrix},\ R = \begin{pmatrix} r_1 & 0 & \dots & 0 \\ 0 & r_2 & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \dots & 0 & r_p \end{pmatrix},\ r_j \in \mathbb{R} \right\}$$

is a maximal flat totally geodesic subspace of $D_{p,q}$. In particular, if $p > 1$ there are sectional curvatures equal to zero. Let F_0^+ denote the subset for which all $r_j \geq 0$.

2. Any geodesic in $D_{p,q}$ is equivalent to one in F_0; every geodesic in F_0 is of the form $({R_t \atop 0})$ with

$$R_t = \begin{pmatrix} \tanh \alpha_1 t & 0 & \dots & 0 \\ 0 & \tanh \alpha_2 t & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \dots & 0 & \tanh \alpha_p t \end{pmatrix}, \qquad \sum \alpha_j^2 = 1.$$

3. Let $X = ({R \atop 0})$ with R a diagonal matrix as in 1. Then

$$d_{D_{p,q}}(O, X) = \sqrt{\sum_{j=1}^{p} \ln^2 \left(\frac{1 + r_j}{1 - r_j} \right)}.$$

4. Given $Z \in D_{p,q}$ there exists $k \in S(U(p) \times U(q))$ such that $k \cdot Z \in F_0^+$.

2.2. Estimates for almost maximal triangles. For $X, Y, Z \in D_{p,q}$ let $\langle X, Y, Z \rangle$ denote a geodesic triangle with vertices X, Y, and Z. By this we mean any 2-simplex whose edges are the geodesic segments joining the vertices. Since all that we will be interested in about these triangles is the integral of the Kähler form ω on them, the way we fill it is immaterial. One could use the center of gravity construction as explained in [DT].

Lemma 2.2. *Let $X, Y, Z \in D_{p,q}$. Then $\left|\int_{\langle X, Y, Z \rangle} \omega\right| < p\pi$.*

Proof. In [DT] the following formula was obtained:

$$\int_{\langle O,X,Y\rangle}\omega = -2\arg(\det(I_p - X^*Y)) = -2\sum_{j=1}^{p}\arg(1-\lambda_j)\ , \tag{2.3}$$

where the λ_j are the eigenvalues of X^*Y. Now note that since $|\lambda_j| < 1$ then $1-\lambda_j \in 1+D$ and so $|\arg(1-\lambda_j)| < \pi/2$. □

Proposition 2.3. *Let $X, Y \in D_{p,q}$ and assume that $t|\int_{\langle O,X,Y\rangle}\omega| \geq p\pi - e^{-t/2}$ for $t \gg 0$. Then:*

(a) *If λ is an eigenvalue of X^*Y then $|\lambda| \geq 1 - e^{-t}/8$.*

(b) *If μ is an eigenvalue of X^*X then $\sqrt{\mu} \geq 1 - \frac{p}{8}e^{-t}$.*

(c) *$d_{D_{p,q}}(O, X) > t$.*

(d) *Let T be the unit tangent vector at O in the direction of X, let α be an eigenvalue of T^*T, and let $K > 0$ be such that $d(O, X) < Kt$. Then $|\alpha| > 1/3K$, and, in particular, T^*T is invertible.*

Proof. (a) Formula (2.3) above implies that $\sum_{j=1}^{p} 2|\arg(1-\lambda_j)| \geq p\pi - e^{-t/2}$. Also we know that for all j, $|\arg(1-\lambda_j)| \leq \pi/2$, hence $2|\arg(1-\lambda)| \geq \pi - e^{-t/2}$; from this now follows the result (Lemma 1.5 in [To]).

(b) We may assume that X is of the form $\binom{R}{0}$, where R is a $p\times p$ diagonal matrix with real entries.

Now, YY^* is of the form $\left(\begin{smallmatrix} C & * \\ * & * \end{smallmatrix}\right)$ with C a $p\times p$ matrix. Note that the determinant of C, $\det C$, is real and smaller than 1; the latter follows because $I_q - YY^*$ (and so, $I_p - C$) is a positive definite matrix. Then

$$\mu > \det X^*X = \det R^2 > \det RCR = \det X^*Y \cdot \det Y^*X = \prod_{j=1}^{p}|\lambda_j|^2,$$

and so $\sqrt{\mu} > (1 - e^{-t}/8)^p$ by (a); or, for t sufficiently large, $\sqrt{\mu} > 1 - \frac{p}{8}e^{-t}$.

