Rigidity of lattices : An introduction

by M. GROMOV and P. PANSU

We present in these lectures basic facts and ideas in the geometry of discrete subgroups in Lie groups with a special emphasis laid upon the rigidity of lattices in the semi-simple Lie groups. We try to give a broad panorama of the field and explain various approaches to the study of the rigidity. These can be roughly divided into two categories. The first approach developed by G.D. Mostow [1967] and H. Fürstenberg [1967] (see $[\mathrm{Mos}]_1$ and $[\mathrm{Für}]_1$) uses the geometry and dynamics of the action of discrete subgroups $\Gamma \subset G$ on an appropriate *ideal boundary* of the ambient Lie group G. This geometro-dynamical line of development culminated in the *superrigidity and arithmeticity* theorems for lattices in the simple Lie groups of $\mathbb{R}$-rank ≥ 2, proven by G.A. Margulis in 1974 (see $[\mathrm{Mar}]_1$).

The second approach, more analytic in nature, uses elliptic P.D.E. and Bochner type integro-differential inequalities. The first results here obtained by E. Calabi and E. Vesentini [1960] and then by A. Weil [1962] (see [Cal], [Cal-Ves], [Wei]), dealt with local and infinitesimal deformations of $\Gamma \subset G$ and the P.D.E's involved were linear. The Calabi-Vesentini method was delinearized by Y.T. Siu in 1980 who found a Kodaira-Bochner-(Siu) identity for harmonic maps of Kähler manifolds into Riemannian manifolds satisfying certain negative curvature conditions. Using Siu's method, N. Mok [1989] was able to give an alternative proof of Margulis' superrigidity for the (lattices in the isometry groups G of) Hermitian symmetric spaces. More recently, K. Corlette [1990] has found a Bochner formula in the quaternionic case and thus established the superrigidy for discrete groups of isometries of the quaternionic hyperbolic spaces as well as for the Cayley plane (see $[\mathrm{Cor}]_1$). Notice that these cases are not covered by Margulis' theory.

Our presentation of both methods, geometro-dynamic and analytic, is quite superficial as we try to avoid all difficult spots as much as possible. Yet we provide the proofs of the basic elementary facts which convey the flavor of the deeper aspects of the theory.

We conclude this introduction with our thanks to Franco Tricerri and Paolo de Bartolomeis who brought us to the comfort of Montecatini where we could calmly present our lectures.

1. Generalities on Lie groups and discrete subgroups.

There are two sharply distinct classes of closed subgroups H in a connected Lie group G. The first case is where H is connected. Then by the classical theory (see [Pont]) H is a Lie (sub)-group which is uniquely determined by its Lie subalgebra $L(H) \subset L(G)$. Thus all geometric properties of the subgroup $H \subset G$ are encoded in the linear algebra of $L(H) \subset L(G)$. The decoding may be difficult at times, but, in principle, it is always possible.

The opposite case is where the subgroup is discrete in G. Here we prefer to use the notation Γ rather than H as this Γ is an animal of quite different nature from G.

The linear structure of $L(G)$ can not be used in the study of discrete subgroups $\Gamma \subset G$ as the infinitesimal information contained in $L(\Gamma)$ literally reduces to zero. Yet as we shall see in these lectures one can use global geometric methods and achieve a fair understanding of Γ in many cases.

1.1. Elementary example. Before going into any kind of general theory we want to revive in the reader's mind the familiar picture of a lattice in the plane. We take $G = \mathbb{R}^2$ and consider the subgroup $\Gamma \subset \mathbb{R}^2$ consisting of all integral combinations of two linearly independent vectors x and y in $\mathbb{R}^2$, that is

$$\Gamma = \{mx + ny\}_{m,n \in \mathbb{Z}} \subset \mathbb{R}^2 .$$

This Γ is obviously discrete and as an abstract group it is isomorphic to $\mathbb{Z}^2$. The quotient space $\mathbb{R}^2/\Gamma$ is compact. In fact, this quotient is, as everybody knows, the 2-torus.

There is another kind of discrete subgroup $\Gamma \subset \mathbb{R}^2$ where the quotient is non-compact. Namely we may take Γ consisting of the multiples of a single non-zero vector in $\mathbb{R}^2$,

$$\Gamma = \{mx\}_{m \in \mathbb{Z}} \subset \mathbb{R}^2 .$$

Here the quotient space is the infinite cylinder $S^1 \times \mathbb{R}^1$.

1.2. Definitions. A (discrete) subgroup $\Gamma \subset G$ is called *cocompact* if the quotient space G/Γ is compact. An equivalent condition is the existence of a *compact* subset $D \subset G$ whose Γ-translates cover G. That is

$$\bigcup_{\gamma \in \Gamma} \gamma D = G .$$

Yet another way to define the cocompactness is to require that Γ is *a net* in G for some (and hence for every) left invariant Riemannian metric in G, where a subset Γ in G (with a given metric) is called a net if there exists (possibly large) $\varepsilon \geq 0$, such that for every $g \in G$ there exists some $\gamma \in \Gamma$ such that $\operatorname{dist}(g,\gamma) \leq \varepsilon$. In other words, the ε-*neighbourhood*

$$U_\varepsilon(\Gamma) \underset{\text{def}}{=} \{g \in G \mid \operatorname{dist}(g,\Gamma) \leq \varepsilon\}$$

equals all of G.

The equivalence of the three definitions is rather obvious. What distinguishes the first definition is the appeal to the quotient space G/Γ (which perversely is sometimes denoted by $\Gamma \backslash G$) which a priori carries less information than the pair $(G, \Gamma \subset G)$. The last definition with nets is interesting as it makes sense for arbitrary metric spaces (where one usually specifies an ε and speaks of ε-nets).

An underlying intuitive idea of the above definition is that cocompact subgroups $\Gamma \subset G$ provide a fair discrete approximation of G with a bounded error and so their properties are expected to be similar to those of G. A more serious mathematical reason for introducing the notion of cocompactness is the existence of certain remarkable subgroups with this property as indicated in 1.4.B.

Our next definition also expresses the idea of approximation of G by Γ but now in terms of measure theory.

A discrete subroup $\Gamma \subset G$ is called a *lattice* (or a *finite covolume* subgroup) if G/Γ has *finite* volume, i.e. the measure on G/Γ induced by the Haar measure on G has finite total mass. An equivalent property is the existence of a non-trivial finite G-invariant measure on G/Γ. The third definition is the existence of a subset $D \subset G$ of finite Haar measure whose Γ-translates cover G. (In these definitions, if we speak of a left invariant Haar measure, we must use the left action of Γ on G. Or we could go to the right measure and action. Fortunately, the existence of a single lattice $\Gamma \subset G$ implies by an easy argument that every left invariant measure on G/Γ is right invariant which makes the left-right precautions unnecessary).