(c) Assume again that

$$X = \begin{pmatrix} R \\ 0 \end{pmatrix}, \qquad R = \begin{pmatrix} r_1 & 0 & \dots & 0 \\ 0 & r_2 & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \dots & 0 & r_p \end{pmatrix}, \qquad r_j \geq 0,$$

for all j. Then

$$\begin{aligned} d_{D_{p,q}}(O, X) &= \sqrt{\sum_{j=1}^{p}\ln^2\left(\frac{1+r_j}{1-r_j}\right)} \\ &\geq \sqrt{p\ln^2\left(\frac{1+(1-\frac{p}{8}e^{-t})}{1-(1-\frac{p}{8}e^{-t})}\right)} \\ &= \sqrt{p}\ln\left(\frac{16}{p}e^t - 1\right) > \sqrt{p}\left(t + \ln\frac{15}{p}\right) > t \end{aligned}$$

for t sufficiently large.

(d) Let X be as in (c). Then

$$T = \begin{pmatrix} A \\ 0 \end{pmatrix}, \qquad A = \begin{pmatrix} \alpha_1 & 0 & \dots & 0 \\ 0 & \alpha_2 & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \dots & 0 & \alpha_p \end{pmatrix},$$

and $r_j = \tanh(\alpha_j s)$ $\forall j$, where $t < s < Kt$.

By (b) $\tanh(\alpha_j Kt) > \tanh(\alpha_j s) = r_j > 1 - \frac{p}{8}e^{-t}$ $\forall j$, which implies that

$$e^{2\alpha Kt} > \frac{15}{p}e^t \Rightarrow 2\alpha Kt > t + \ln\frac{15}{p} \Rightarrow \alpha > \frac{1}{2K} + \frac{1}{2Kt}\ln\frac{15}{p}.$$

Hence, for t sufficiently large, $\alpha > 1/3K$. □

3. Measure cycles

The cohomology of S may be computed using the complex of "straight cochains". This is the complex of Borel measurable functions $b\colon D\times\cdots\times D \to \mathbb{R}$ that are invariant under the action of Γ. Straight k-cochains are therefore functions on $\Gamma\backslash D^{k+1}$. The coboundary operator in this complex is defined by the usual formula

$$\delta b(x_0, \dots, x_{p+1}) = \sum(-1)^i b(x_0, \dots, \hat{x}_i, \dots, x_{p+1}).$$

Similarly, finite Borel measures on $\Gamma\backslash D^{k+1}$ are called straight chains; the homology of S can be obtained using this complex, whose boundary operator is defined by duality from the above δ operator.

The class of the 2-form $f^*\omega$ defined in the introduction may be represented by the straight 2-cochain, denoted also $f^*\omega$, given by the formula

$$f^*\omega\langle x_0, x_1, x_2\rangle = \int_{\langle fx_0, fx_1, fx_2\rangle} \omega,$$

where $\langle fx_0, fx_1, fx_2\rangle$ denotes a geodesic triangle in $D_{p,q}$ with vertices fx_0, fx_1, fx_2. The reader is refered to [Gr and Th] for more details on measure cycles.

We write G for the full group of isometries of D; it has two components, the identity component being $PSL(2, \mathbb{R})$. Let Ω denote the quotient space $\Gamma\backslash G$ and let μ denote the Haar measure on G, normalized so that $\mu(\Omega) = \mathrm{Area}(S) = 4\pi(g-1)$.

Choose any geodesic ray from 0 in D, and let σ_n be the equilateral triangle in D, with sides of length n, first vertex at 0, and second vertex on the geodesic ray; let A_n denote its area. Let ζ_n denote the measure cycle consisting of all G-translates of σ_n, each weighted with coefficient $1/A_n$. That is, ζ_n is the linear functional on straight 2-cochains on S whose value on the cochain b is given by

$$b(\zeta_n) = \int_\Omega \frac{1}{A_n} b(g\sigma_n)\, d\mu(g)\,.$$

This measure chain is actually a cycle because each edge in its boundary belongs to precisely two translates of σ_n, with opposite orientations. To compute the homology class it represents, we evaluate it on dA, the area form of S:

$$dA(\zeta_n) = \mu(\Omega) = 4\pi(g-1) = dA([S])\,;$$

therefore, ζ_n represents the fundamental cycle $[S]$.

From this it follows that, if ρ is a representation with maximum characteristic number, then for all n

$$4p\pi(g-1) = |c(\rho)| = |f^*\omega(\zeta_n)| = \left|\int_\Omega \frac{1}{A_n} f^*\omega(g\sigma_n)\, d\mu\right|\,,$$

or equivalently,

$$\left|\int_\Omega \int_{st(fg\sigma_n)} \omega\, d\mu\right| = pA_n\mu(\Omega)\,, \tag{3.1}$$

where $st(fg\sigma_n)$ denotes a geodesic triangle in $D_{p,q}$ with vertices f (vertices of $g\sigma_n$).

The remainder of the section is devoted to the proof of two technical lemmas.

Lemma 3.1. *$\pi - A_n \sim 6e^{-n/2}$ as $n \to \infty$.*

Proof. If α is an interior angle of σ_n, then the hyperbolic law of cosines implies that

$$\cos\alpha = \frac{\cosh n}{1+\cosh n}\,;$$

so, for large n,

$$1 - \frac{\alpha^2}{2} \sim \frac{\cosh n}{1+\cosh n}\,,$$

or equivalently, $\alpha \sim 2e^{-n/2}$. Now, $A_n = \pi - 3\alpha$. □

Lemma 3.2. *Let $\varepsilon_n = p(\pi - A_n)$, and let $\Psi = \{g \in \Omega : |\int_{st(fg\sigma_n)} \omega| < pA_n - n^2\varepsilon_n\}$. Then $\mu(\Psi) < \mu(\Omega)/n^2$.*

Proof. Let $\phi(g) = \left|\int_{st(fg\sigma_n)} \omega\right|$. Since $\phi(g) < pA_n - n^2\varepsilon_n$ on Ψ, and $\phi(g) < pA_n + \varepsilon_n = p\pi$ on Ω, it follows from (3.1) that

$$\begin{aligned} pA_n\mu(\Omega) &= \int_\Omega \phi(g)\, d\mu = \int_\Psi \phi(g)\, d\mu + \int_{\Omega-\Psi} \phi(g)\, d\mu \\ &< pA_n\mu(\Omega) + \varepsilon_n(\mu(\Omega - \Psi) - n^2\mu(\Psi))\,; \end{aligned}$$

therefore, $\mu(\Omega) - n^2\mu(\Psi) > \mu(\Omega - \Psi) - n^2\mu(\Psi) > 0$. □

4. The extension to the boundary

In this section we prove the extension theorem for the map $f\colon D \to D_{p,q}$ mentioned in the introduction. In order to do this, besides the estimates of the previous sections, it will be necessary to know the rate of decay (with respect to the Euclidean metric of $\mathbb{C}^{pq}$) of certain geodesic balls in $D_{p,q}$. The precise statement is the following

Lemma 4.1. *Let $X \in D_{p,q}$ and let $T \in T_O D_{p,q}$ be the unit tangent vector at O in the direction of X. Let $B_r(X) = \{Y \in D_{p,q} \mid d_{D_{p,q}}(X, Y) < r\}$ and let d = Euclidean diameter of $B_r(X)$ (i.e., $d = \sup_{Z, Z' \in B_r(X)} \|Z - Z'\|_E$, where $\|Z\|_E = \sqrt{\frac{1}{p}\operatorname{tr} Z^* Z}$). Let $t = d_{D_{p,q}}(O, X)$ and assume that $\alpha > 0$ is the smallest eigenvalue of T^*T. Then $d < Ce^{-\alpha t}$, where $C > 0$ only depends on r.*

Proof. The operator norm, $\|Z\| = \sup_{x \neq 0} \sqrt{x^* Z^* Z x / x^* x}$, is equivalent to the Euclidean norm above; hence, it is enough to prove the result for it.

Since the isotropy subgroup at O of $\operatorname{Iso}(D_{p,q})$ acts as isometries for the operator norm as well, we can assume that

$$X = \begin{pmatrix} R \\ 0 \end{pmatrix}, \qquad T = \begin{pmatrix} A \\ 0 \end{pmatrix},$$

$$R = \begin{pmatrix} r_1 & 0 & \dots & 0 \\ 0 & r_2 & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \dots & 0 & r_p \end{pmatrix}, \qquad A = \begin{pmatrix} \alpha_1 & 0 & \dots & 0 \\ 0 & \alpha_2 & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \dots & 0 & \alpha_p \end{pmatrix},$$

and $r_j = \tanh(\alpha_j t) > 0$ for all j.