1.3. Simple example. Let us try the above definition on discrete subgroups Γ of the group $G = \mathbb{R}^n$. Every such Γ equals the integral span of some linearly independent vectors $x_1,\dots,x_k$ in $\mathbb{R}^n$, as an easy (and well known) argument shows. Then the quotient $\mathbb{R}^n/\Gamma$ is the Cartesian product of the torus $T^k = \mathbb{R}^k/\Gamma$ by $\mathbb{R}^{n-k}$ where $\mathbb{R}^k \subset \mathbb{R}^n$ is the span of $x_1,\dots,x_k$ and $\mathbb{R}^{n-k} = \mathbb{R}^n/\mathbb{R}^k$. Now one sees that the following four conditions are equivalent:

1. Γ is cocompact in $\mathbb{R}^n$.

2. Γ has finite covolume.

3. Γ is isomorphic, as an abstract group, to $\mathbb{Z}^n$.

4. The linear span of Γ in $\mathbb{R}^n$ equals $\mathbb{R}^n$.

1.4. Non-cocompact lattices. The above example may make one believe that

cocompact $\Leftrightarrow$ finite covolume.

In fact the implication

cocompact $\Rightarrow$ finite covolume

is true and obvious as every *compact* subset $D \subset G$ has finite haar measure. The opposite implication is known to be true for nilpotent and solvable Lie groups (see [Rag]) as well as for $G = \mathbb{R}^n$ treated above. Yet this is not so for *semisimple* Lie groups and the counter example is provided by the following

1.4.A. Most remarkable lattice. Take $G = SL_n\mathbb{R}$ and let $\Gamma = SL_n\mathbb{Z} \subset SL_n\mathbb{R}$ that is the subgroup consisting of the matrices with integral entries and having determinant one. For example, if $n = 2$, then $\Gamma = SL_2\mathbb{Z}$ equals the set of the *integral* solutions a, b, c, d to the equation $ab - cd = 1$.

1.4.A'. Theorem. $SL_n\mathbb{Z}$ *has finite covolume in* $SL_n\mathbb{R}$ *for* $n \geq 2$ *but is not cocompact.*

The proof of the finite covolume property follows from the Hermite-Minkowski reduction theory which provides an "almost orthogonal"basis in every lattice in $\mathbb{R}^n$ (see [Rag], [Cas]).

Notice that the finite covolume property of $\Gamma \subset G$ implies Γ is infinite in the (interesting) case where G is non-compact as $\infty = \text{Vol}\, G = (\#\,\Gamma)\, \text{Vol}(G/\Gamma)$. For example the Diophantine equation $ab - cd = 1$ has infinitely many solutions. This can also be seen directly by successively multiplying the matrices $\begin{pmatrix}1&0\\1&1\end{pmatrix}$ and $\begin{pmatrix}1&1\\0&1\end{pmatrix}$ in $SL_2\mathbb{Z}$.

1.4.A". Geometric interpretation of the quotient space $SL_n\mathbb{R}/SL_n\mathbb{Z}$. We claim that $GL_n\mathbb{R}/SL_n\mathbb{Z}$ equals the space of unimodular lattices in $\mathbb{R}^n$. To see that we observe the following two facts

(a) $SL_n\mathbb{Z}$ consists of exactly those linear maps $\mathbb{R}^n \to \mathbb{R}^n$ with determinant one which send the subgroup $\mathbb{Z}^n \subset \mathbb{R}^n$ (necessarily bijectively) into itself.

(b) Let $\mathcal{L} \subset \mathbb{R}^n$ be a *unimodular* lattice in $\mathbb{R}^n$, that is $\mathrm{Vol}\ \mathbb{R}^n/\mathcal{L} = 1$ for the standard Haar (Lebesgue) measure on $\mathbb{R}^n$. Then there exists a linear transformation g of $\mathbb{R}^n$ with $\det g = 1$ which maps $\mathcal{L}$ onto $\mathbb{Z}^n$.

Proof. As we have mentioned earlier, $\mathcal{L}$ is integrally spanned by some vectors $x_1,...,x_n$ in $\mathbb{R}^n$ (one may use any $x_1,...,x_n$ which generate $\mathcal{L}$ as a group). Then we use a transformation g which send $x_1,...,x_n$ to the standard basis of $\mathbb{R}^n$ (which integrally generates $\mathbb{Z}^n$).

Now we see that the action of $SL_n\mathbb{R}$ on the space $\{\mathcal{L}\}$ of the unimodular lattices in $\mathbb{R}^n$ is transitive and according to (a) the isotropy group at $\mathcal{L} = \mathbb{Z}^n$ equals $SL_n\mathbb{Z}$. Thus

$$\{\mathcal{L}\} = SL_n\mathbb{R}/SL_n\mathbb{Z},$$

as we have stated.

Now, to show that $SL_n\mathbb{R}/SL_n\mathbb{Z}$ is non-compact, we take a sequence of unimodular lattices $\mathcal{L}_i$, $i = 1,2,...$, where $\mathcal{L}_i$ is generated by vectors $x_1^i, x_2^i,...x_n^i$, such that $\|x_1^i\| \to 0$ as $i \to \infty$. Clearly, such a sequence diverges in $SL_n\mathbb{R}/SL_n\mathbb{Z}$ with the quotient space topology.

Remark. The effect of the condition $\|x_1^i\| \to 0$ can best be seen in the geometry of the torus $T_i^n = \mathbb{R}^n/\mathcal{L}_i$. Namely, the geodesic loop in T_i^n covered by the segment $[0, x_1^i] \subset \mathbb{R}^n$ has length $= \|x_1^i\| \to 0$ and so $\mathrm{Inj\ Rad}\ T_i^n \to 0$ for $i \to \infty$. On the other hand, the convergence $\mathcal{L}_i \to \mathcal{L}$ would imply a convergence (see below) of T_i^n to some flat n-dimensional torus T^n. Such a torus, of course, has $\mathrm{Inj\ Rad}\ T^n = \rho > 0$ which contradicts to the convergence $\mathrm{Inj\ Rad}\ T_i^n \to 0$ as Inj Rad is a *continuous* function on the space of flat tori, with the following topology corresponding to that in $SL_n\mathbb{R}/SL_n\mathbb{Z}$: a sequence T_i^n converges to T^n if and only if there exist linear diffeomorphisms $f_i : T_i^n \to T^n$, such that the induced metrics $f_i^*(g)$ converge to g_i, where g and g_i are the flat Riemannian metrics on T^n and T_i^n coming from the Euclidean metric on $\mathbb{R}^n$.

1.4.B. Arithmetic lattices. Starting from $SL_n\mathbb{Z}$ in $SL_n\mathbb{R}$ on may construct further (arithmetic) lattices in Lie groups. The construction goes in three steps.

(1) A *lattice* $\Gamma \subset G$ is called 1-*arithmetic* if there exists a continuous homomorphism h_1 of G into $GL_n\mathbb{R}$ for some n, such that $\Gamma = h_1^{-1}(SL_n\mathbb{Z})$.

(2) A lattice $\Gamma \subset G$ is called 2-*arithmetic* if there exists a 1-arithmetic lattice Γ_1 in some Lie group G_1 for which there exists a continuous homomorphism $h_2 : G_1 \to G$ with *compact* kernel, such that $h_2(\Gamma_1) = \Gamma$.