Let

$$g = \begin{pmatrix} (I_q - XX^*)^{-1/2} & (I_q - XX^*)^{-1/2} X \\ X^*(I_q - XX^*)^{-1/2} & (I_p - X^*X)^{-1/2} \end{pmatrix},$$

g is an isometry mapping O to X.

$$XX^* = \begin{pmatrix} R^2 & 0 \\ 0 & 0 \end{pmatrix} \Rightarrow g = \begin{pmatrix} (I - R^2)^{-1/2} & 0 & (I - R^2)^{-1/2} R \\ 0 & I & 0 \\ R(I - R^2)^{-1/2} & 0 & (I - R^2)^{-1/2} \end{pmatrix}.$$

If

$$Z = \begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix} \in B_r(O),$$

then

$$g(Z) = \begin{pmatrix} (I - R^2)^{-1/2}(Z_1 + R) \\ Z_2 \end{pmatrix} \left[(I - R^2)^{-1/2}(RZ_1 + I)\right]^{-1}.$$

Now, $\|X - g(Z)\|^2 \le \|M\|^2 + \|N\|^2$, where

$$M = R - (I - R^2)^{-1/2}(Z_1 + R)(I + RZ_1)^{-1}(I - R^2)^{1/2},$$

$$N = Z_2(I + RZ_1)^{-1}(I - R^2)^{1/2}.$$

Then

$$\begin{aligned} M &= R - (I - R^2)^{-1/2} R^{-1} (I + RZ_1 + R^2 - I)(I + RZ_1)^{-1} (I - R^2)^{1/2} \\ &= R - (I - R^2)^{-1/2} R^{-1} (I - R^2)^{1/2} \\ &\quad - (I - R^2)^{-1/2} R^{-1} (R^2 - I)(I + RZ_1)^{-1} (I - R^2)^{1/2} \\ &= R - R^{-1} + R^{-1} (I - R^2)^{1/2} (I + RZ_1)^{-1} (I - R^2)^{1/2}, \end{aligned}$$

and so

$$\|M\| \leq \|R - R^{-1}\| + \|R^{-1}\| \, \|I - R^2\| \, \|(I + RZ_1)^{-1}\|$$

and

$$\|N\| \leq \|Z_2\| \, \|(I + RZ_1)^{-1}\| \, \|I - R^2\|^{1/2}.$$

Since $Z \in B_r(O)$, there is a constant C_0 depending only on r such that $\|Z\| < C_0 < 1$, hence $\|(I + RZ_1)^{-1}\| < 1/(1 - C_0)$. Also, for all t greater than a fixed $t_0 > 0$:

(i) $\|R^{-1}\|$ is bounded (e.g., $\|R^{-1}\| < 2$ if $\tanh \alpha t > 1/2$).

(ii) $\|I - R^2\| \leq 1/\cosh^2 \alpha t < 4e^{-2\alpha t}$.

(iii) $\|R - R^{-1}\| \leq 1/\sinh \alpha t \cosh \alpha t < C_1 e^{-\alpha t}$.

Hence,

$$\|X - g(Z)\| \leq C(r) e^{-\alpha t}. \quad \square$$

The facts that Γ is co-compact and f is Γ-equivariant imply that f is a Lipschitz map. Also, there is no loss in assuming that $f(0) = O$ (to see this, let g be an isometry mapping $f(0)$ to O; replace f by $f' = gf$ and ρ by $\rho' = g\rho g^{-1}$, then f' is ρ'-equivariant and if D_0 is invariant under ρ' then $g^{-1} D_0$ is invariant under ρ).

Theorem 4.2. *The map f has radial limits for almost every geodesic ray emanating from* 0. *Moreover, all these limits lie in the Šilov boundary of $D_{p,q}$.*

Proof. Using the estimates at the end of §3 one sees that, for every N sufficiently large and for all triangles $g\sigma_n$, except for those in a set of measure $4\pi(g - 1) \sum_{j=N}^{\infty} 1/j^2$, the inequality

$$\left| \int_{st(fg\sigma_n)} \omega \right| \geq p\pi - e^{-n/2}$$

holds $\forall n > N$. Letting $N \to \infty$ we find that for almost every $g \in \Omega$ the above inequality holds for all $n > N_g$.

In particular, there is a point, say 0, for which the inequality holds for almost every triangle $g\sigma_n$ ($n > N_g$) based at 0. Looking at the first side of each of these triangles, we conclude that for almost every geodesic ray γ emanating from 0, the following are true:

1. Proposition 2.3(c) implies that for $n \gg 0$, $d_{D_{p,q}}(O, f \circ \gamma(n)) > n$.

2. Since f is Lipschitz, there exists $K > 0$ such that $d_{D_{p,q}}(O, f \circ \gamma(n)) < Kn$.

3. Proposition 2.3(d) now implies that for $n \gg 0$ the geodesic joining O and $Z_n = f \circ \gamma(n)$ has tangent vector at O, T_n, such that the eigenvalues of $T_n^* T_n$ are all bounded below by $1/3K$.

If $p = 1$ all sectional curvatures of $D_{1,q}$ are $\leq -1/4$, so the sequence of angles $\angle(T_n, T_{n+1})$ decreases exponentially and $\{Z_n\}$ converges.

In the case that $p > 1$, we do not have control anymore over the sequence of angles. For this case we now invoke Lemma 3.1 above, which tells us that the Euclidean distance between Z_n and Z_{n+1} decreases as $e^{-n/3K}$.

Let h_n be the geodesic ray in the direction of T_n, and let $U_n = h_n(+\infty)$ be the point where it intersects the boundary of $D_{p,q}$. Condition 3 above implies that for $n \gg 0$, $T_n^* T_n$ is invertible and, so, that U_n is in $\check{S}$. To see this note that the isotropy group at O leaves invariant $\check{S}$ and preserves the nonsingularity of $T_n^* T_n$.

The fact that $d(O, Z_n) > n$ gives that $\|U_n - Z_n\|$ decreases like $e^{-n/3K}$. Now it follows that $\|U_n - U_{n+1}\|$ goes to zero exponentially and, hence, that $\{U_n\}$ converges to some U; the compactness of $\check{S}$ implies that U is also inside it.

It is clear now that $Z_n \to U$. $\square$

Theorem 4.2 above provides us with a measurable extension that we will denote by $\overline{f}$. Note that, since f is Lipschitz, $\overline{f}$ will be continuous along every geodesic converging to almost every point on ∂D, and not only along rays from 0.

5. The action of $\rho(\Gamma)$ on $D_{p,q}$

5.1. Boundary behavior. Let $\check{S}$ denote the Šilov boundary of $D_{p,q}$; also let

$$D_0 = \left\{ \left(\begin{smallmatrix} Z \\ 0 \end{smallmatrix}\right) | Z \text{ is a } p \times p \text{ complex matrix with } I_p - Z^* Z > 0 \right\}.$$

Let $U, V \in \check{S}$; under the Borel embedding they take the form $[\begin{smallmatrix} U \\ I_p \end{smallmatrix}]$ and $[\begin{smallmatrix} V \\ I_p \end{smallmatrix}]$. Define Π_{UV} as the subspace of $\mathbb{C}^{p+q}$ spanned by the columns of $[\begin{smallmatrix} U \\ I_p \end{smallmatrix}]$ and $[\begin{smallmatrix} V \\ I_p \end{smallmatrix}]$.

Let $h(z, w) = -z_1\overline{w}_1 - \cdots - z_q\overline{w}_q + z_{q+1}\overline{w}_{q+1} + \cdots + z_{q+p}\overline{w}_{q+p}$ be the Hermitian form in $\mathbb{C}^{p+q}$ preserved by $SU(p, q)$. Then $h|_{\Pi_{UV}}$ can be represented by the matrix

$$A = \begin{pmatrix} 0 & I - U^*V \\ I - V^*U & 0 \end{pmatrix}.$$

Lemma 5.1. *Let U, V, and A be as above. Then*

(1) $\det(A) = 0$ *if and only if* $\det(I - U^*V) = 0$.

(2) *The signature of A is zero.*

Proof. (1) $\det(A) = 0 \Rightarrow \exists (x, y) \neq (0, 0)$ such that $A\left(\begin{smallmatrix} x \\ y \end{smallmatrix}\right) = \left(\begin{smallmatrix} 0 \\ 0 \end{smallmatrix}\right)$; hence $(1 - U^*V)y = (1 - V^*U)x = 0$.

Conversely, if $\exists x \neq 0$ such that $(I - U^* V)x = 0$ then $A\binom{0}{x} = \binom{0}{0}$.

(2) The result will follow from the fact that for any $2p \times 2p$ matrix M with block decomposition $\begin{pmatrix} 0 & M_1 \\ M_2 & 0 \end{pmatrix}$, its characteristic polynomial has only even power terms.