(3) A lattice $\Gamma \subset G$ is called *arithmetic* if it is *commensurable* with a 2-arithmetic lattice $\Gamma_2 \subset G$, where the commensurability means that the intersection $\Gamma \cap \Gamma_2$ has a finite index in Γ as well as in Γ_2.

Remarks. (1) Our definition of arithmeticity is slightly different from the usual one (see $[\mathrm{Bor}]_1$ [Rag]) but it is equivalent to the standard definition for semisimple groups G.

(2) If we want to use this construction to produce an (arithmetic) lattice in G we may first enlarge G by (essentially) multiplying it by a compact group K and then we need a homomorphism h_1 of the enlarged group $G_1 = G \times K$ into some $GL_n\mathbb{R}$, such that h_1^{-1} is a *lattice* in G_1. Constructing such G_1 and then h_1 is by no means trivial. However, A. Borel (see $[\mathrm{Bor}]_2$) has proven by this method that every non-compact simple Lie group G contains some arithmetic cocompact lattice and also some non-cocompact lattice.

Example. Let

$$G = O(\varphi_o) = O(p,q) \subset GL_n \ , \ n = p+q \ ,$$

be the orthogonal group consisting of the linear transformations of $\mathbb{R}^k$ fixing the form

$$\varphi_o = \sum_{i=1}^{p} x_i^2 - \sum_{i=p+1}^{n} x_i^2 \ .$$

Then we take another form, say φ_1 of the same signature as φ_o say

$$\varphi_1 = \sum_{i=1}^{p} a_i x_i^2 - \sum_{i=p+1}^{n} b_i x_i^2 \ ,$$

for $a_i > 0$ and $b_i > 0$, and denote by $O(\varphi_1)$ the orthogonal group corresponding to φ_1. The two forms are linearly equivalent and the linear transformation $g \in GL_n$ which moves φ_o to φ_1 also moves $O(\varphi_o)$ to $O(\varphi_1)$, by

$$gO(\varphi_o)g^{-1} = O(\varphi_1) .$$

Thus we obtain with φ_1 a new embedding to GL_n with the image $O(\varphi_1)$, say $h_1 : O(p,q) \to GL_n$. Now we ask ourselves when $h_1^{-1}(SL_n\mathbb{Z})$ (which is the same as $O(\varphi_1) \cap (SL_n\mathbb{Z})$) is a lattice in $O(p,q)$. The answer is provided by the following

<u>Theorem</u>. *Let* $n \geq 3$ *and both numbers* p *and* q *be positive.* (If one of them is zero the group $O(\varphi)$ is compact and every $\Gamma \subset O(\varphi)$ is a cocompact lattice). *Then* $\Gamma = SL_n\mathbb{Z} \cap O(\varphi_1)$ *is a lattice in* $O(\varphi_1)$ *if and only if* φ_1 *is proportional to a rational form,* $\varphi_1 = \alpha\varphi'_1$ for

$$\varphi'_1 = \sum_{i=1}^{p} a'_i x_i^2 - \sum_{i=p+1}^{n} b'_i x_i^2$$

where a'_i *and* b'_i *are rational numbers. Furthermore,* Γ *is cocompact if the only rational solution to the equation*

$$\varphi'_1(x_1,x_2 \ldots x_n) = 0$$

is

$$x_1 = x_2 = \ldots = x_n = 0 .$$

The proof follows from the general theory of arithmetic groups and can be found in $[Bor]_2$ and [G-P].

<u>Remarks</u>. (a) The theorem implies, in particular, that the group $\Gamma = SL_n\mathbb{Z} \cap O(\varphi_1)$ is *infinite* for all rational indefinite forms φ_1.

(b) The theorem remains valid for $n = 2$ unless the form φ'_1 splits into the product of two rational linear forms. For example Γ is a (cocompact) lattice for $\varphi = x_1^2 - bx_2^2$, where b is a positive rational number which is *not* a square of a rational number. Now, look at the orbit $\gamma(e_1) \subset \mathbb{R}^2$ for $e_1 = (1,0) \in \mathbb{R}^2$ and all $\gamma \in \Gamma$. As Γ is infinite this orbit is also infinite. On the other hand, every $\gamma(e_1)$ has *integer* coordinates x_1 and x_2 such that

$$x_1^2 - bx_2^2 = \varphi(\gamma(e_1)) = \varphi(e_1) = 1 .$$

Thus we recapture the classical result on the infinity of integral solutions to the Pell equation $x_1^2 - bx_2^2 = 1$.

(c) If $n \geq 5$, then the equation $\varphi'_1(x_1,...,x_n) = 0$ always has a non-trivial rational solution (see [B-S]) and so our $\Gamma \subset O(\varphi)$ is not cocompact. Yet one can obtain cocompact lattices by using quadratic forms over *finite extensions* of $\mathbb{Q}$ and invoking the second step of the definition (see 2.9.C_2, $[\text{Bor}]_1$, [G-P]).

1.5. Discrete groups and locally homogeneous spaces. Let the Lie group G in question equal the isometry group of a simply connected homogeneous Riemannian manifold. Then the structure of every discrete subgroup $\Gamma \subset G = \text{Iso } X$ is adequately reflected in the geometry of the quotient space $V = X/\Gamma$. For example, if Γ has *no torsion* and thus the action of Γ on X is free, then X equals the *universal covering* of V and Γ appears as the deck transformation (Galois) group of the covering $X \to V$. (In fact this remains true in the torsion case if V is given the quotient *orbifold* structure).

1.5.A. *Basic Example*. Let X be a symmetric space of negative Ricci curvature. Then X necessarily is simply connected and has non-positive sectional curvature (see [B-G-S]). Thus X is homeomorphic to a Euclidean space. Furthermore, the isometry group $G = \text{Iso } X$ is semisimple and $X = G/K$ for the maximal compact subgroup $K \subset G$. Now, one sees that the compactness of K implies that

$$\text{compactness } G/\Gamma \Leftrightarrow \text{compactness } X/\Gamma$$

$$\text{Vol } X/\Gamma < \infty \Leftrightarrow \text{Vol } G/\Gamma < \infty .$$

Hence, the study of torsion free cocompact discrete subgroups $\Gamma \subset G$ reduces to the geometry of compact locally symmetric spaces V locally isometric to X. Similarly, lattices $\Gamma \subset G$ correspond to complete manifolds V of finite volume.

Remarks. (a) The torsion free condition is not very restrictive for the following two reasons. First, every lattice Γ in our G contains a torsion free subgroup $\Gamma' \subset \Gamma$ of finite index (see [Rag]) and many questions concerning Γ can be easily readdressed to Γ'. However, the existence of Γ' is a highly non-trivial matter and, in fact, one can easily do without it by slightly enlarging the category. Namely, one should allow the *singular* spaces $V = X/\Gamma$, where Γ may have torsion. These can be treated as *orbifolds* (see [Th]) or as *metric* spaces, where the Riemannian metric X defines a curve length and consequently a metric in V. Notice that V contains a well defined nonsingular locus $U \subset V$ which is open and dense in V and which is characterized by the following property. The lift Y of U to X (for the quotient map $X \to V$) is the maximal subset in X on which Γ acts *freely*. Observe that this U is a non-complete

Riemannian manifold which is locally isometric to X and the metric completion of U equals V .

(b) Not every semisimple group G appears as Iso X for some symmetric space X . For example, the universal covering of $SL_2\mathbb{R}$ is not like that. However, every G is *locally isomorphic* to G' = Iso X for some X and the local isomorphism (i.e. the isomorphism between the universal coverings) is as good for our problems as an actual isomorphism.

Summarizing (a) and (b) we come to the conclusion that one can see much of discrete groups Γ in semisimple Lie groups G by looking at complete locally symmetric spaces V . Conversely, the geometry of these V reduces to that of discrete groups Γ isometrically acting on the universal coverings X of V .

Although the two view points are essentially equivalent they bring along quite different geometric images and techniques. When we look at a lattice Γ in a non-compact Lie group G we see a periodic set of stars in an infinite sky. Mathematically one thinks of Γ as a kind of a discrete approximation to G and (or) regards G as a kind of a continuous envelope (or hull) of Γ . But nothing of this is seen in $V = X/\Gamma$. This is just a complete manifold of finite volume (which is compact if $\Gamma \subset G$ is cocompact). The study of such V can be naturally conducted in the general framework of Riemannian geometry and analysis on (compact) manifolds.

§2. An overview of the rigidity problems.

The word "rigidity" applies to (discrete) subgroups $\Gamma \subset G$ as well as to locally homogeneous manifolds (and orbifolds) V. This expresses the idea that Γ is "unmovable" in G and in terms of V that the topology of V determines the (locally homogeneous) geometry. One usually distinguishes three different kinds of rigidity.

I. *Ordinary* or *local rigidity* which is sometimes called just *rigidity*. (See 2.1.).

II. *Strong (Mostow) rigidity*. (See 2.2.A.)

III. *Superrigidity* discovered by Margulis. This has two aspects, Archimedean and non-Archimedean (see 2.9. and 2.10.).

2.1. Rigidity of homomorphisms. Let Γ be an abstract countable group and G be a Lie group. A homomorphism $\rho : \Gamma \to G$ is called (locally) *rigid* if every deformation of ρ is induced by automorphisms of the ambient group G. Let us explain it in more details. First, a *one parameter deformation* ρ_t, $t \in \mathbb{R}$, is a family of homomorphisms such that $\rho_o = \rho$ and $\rho_t(\gamma) \in G$ is continuous in t for every $\gamma \in \Gamma$.

There is an obvious way to deform any $\rho = \rho_o$. Just take a path in G, that is $g_t \in G$, $t \in \mathbb{R}$, with $g_o = \mathrm{id}$ and compose ρ_o with the conjugations by g_t,

$$\rho_t = g_t \rho \ g_t^{-1} \quad . \qquad (*)$$

More generally, one can take a one parameter family of automorphisms a_t of G with $a_o = \mathrm{Id}$ and then deform ρ by

$$\rho_t = a_t \circ \rho \, . \qquad (**)$$

The deformations of this kind (**) have little to do with the specific of Γ and ρ and they are called *trivial*. (Sometimes one reserves the word "trivial" to inner automorphisms as in (*)). Now, a homomorphism ρ is called *rigid* if every deformation of ρ is trivial.

2.1.A. The space of homomorphisms. Denote by $\mathfrak{R}$ the space of homomorphisms $\rho : \Gamma \to G$ and then take the quotient space $\bar{\mathfrak{R}} = \mathfrak{R}/\mathrm{Aut}\, G$. Then the rigidity of ρ can be expressed by saying that $\bar{\rho} \in \bar{\mathfrak{R}}$ (i.e. the image of ρ under the tautological map $\mathfrak{R} \to \bar{\mathfrak{R}}$) is an *isolated* point of $\bar{\mathfrak{R}}$. To be rigorous here one should be careful with the topology in $\bar{\mathfrak{R}}$. The subtlety comes from possible non-compactness of the group $\mathrm{Aut}\, G$ which can make the space $\bar{\mathfrak{R}}$ non-

Hausdorff. Yet the difficulty is not serious and one can equate with little foundational work the following three properties.

1. ρ is rigid.

2. $\bar{\rho} \in \bar{\mathfrak{R}}$ is an isolated point.

3. The dimension of $\bar{\mathfrak{R}}$ at $\bar{\rho}$ equals zero. (This definition suggests $\dim_{\bar{\rho}} \bar{\mathfrak{R}}$ for a measure of non-rigidity of $\bar{\rho}$ in the case this dimension is positive).

2.2. Rigidity of subgroups. A subgroup $\Gamma \subset G$ is called *rigid* if the inclusion $\Gamma \subset G$ is a rigid homomorphism in the above sense.

2.2.A. Strong rigidity of Mostow. A lattice Γ in G is called (strongly) *Mostow rigid* if for another lattice in an arbitrary (connected as usual) Lie group, say $\Gamma' \subset G'$, every isomorphism between the lattices, $\Gamma \leftrightarrow \Gamma'$ extends to a unique isomorphism of the ambient Lie groups, $G \leftrightarrow G'$.

An essentially equivalent rigidity property for locally homogeneous spaces V of finite volume says that every V' *with finite volume* which is homotopy equivalent to V is isometric to V.

2.2.A'. If $\Gamma \subset G$ is a *cocompact* lattice then the Mostow rigidity is stronger than the ordinary (local) rigidity. In fact, a small deformation $\Gamma_\varepsilon \subset G$ of Γ is again discrete and cocompact by a simple argument (an exercise to the reader). Then the strong rigidity provides automorphisms a_ε of G_ε sending $\Gamma = \Gamma_0$ to Γ_ε which are (by another simple argument) continuous in ε. But, amazingly, there exist non-cocompact Mostow rigid lattices which are not locally rigid (see [Th], $[Gr]_1$).

2.3. Rigidity of G-actions. Let a Lie group G smoothly act on a manifold W and let $A : G \times W \to W$ denote this action. We say that A is *rigid* if for every smooth one-parameter family of actions A_t of G on W there exists a one-parameter family of diffeomorphisms $h_t : W \to W$ which conjugate A to A_t for all t in a small interval $(-\varepsilon,\varepsilon) \subset \mathbb{R} \to t$.

2.3.A'. Example. Let Γ be a cocompact lattice in G and consider the (compact) G-homogeneous space $W = G/\Gamma$. The topology of this space does not change if we slightly deform Γ (this must be obvious to the readers who solved the exercise indicated in 2.2.A'). Thus deformations Γ_ε of Γ naturally correspond to deformations A_ε of the original action A of G on $W = G/\Gamma$. Next, one can easily see that a diffeomorphism h_ε of W changing the action A to A_ε is essentially the same thing as a conjugation of Γ to Γ_ε in G. Thus the rigidity of the action implies the rigidity of Γ with the additional property of the implied automorphism $a_\varepsilon : G \to G$ being *inner*. (Making a detailed proof amounts to unraveling the pertinent definitions concerning groups, homogeneous spaces etc.).