Looking at the matrix $\lambda I - M$ we see that if one of the λ's of the upper left block does not appear in one of the monomials of the determinant, then either the monomial is zero or it contains an element of M_1 on the same row as λ; but then the λ on the same column as this element cannot be present in the monomial either. This shows that the number of λ's appearing in each nonzero monomial is even. Therefore the signature is zero. □

Corollary 5.2. *If $U, V \in \check{S}$ are such that* $\det(I - U^* V) \neq 0$, *then $h|_{\Pi_{UV}}$ is a nondegenerate Hermitian form of type (p, p).* □

Lemma 5.3. *Let $U, V \in \check{S}$. Then there exists an isometry g that takes both U and V into $\check{S} \cap \partial D_0$.*

Proof. By Lemma 5.1, $h|_{\Pi_{UV}}$ has at most p negative eigenvalues. Hence, there exist at least $q - p$ orthonormal (i.e., h-orthogonal and h-length -1) vectors orthogonal to Π_{UV}. Complete these vectors to a $\pm$-orthonormal basis of $\mathbb{C}^{p+q}$. This is the desired isometry. □

Now, given $p \times p$ unitary matrices U and V, one can find two matrices P, Q, also unitary, such that $PUQ^* = I_p$ and $PVQ^* = \Lambda$ with Λ also diagonal.

Lemma 5.4. *Let U, V be two points in $\check{S}$. Then, if* $\det(I - U^* V) \neq 0$ *there exists a geodesic in $D_{p,q}$ joining U and V.*

Proof. First of all, notice that the rank of the matrix $I - U^* V$ is invariant under isometries. So, by Lemma 5.3 above it is enough to prove the result for the case $U = \binom{I_p}{0}$, $V = \binom{\Lambda}{0}$, where Λ is a diagonal matrix. The condition $\det(I - U^* V) \neq 0$ says that the diagonal entries of Λ are all different from one; it is clear now that there exist geodesics inside the polydisk D^p joining U and V. □

Corollary 5.5. *If $U, V \in \check{S}$ and if* $\det(I - U^* V) \neq 0$, *then there exists an isometry sending U to $\binom{I_p}{0}$ and V to $\binom{-I_p}{0}$.*

Proof. First note that any geodesic through O is the image, under an element of the isotropy group at O, of one of the form $\binom{R_t}{0}$ with

$$R_t = \begin{pmatrix} \tanh \alpha_1 t & 0 & \dots & 0 \\ 0 & \tanh \alpha_2 t & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \dots & 0 & \tanh \alpha_p t \end{pmatrix},$$

and so its endpoints are antipodal.

U and V are joined by a geodesic. Send it to a geodesic through O. This maps U and V onto antipodal points; finally follow by a suitable element of the isotropy group at O. $\square$

We want to show now that given $U, V \in \check{S}$, $\det(I - U^* V) \neq 0$ implies that there exists a unique totally geodesic space, isometric to $D_{p,p}$, containing U and V in its boundary. By the preceding corollary it is enough to do so for $\binom{I_p}{0}$ and $\binom{-I_p}{0}$.

Proposition 5.6. *If $U, V \in \check{S}$, then $\det(I - U^* V) \neq 0$ implies that there exists a unique totally geodesic space, isometric to $D_{p,p}$, containing U and V in its boundary.*

Proof. Suppose $\binom{I_p}{0}$ and $\binom{-I_p}{0}$ are contained in the boundary of gD_0. Then there exist two points, L, M, in the Šilov boundary of D_0 such that $gL = \binom{I_p}{0}$ and $gM = \binom{-I_p}{0}$.

By Corollary 5.5 applied to D_0, there exists an isometry h leaving D_0 invariant and such that $h\binom{I_p}{0} = L$ and $h\binom{-I_p}{0} = M$. Since $gD_0 = ghD_0$, we may as well assume that g fixes $\binom{I_p}{0}$ and $\binom{-I_p}{0}$. Such a g must be of the form

$$\begin{pmatrix} A_1 & A_2 & B_1 \\ 0 & A_4 & 0 \\ C_1 & C_2 & D \end{pmatrix},$$

but then clearly $gD_0 = D_0$. $\square$

5.2. The limit set. Recall that we have defined a measurable map $\overline{f}\colon \overline{D} \longrightarrow D_{p,q} \cup \check{S}$, given almost everywhere on ∂D by radial limits.

Let $\mathscr{R} \subset \check{S}$ denote the set of radial limits, referred to from now on as the limit set of ρ.

From (2.3) for a triangle based at O one can easily derive the general formula

$$\int_{\langle X, Y, Z\rangle} \omega = -2\sum_{j=1}^{p}(\arg(1-\lambda_j) + \arg(1-\mu_j) + \arg(1-\nu_j)),$$

where λ_j, μ_j, ν_j are the eigenvalues of X^*Y, Y^*Z, Z^*X, respectively.