2.4. Rigidity of finite groups. If Γ is a *finite* group, then every homomorphism $\Gamma \to G$ is locally rigid. This is a classical fact and it can be seen from various angles. Here are some of them.

2.4.A. If G is a subgroup in GL_n, then homomorphisms of Γ to G induce those to GL_n that are *linear representations* of Γ. One knows since the last century that n-dimensional representations of finite (or compact) group form a finite set modulo conjugation in GL_n. This implies the rigidity in GL_n and with some extra work one can take care of all $G \subset GL_n$.

2.4.B. Consider two homomorphisms of Γ to G, denoted $\gamma \mapsto \bar{\gamma}$ and $\gamma \mapsto \hat{\gamma}$, and let us try to find a conjugation of $\bar{\Gamma}$ to $\hat{\Gamma}$. That is, we are after an element $g \in G$ satisfying $g\bar{\gamma}g^{-1} = \hat{\gamma}$, for all $\gamma \in \Gamma$, or, equivalently

$$g = \hat{\gamma} g \bar{\gamma}^{-1} \qquad (*)$$

The equation (*) should be thought of as the fixed point condition for the action

$$g \mapsto \hat{\gamma} g \bar{\gamma}^{-1} \qquad (**)$$

of Γ on G.

If $\bar{\Gamma} = \hat{\Gamma}$ then the obvious fixed point is $g = \mathrm{Id} \in G$. Now, if $\bar{\gamma}$ is close to $\hat{\gamma}$ for all $\gamma \in \Gamma$ we can invoke the following classical

2.4.B'. Stability theorem. *Let a finite (or compact) group Γ act on a smooth manifold W and $w_0 \in W$ be a fixed point of this action. Then w_0 is stable under small smooth perturbation of the action say A_ε of A, in the following sense. Every action A_ε close to $A = A_0$ has a fixed point $w_\varepsilon \in W$ which is continuous in ε for small ε.*

Idea of the proof. If the space W and the actions A_ε are linear, then the fixed point w_ε can be obtained by the averaging of the orbit $\gamma_\varepsilon(w_0)$ over Γ, where γ_ε denotes the action of Γ according to A_ε. (This is what is used in the representation theory mentioned above). Then the general nonlinear case follows by using a linearization of the action and a suitable implicit function theorem. Alternatively, one may use an invariant Riemannian metric on W (which can also be obtained by averaging an arbitrary metric) and then take the center of mass of the orbit $\gamma_\varepsilon(w_0)$ for such a metric (see [Kar] for a center of mass construction).

2.4.C. *If* G *is a compact group then every smooth action of* G *on a compact manifold* W *is rigid*. In fact, the diffeomorphisms h moving one action to another, say $\bar{g}$ to $\hat{g}$, are the fixed points of the action $h \mapsto \hat{g} \circ h \circ \bar{g}^{-1}$ of G on the space of maps $h : W \to W$. The existence of a fixed point in this (infinite dimensional) space can be obtained by the argument indicated in 1.4.B'., which implies the desired rigidity according to Mostow, Palais etc. (see $[\text{Mos}]_2$, [Pal]). Notice that this rigidity implies that for discrete (and hence finite) subgroups $\Gamma \subset G$.

2.4.D. <u>On the failure of the strong rigidity</u>. Two isomorphic finite subgroups Γ and Γ' in G do not have to be conjugate or be obtainable one from the other by an automorphism of G. A trivial example is that of two different representations of the cyclic group $\Gamma = \mathbb{Z}/p\mathbb{Z}$ into the unitary group $U(1) = S^1 \subset \mathbb{C}^\times$. Every such representation is determined by where a given generator $\gamma_0 \in \Gamma$ goes. The image of γ_0 is a p-th root of unity $\mathbb{C}^\times$ and by taking the representations $\gamma_0 \mapsto \alpha_0$ and $\gamma_0 \mapsto \alpha'_0$ for different roots of unity we obtain non-equivalent representations.

This example is not very convincing as the *images* of the representations are the same, provided the representations in question are injective (i.e. the roots α are primitive). Namely, the image consists of *all* p-th roots of unity in such a case. A more interesting situation arises for cyclic subgroups Γ in U(2). Such a subgroup can be generated by a diagonal matrix $\begin{pmatrix} \alpha_o & 0 \\ 0 & \beta_o \end{pmatrix}$ where α_0 and β_0 are p-th roots of unity. Now, for a fixed α_0 we can vary β_0 thus obtaining non-conjugate subgroups in U(2) (Two subgroups can be distinguished by the trace, $\gamma_0^i \mapsto \alpha_0^i + \beta_0^i$, $i = 1,2,\dots,p$, which is a function on $\Gamma = \{\gamma_0,\dots,\gamma_0^p\}$ invariant under conjugations in U(2)).

One can see everything more geometrically by looking on how Γ acts on the unit sphere $S^3 \subset \mathbb{C}^2$. If this action is free (which is the case for most representations $\Gamma \to U(2)$) one takes the quotient manifold S^3/Γ which is called a *lens space* and which carries a (locally homogeneous) metric of constant curvature $+1$. The above discussion shows that two different lens spaces do not have to be isometric. (We suggest to the reader to work out a specific example and describe, geometrically, an invariant distinguishing S^3/Γ and S^3/Γ' for Γ and Γ' isomorphic to $\mathbb{Z}/5\mathbb{Z}$).

2.4.D'. The geometry and topology of lens spaces goes much deeper than one might think. For example two such spaces can be homotopy equivalent but not diffeomorphic (see $[\text{Mil}]_1$).

2.5. <u>Rigidity of lattices in Abelian and nilpotent groups</u>. Every isomorphism between two lattices in $\mathbb{R}^n$ is induced by a unique automorphism of $\mathbb{R}^n$. In fact, continuous automorphisms of $\mathbb{R}^n$ are just linear automorphisms. As every lattice $\Gamma \subset \mathbb{R}^n$ linearly spans $\mathbb{R}^n$, the homomorphisms $h : \Gamma \to \mathbb{R}^m$ uniquely extend to $\mathbb{R}^n$ by linearity : each $x = \sum_i c_i\gamma_i \in \mathbb{R}^n$ goes to $h(x) \underset{\text{def}}{=} \sum_i c_i h(\gamma_i) \in \mathbb{R}^m$. Of course, one should check that $h(x)$ does not depend on how x is combined out of γ_i, i.e. the equality $\sum_i c_i\gamma_i = \sum_i c'_i\gamma_i$ must imply $\sum_i c_i h(\gamma_i) = \sum_i c'_i h(\gamma_i)$. This is indeed true and easy to prove.

2.5.A. Let us indicate a more geometric way of extension of h from Γ to $\mathbb{R}^n$. First we consider the group $\Gamma_{\mathbb{Q}} \subset \mathbb{R}^n$ of Γ-*rational* points, where $x \in \mathbb{R}^n$ is called Γ-*rational* if $nx \in \Gamma$ for some $x \in \mathbb{Z}$. In other words

$$\Gamma_{\mathbb{Q}} = \bigcap_{n \in \mathbb{Z}} n^{-1}\Gamma ,$$

where

$$\alpha\Gamma \underset{\text{def}}{=} \{\alpha\gamma \mid \gamma \in \Gamma\} .$$

Now, every h extends from Γ to $\Gamma_{\mathbb{Q}}$ by $h(n^{-1}\gamma) = n^{-1} h(\gamma)$ for all $\gamma \in \Gamma$ and $n \in \mathbb{Z}$. Here again one must check that

$$n_1^{-1}\gamma_1 = n_2^{-1}\gamma_2 \Rightarrow n_1^{-1} h(\gamma_1) = n_2^{-1} h(\gamma_2) ,$$

but this is quite easy. Moreover, the extension

$$h_{\mathbb{Q}} : \Gamma_{\mathbb{Q}} \to \mathbb{R}^m$$

is a homomorphism which (by an extra easy argument) is continuous for the induced topology of $\Gamma_{\mathbb{Q}} \subset \mathbb{R}^n$. Thus $h_{\mathbb{Q}}$ extends by continuity to the desired homomorphism, say

$$h_{\mathbb{R}} : \mathbb{R}^n \to \mathbb{R}^m.$$

Exercise. Show that the existence of an extension of h from Γ to $\mathbb{R}^n$ remains valid for non-cocompact discrete subgroups Γ but then such an extension is non-unique.

2.5.B. The rigidity of Γ can be expressed in terms of the flat torus $T^n = \mathbb{R}^n/\Gamma$ as follows : every automorphism of $\Gamma = \pi_1(T^n)$ is induced by an affine automorphism of T^n that is a diffeomorphism preserving the Levi Civita connection of the flat metric on T^n. Furthermore, this automorphism is unique up to translations of T^n.

Remarks. (a) We can not replace the affine automorphisms by isometries. In fact as we deform a lattice $\Gamma \subset \mathbb{R}^n$ the isometry type of T^n may change.

(b) One can slightly modify and generalize the above as follows :

Every continuous map between flat tori, say $T^n \to T^m$ *is homotopic to an affine map and this affine map is unique up to translations of* T^n *on* T^m. (Here, affine maps don't have to be diffeomorphic, but just continuous sending geodesics to geodesics).

The proof of this mapping theorem is quite easy and we suggest the reader to work it out. A more difficult exercise to the reader is to prove a similar result for maps between arbitrarily compact flat Riemannnian manifolds.

We shall see later on in §4 that a similar mapping theorem holds true for some (non flat!) locally symmetric spaces of non-compact type but the required proof is significantly deeper.

2.5.C. Nilpotent groups. A group G is called nilpotent of (nilpotency) degree k if for arbitrary $k+1$ elements $g_o, g_1, \dots, g_k$ the iterated commutator of g_i equals id,

$$[g_o,[g_1,\dots,[g_{k-1}, g_k]]] = \mathrm{id} .$$

Thus "nilpotent of degree one" means "abelian" as

$$[g_o,g_1] \underset{\mathrm{def}}{=} g_o g_1 g_o^{-1} g_1^{-1} = \mathrm{id} ,$$

for all g_o and g_1 in G. The nilpotency of degree 2 requires

$$[g_o,[g_1, g_2]] = \mathrm{id}$$

for all g_o, g_1 and g_2 in G and so on. Finally, just "nilpotent" means "nilpotent of some degree k ".

2.5.C'. Example. The group Δ^{k+1} of upper triangular $(k+1) \times (k+1)$-matrices with unit diagonal entries is nilpotent of order k. This is seen immediately. The first interesting example is the Heisenberg (abc)-group $\Delta^3 = \left\{\begin{pmatrix} 1 & a & c \\ 0 & 1 & b \\ 0 & 0 & 1 \end{pmatrix}\right\}$. Here the matrices $\underline{\mathbf{a}} = \begin{pmatrix} 1 & a & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$, $\underline{\mathbf{b}} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & b \\ 0 & 0 & 1 \end{pmatrix}$ and $\underline{\mathbf{c}} = \begin{pmatrix} 1 & 0 & c \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$ satisfy $[\underline{\mathbf{a}},\underline{\mathbf{c}}] = [\underline{\mathbf{b}},\underline{\mathbf{c}}] = 1$ and $[\underline{\mathbf{a}},\underline{\mathbf{b}}] = \begin{pmatrix} 1 & 0 & ab \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$.

2.5.C_1. Malcev uniqueness (rigidity) theorem. *let* G *and* G' *be simply connected nilpotent Lie groups and* $\Gamma \subset G$ *and* $\Gamma' \subset G'$ *be lattices. Then every isomorphism* $\Gamma \leftrightarrow \Gamma'$ *uniquely extends to an isomorphism* $G \leftrightarrow G'$.

The corresponding existence theorem is also due to Malcev. It reads

2.5.C_2. *For every finitely generated torsionless nilpotent group* Γ *there exists a simply connected Lie group* G *containing* Γ *as a lattice.*

The idea of the proof of 2.5.C_1 is the same as the one used in the abelian case. First one defines $\Gamma_{\mathbb{Q}} \subset G$ as the set of all $g \in G$ satisfying $g^n \in \Gamma$ for some $n \in \mathbb{Z}$, and one shows that $\Gamma_{\mathbb{Q}}$ is a dense subgroup in G. Then one extends an isomorphism from Γ to $\Gamma_{\mathbb{Q}}$ by a purely algebraic argument. The final extension from $\Gamma_{\mathbb{Q}}$ to G is obtained by continuity. Notice that, as in the abelian case, everything works here for an arbitrary homomorphism $h : \Gamma \to G'$, and allows a unique extension to a homomorphism between the Lie groups, say $h_{\mathbb{R}} : G \to G'$.

The proof of the Malcev existence theorem is similar to how one builds real numbers starting from integers. First one abstractly constructs the group $\Gamma_{\mathbb{Q}}$ by introducing the "roots" $\gamma^{1/n}$, for all $\gamma \in \Gamma$ and $n \in \mathbb{Z}$. The group $\Gamma_{\mathbb{Q}}$ is then given an appropriate topology and $G = \Gamma_{\mathbb{R}} \supset \Gamma_{\mathbb{Q}}$ is obtained by completing $\Gamma_{\mathbb{Q}}$.