Let $U, V, W \in \mathscr{R}$ be three limit points. Then the estimates on equilateral triangles in §3 show that we can find sequences $X_n, Y_n, Z_n \in D_{p,q}$ such that $X_n \to U$, $Y_n \to V$, $Z_n \to W$, and

$$\left|\int_{\langle X_n, Y_n, Z_n\rangle} \omega\right| \to p\pi.$$

Definition 5.7. We say that the set of limit points $\mathscr{R}$ is degenerate if for any pair of points $U, V \in \mathscr{R}$, $\det(I - U^* V) = 0$. Otherwise we say it is nondegenerate.

Assume that $\mathscr{R}$ is nondegenerate; then, by Corollary 5.5, we can further assume that $U = \binom{I_p}{0}$ and $V = \binom{-I_p}{0}$ belong to $\mathscr{R}$ (by replacing, if necessary,

f by gf, $\mathscr{R}$ by $g\mathscr{R}$, and ρ by $g\rho g^{-1}$; and noticing that D_0 is $g\rho g^{-1}$-invariant if and only if $g^{-1}D_0$ is ρ-invariant).

Note that $\int_{\langle X, Y, Z\rangle} \omega$ as a function defined over $D_{p,q} \times D_{p,q} \times D_{p,q}$ extends continuously to

$$\overline{D}_{p,q} \times \overline{D}_{p,q} \times \overline{D}_{p,q} \backslash \{(U, V, W) | \det(I - U^* V)(I - V^* W)(I - W^* U) = 0\}.$$

In particular, $|\int_{\langle U, V, Z_n\rangle} \omega| \to p\pi$; on the other hand, the left-hand side converges to $-2\sum_{j=1}^{p}(\arg(1+\lambda_j) + \arg(1-\overline{\lambda}_j))$, where the λ_j are the eigenvalues of $(I\ 0)W$. This shows that these eigenvalues must all have norm 1, and so, that $(I\ 0)W$ is a unitary matrix; hence W must also be contained in ∂D_0. In other words:

Lemma 5.8. *If $\mathscr{R}$ is nondegenerate, then it is contained in ∂D_0.*

Theorem 5.9. *If ρ is a maximal representation whose limit set is nondegenerate, then the group $\rho(\Gamma)$ leaves invariant the subspace D_0.*

Proof. Let $g \in \rho(\Gamma)$; then $g^{-1}\binom{I_p}{0}$ and $g^{-1}\binom{-I_p}{0}$ belong both to $\mathscr{R}$. Hence, $\binom{I_p}{0}, \binom{-I_p}{0} \in gD_0$, but this implies $gD_0 = D_0$. □

5.3. Degeneracy of the limit set and parabolic groups. Here we finish the proof of Theorem 1.2 by showing that the limit set $\mathscr{R}$ of a maximal representation ρ in $PSU(2, q)$ is always nondegenerate (compare with Proposition 2.1 of [C]).

We will assume now that for all pairs $U, V \in \mathscr{R}$, 1 is an eigenvalue of U^*V. This includes the case when the extension map $\overline{f}$ is constant along ∂D.

Lemma 5.10. *If $p = 2$ then, $\mathscr{R}$ is contained in the boundary of a boundary component.*

Proof. We may assume that

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \\ \vdots & \vdots \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & \lambda \\ 0 & 0 \\ \vdots & \vdots \\ 0 & 0 \end{pmatrix} \in \mathscr{R}.$$

Let $V = \binom{V_1}{V_2} \in \mathscr{R}$ be another point, with $V_1 = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$. $V \in \mathscr{R}$ implies that $V_1 = (I_2\ 0)V$ has an eigenvalue 1, and so

$$1 - (a + d) + \det V_1 = 0.$$

Also $\begin{pmatrix} 1 & 0 \\ 0 & \overline{\lambda} \end{pmatrix} V_1$ must have an eigenvalue 1; hence

$$1 - (a + \overline{\lambda} d) + \overline{\lambda} \det V_1 = 0.$$

These two equations imply that $\lambda(1 - a) = 1 - a$, and since $\lambda \neq 1$, we must have $a = 1$. Since $V^* V = I_2$, $b = 0$.