The actual proof of the Malcev theorem is somewhat involved and we do not explain it here (see [Mal]). But we want to indicate a somewhat different geometric construction of a Lie group starting from a discrete nilpotent group Γ. The rough idea is to view Γ as a discrete set looked at from a certain point v "outside" Γ. As the view point v goes further and further away from Γ the "visual gaps" between points in Γ become smaller and smaller. Then, in the limit for $v \to \infty$, we shall see no gaps at all but rather a continuous object, which, in fact, can be given a structure of a nilpotent Lie group. In more technical terms this limit group is obtained with some auxiliary (word) metric ρ on Γ by taking

$$\Gamma^* = \lim_{n \to \infty} (\Gamma, n^{-1}\rho),$$

where one uses the (pointed) *Hausdorff topology* to go to the limit. (See [G-L-P], $[\mathrm{Gr}]_2$). The above limit group Γ^* is not the same as $G = \Gamma_{\mathbb{R}}$ obtained by completing $\Gamma_{\mathbb{Q}}$, though both are Lie groups of the same dimension and nilpotency degree. Yet, Γ^* is not always isomorphic to G (see $[\mathrm{Pan}]_1$ for an extensive discussion).

Examples. (a) If Γ is a free abelian group of rank n , then, of course, $\Gamma^* = \mathbb{R}^n$. Yet even in this case there is no canonical embedding of Γ to Γ^*.

(b) Let $\Gamma \subset \Delta^{k+1}$ consist of the triangular matrices with integer entries. Then $\Gamma_{\mathbb{Q}}$ consists of the matrices with rational entries (this is easy to show) and the Malcev theorem becomes especially transparent. Also in this case Γ^* is isomorphic to $\Gamma_{\mathbb{R}} = \Delta^{k+1}$.

Remarks . (a) The above discussion suggests the existence of certain universal constructions (functors) leading from discrete to continuous groups, but no direct generalization of $\Gamma \to \Gamma_{\mathbb{R}}$ and $\Gamma \to \Gamma^*$ seems possible for non (virtually) nilpotent groups, such as lattices in semisimple groups. Yet there are other constructions available as we shall see in the course of these lectures.

(b) The geometry of lattices in simply connected solvable Lie groups is similar to that for nilpotent groups although the final answers are less functorial. An interested reader may consult $[\mathrm{Mos}]_3$, [Rag].

2.6. Representations of finitely presented groups. Let Γ be given by a finite presentation $\{\gamma_1,\dots,\gamma_k \mid r_1,\dots,r_\ell\}$, where $\gamma_1,\dots,\gamma_k$ are generators of Γ and $r_1,\dots,r_\ell$ are relations. Recall that every relation has the form

$$\gamma_{i_1}^{\varepsilon_1}\gamma_{i_2}^{\varepsilon_2}\dots\gamma_{i_n}^{\varepsilon_n} = 1 ,$$

where ε_i are arbitrary integers, where the values of the indices i_j run through $1,\dots k$ possibly with repetitions (so that n can be any number) and where 1 denotes the neutral element. Then every homomorphism $\Gamma \to GL_n$, denoted $\gamma \mapsto \bar\gamma$, is determined by k matrices $\bar\gamma_1,\dots\bar\gamma_k$ in GL_n which satisfy the relations r_i with the matrices $\bar\gamma_i$ in place of γ_i . Notice that a single matrix relation amounts to n^2 algebraic equations imposed on the entries of the matrices $\bar\gamma_i$. As every $\bar\gamma_i$ has n^2 entries, the set of all representations $\Gamma \to GL_n$ appears as an algebraic subvariety $\mathcal{R}$ in the kn^2-dimensional Euclidean space given by ℓn^2 equation. Thus the expected dimension of $\mathcal{R}$ is $(k-\ell)n^2$. Recall that the real object of interest is not $\mathcal{R}$ but $\mathcal{M} = \mathcal{R}/GL_n$ where GL_n acts on $\mathcal{R}$ by conjugation. For $k \ge 2$ the action is free almost everywhere on the kn^2-dimensional space containing $\mathcal{R}$. Thus the expected dimension of $\mathcal{M}$ is $(k-\ell-1)n^2$.

If we work over *complex* numbers and our GL_n is, in fact, $GL_n\mathbb{C}$ then this heuristic computation does give rigorous results. Namely, if $k-\ell \ge 2$, then the dimension of $\mathcal{M}$ is at least n^2 at every point (by elementary intersection theory, see [Sha]) . In particular, *no complex representation of* Γ *is rigid.*

Notice that the above argument remains valid for all complex Lie groups G in place of $GL_n\mathbb{C}$ and the conclusion often holds true for real groups as well. Thus we see that the rigidity of $\Gamma \subset G$ implies that the number ℓ of relations of Γ is almost as large as the number k of generators of Γ, namely

$$\ell \geq k-1 .$$

2.6.A. Example. $G = SL_2\mathbb{R}$. This group has dimension 3 and so the expected dimension of $\mathfrak{M}$ is $3(k-\ell-1)$. In particular, if Γ is the fundamental group of a surface of genus g , then $k = 2g$ and $\ell = 1$, and the expected dimension of $\mathfrak{M}$ is $6g-6$. Now, a *discrete cocompact* surface group Γ in $SL_2\mathbb{R}$ gives us a surface V of constant negative curvature, that is

$$V = SO(2)\setminus SL_2\,\mathbb{R}/\Gamma .$$

The space of such surfaces of genus $g \geq 2$ is known (by elementary Teichmüller theory) to have dimension 6g-6 . Thus our heuristic computation gave us the right answer for $g \geq 2$. Notice that this answer implies non-rigidity of (most) lattices in $SL_2\mathbb{R}$.

2.6.B. Example : $G = SL_2\mathbb{C}$. For this group the corresponding symmetric space $X = SU(2) \setminus SL_2\mathbb{C}$ is the 3-dimensional hyperbolic space. Thus lattices Γ in $SL_2\mathbb{C}$ correspond to 3-dimensional manifolds (*orbifolds*, if Γ has torsion) V of constant negative curvature and finite volume. If V is compact then the corresponding Γ (cocompact in this case) admits a presentation with the number ℓ of relations equal to k , the number of generators. In fact, this is true for the fundamental group Γ of an arbitrary closed 3-manifold. To see this take a Morse function on V with a single maximum and single minimum. Then the number k of critical points of index 1 equals the number ℓ of the points of index 2, since

$$1 + k - \ell - 1 = \chi(V) = 0 .$$

It follows that the associated Morse complex has k one-cells and ℓ two-cells, which gives us the desired presentation of $\Gamma = \pi_1(V)$. Now, our heuristic computation gives

$$\dim_{\mathbb{C}} \mathfrak{M} = 3(k-\ell-1) = -3$$

which suggests that $\mathfrak{M}$ does not exist at all (or reduces to a single point corresponding to the trivial representation of Γ).

To see what really happens we consider a non-cocompact lattice Γ without torsion and look at the corresponding non-compact manifold V . One knows (see [Th], $[Gr]_1$) that this V is diffeomorphic to a compact manifold, say $\bar{V}$, whose boundary is the union of several, say m , 2-tori. The Euler characteristic of this V is (by an obvious argument)

$$\chi(V) = -m ,$$

and the above discussion gives a presentation of $\Gamma = \pi_1(V)$ with k generators and ℓ relation, such that $k = \ell + m$. (By attaching m solid tori to the boundary one adds m relations to Γ. But the resulting manifold V' is compact and so its fundamental group Γ' with $k' = k$ and $\ell' = \ell + m$ has $k' = \ell'$. This gives an explanation to the relation $k = \ell + m$). Now our heuristic argument predicts that

$$\dim_{\mathbb{C}} \mathcal{M} = m - 3 ,$$

but the correct answer is

$$\dim_{\mathbb{C}} \mathcal{M} = m ,$$

where the dimension is measured at the point in $\mathcal{M}$ corresponding to our Γ (see [Th]). In particular, cocompact lattices are rigid while non-cocompact ones are not. (However, all lattices in $SL_n\mathbb{C}$ are Mostow rigid, see §3).