This shows that

$$\mathscr{R} \subset \partial\mathscr{B}, \qquad \mathscr{B} = \left\{ \begin{pmatrix} 1 & 0 \\ 0 & w_1 \\ 0 & 0 \\ \vdots & \vdots \\ 0 & w_{q-1} \end{pmatrix} : \sum_{i=1}^{q-1} \overline{w}_i w_i < 1 \right\}. \quad \square$$

Corollary 5.11. $\rho(\Gamma)$ *leaves invariant a boundary component*

Proof. First of all, note that $\mathscr{R}$ is invariant under the action of $\rho(\Gamma)$. If $\mathscr{R}$ consists of a single point, then $\mathscr{R}$ itself is a boundary component. If $\mathscr{R}$ has more than one point, then $\mathscr{R} \subset \partial\mathscr{B}$, and so for every element $g \in \rho(\Gamma)$, $\mathscr{R} \subset \partial\mathscr{B} \cap \partial g\mathscr{B}$. This implies $\mathscr{B} = g\mathscr{B}$. $\square$

This corollary shows that $\rho(\Gamma)$ lies in the stabilizer P of a boundary component. It is shown in [Wf] that such a group is a maximal parabolic subgroup and so it can be decomposed as a semidirect product of a reductive subgroup and a unipotent subgroup of $PSU(2, q)$. Moreover, J. Wolf also shows [Wf, Boundary Group Theorem] that the Lie algebra of the reductive factor of P is the set

$$\{X \in \mathfrak{su}(2, q) | \, [X, E_{q\,q+1} + E_{q+1\,q}] = 0\},$$

where E_{ij} is the matrix with a 1 on the (i, j) entry and zero everywhere else. More explicitly, let $X = \left(\begin{smallmatrix} A & B \\ B^* & D \end{smallmatrix}\right)$, $A \in \mathfrak{u}(q)$, $D \in \mathfrak{u}(2)$, and $\operatorname{tr} A + \operatorname{tr} D = 0$. Then the condition $[X, E_{q\,q+1} + E_{q+1\,q}] = 0$ implies that X is of the form

$$X = \begin{pmatrix} A' & 0 & 0 & b \\ 0 & is & r & 0 \\ 0 & r & is & 0 \\ b^* & 0 & 0 & it \end{pmatrix},$$

where $r, s, t \in \mathbb{R}$ and $A' \in \mathfrak{u}(q-1)$. From this it follows that the semisimple factor of P is isomorphic to $PSU(1, q-1)$.

Hence, the representation ρ can be deformed to another one, ρ', with image in $PSU(1, q-1)$. Inequality (1.1) applied to ρ' implies that $|c(\rho')| \leq 4\pi(g-1)$, but $c(\rho) = c(\rho')$, contradicting the maximality of ρ.

Acknowledgments. I want to thank my friend and adviser Domingo Toledo for his help and encouragement while this work was being done. I would like to express my gratitude to the Mathematics Department of the University of Utah for providing such a friendly and stimulating atmosphere.

References

[C] K. Corlette, *Rigid representations of Kählerian fundamental groups*, Preprint.

[DT] A. Domic and D. Toledo, *The Gromov norm of the Kähler class of symmetric domains*, Math. Ann. **276** (1987), 425–432.

[Gol] W. Goldman, *Discontinuous groups and the Euler class*, Doctoral Dissertation, University of California, Berkeley.

[Go2] ———, *Representations of fundamental groups of surfaces*, Lecture Notes in Math., vol. 1167, Springer, 1985, pp. 95–117.

[Gr] M. Gromov, *Volume and bounded cohomology*, Inst. Hautes Études Sci. Publ. Math. **56** (1982).

[K] H. Kneser, *Die kleinste Bedeckungszahl innerhalb einer Klasse von Flächenabbildungen*, Math. Ann. **103** (1930), 347–358.

[M] J. W. Milnor, *On the existence of a connection with curvature zero*, Comment Math. Helv. **32** (1957), 216–223.

[NS] M. S. Narasimhan and C. S. Seshadri, *Stable and unitary vector bundles on a compact Riemann surface*, Ann. of Math. **82** (1965), 540–567.

[P] I. I. Pyatetskii-Shapiro, *Automorphic functions and the geometry of classical domains*, Gordon and Breach, World Science Publishers, 1969.

[Th] W. Thurston, *The geometry and topology of 3-manifolds*, Princeton University notes.

[To] D. Toledo, *Representations of surface groups in $PSU(1, n)$ with maximum characteristic number*, J. Differential Geom. **29** (1989), 125–134.

[Wd] J. W. Wood, *Bundles with totally disconnected structure group*, Comment. Math. Helv. **51** (1976), 183–199.

[Wf] J. Wolf, *Fine structure of Hermitian symmetric spaces*, Symmetric Spaces (W. Boothby and G. Weiss, eds.), Marcel Dekker, New York, 1972.

[Z] R. J. Zimmer, *Ergodic theory and semi-simple groups*, Birkhäuser, Boston, Mass., 1984.

DEPTARTMENT OF MATHEMATICS, UNIVERSITY OF CHICAGO, CHICAGO, ILLINOIS 60637