Remark. The (rigidity) equality $\dim_{\mathbb{C}} \mathcal{M} = 0$ for cocompact lattices in $SL_2\mathbb{C}$ should be thought of as a kind of uniqueness theorem (compare Malcev theorems above). The corresponding existence result (which is by far deeper than the rigidity and is due to Thurston) claims that a compact 3-manifold satisfying certain topological conditions admit metrics of constant negative curvature (see [Th]). Notice that the topological conditions needed for Thurston's existence proof do not seem the best possible. The best possible result (may be, too good to be true) would be as follows.

Conjecture. A closed 3-manifold V admits a metric of constant negative curvature if and only if $\pi_2(V) = \pi_3(V) = 0$ and every solvable subgroup in $\pi_1(V)$ is cyclic.

2.6.C. Rigidity of lattices and the rigidity of polyhedra. There is formal similarity between the deformation problem for discrete subgroups and that for polyhedra in the spaces X of constant curvature. For example, let P be a convex polyhedron in $\mathbb{R}^3$ and let us try to deform P without changing the geometry of its 2-faces. Of course, there are *trivial* deformations corresponding to the rigid motions of P (that come from the group $\mathrm{Iso}\,\mathbb{R}^3$), but we are after the real deformations which change the dihedral angles between the faces. In fact, every such deformation is given by the numbers attached to the edges of P and measuring (the change of) these angles. Denote by e the number of the edges and observe that every vertex of P imposes 3 relations on these angles. For example, if there are exactly 3 faces adjacent to some vertex of P, then all three dihedral angles are uniquely determined by the planar angles of the faces themselves at this point. If there are 4 faces, then there is one degree of freedom left and so on.

By the Euler characteristic formula the numbers of the vertices, faces and edges satisfy

$$v - e + f = 2 .$$

Furthermore, if all faces are triangular, then, obviously, $f = \frac{2}{3} e$ and so $e - 3v = -6$, which *suggests* there is no non-trivial deformations of P . In fact, the above computation "overcounts" the number of relations (as also happens for our heuristic count of the dimension of the deformation spaces in $SL_2\mathbb{C}$) . To make the count more precise, we fix a 2-face F_0 of P and observe that the angles at the three edges of F_0 are uniquely determined by the remaining dihedral angles. Thus we reduce the situation to $k = e - 3$ parameters and $\ell = 3v-9$ relations (as we do not need anymore the relations coming from the vertices of F_0). Now we have a perfect match

$$k = \ell ,$$

which suggests both the rigidity (uniqueness) as well as the existence of polyhedra with given geometry of the faces. In fact, the rigidity is the classical theorem of Cauchy and Dehn and the existence is also valid by the work of A.D. Alexandrov (see [Ale]).

2.6.C_1. Reflection groups in H^2 . Let P be a convex n-gon, $n \geq 5$, in the hyperbolic plane H^2 all of whose angles are 90°. Then the k reflexions $\gamma_1,...,\gamma_k$ of H^2 with respect to the edges of P (or rather the straight lines ℓ_i extending the edges) generate a discrete group Γ with the fundamental domain P . (Notice that P is a distinguished fundamental domain as it is bounded by the lines ℓ_i = Fixed-point-set-of γ_i) . Now we want to deform Γ by deforming P keeping all angles 90°. The geometry of such a P is determined by the length of the edges, that are n real numbers, which are subject to three relations. For example, every triangle P in H^2 is uniquely (up to $\mathrm{Iso}\, H^2$) determined by its angles (of course these can not be all 90°). Thus the space $\mathfrak{M}$ for Γ has $\dim \mathfrak{M} = n - 3$.

2.6.C_2. Reflection groups in H^3 . Let us count the dimension of the space of deformations of a convex polyhedron $P \subset H^3$ which do not change the dihedral angles of P . We assume here that P is dual to a simplicial polyhedron, i.e. there are exactly 3 edges at every vertex of P . Then the deformation is controlled by the lengths of the edges, where the three edges at a fixed vertex V_0 are determined by the rest. Thus we have $k = e - 3$ parameters restricted by $\ell = 3(f - 3)$ relations, which gives us the identity $k = \ell$ (since $v-e+f = 2$ and $v = \frac{2}{3} e$) . This identity *suggests* both the rigidity (uniqueness) or P as well as the existence theorem, and a theorem by Andreev (see [And], [Th]) confirms this prediction under some mild combinatorial restrictions on P .

Now let P have all dihedral angles 90°. Then the group Γ generated by the reflections around the faces of P is discrete and cocompact (we have assumed all along P is compact) and the above-mentioned Andreev theorem insures the rigidity of Γ. Moreover, the existence part of Andreev's theorem provides a group Γ whose fundamental domain is combinatorially isomorphic to a given abstract polyhedron P satisfying certain necessary conditions.

Finally notice that the structure of the equations in the Andreev (Thurston) theory is very much similar to that in the Cauchy-Dehn-Alexandrov theorem and the proofs in both cases follow the same strings of ideas.

2.7. Local rigidity theorem for (semi)simple Lie groups. *Let Γ be a lattice in a non-compact simple Lie group G which is not locally isomorphic to $SL_2\mathbb{R}$ or to $SL_2\mathbb{C}$. Then Γ is locally rigid. Furthermore, if Γ is ocompact, then the rigidity remains valid for $SL_2\mathbb{C}$ and $PSL_2\mathbb{C}$. (These are the two groups locally isomorphic to $SL_2\mathbb{C}$).*

We do not prove this rigidity in our lectures.

2.7.A. Historical remarks and references. The first rigidity theorem, due to Selberg [1960], applies to $\Gamma = SL_n\mathbb{Z} \subset SL_n\mathbb{R}$ for $n \geq 3$. His proof is geometric in nature and is based on the study of maximal Abelian subgroups (isomorphic to $\mathbb{Z}^{n-1}$) in Γ (see [Sel]). The rigidity of cocompact lattices in O(n,1) is due to Calabi [1961] and the cocompact Hermitian case (where the corresponding symmetric space is Hermitian) is covered by a theorem of Calabi and Vesentini [1960] (see [Cal], [Cal-Ves]). The general cocompact case is due to A. Weil [1962] (see [Wang] and the end of Ch. VII in [Rag]).

The method of Calabi-Vesentini and Weil is based on Hodge theory and Bochner type inequalities.

2.7.B. The above rigidity theorem extends to semi-simple Lie groups without compact normal subgroups where one should take special care of the components locally isomorphic to $SL_2\mathbb{R}$ and $SL_2\mathbb{C}$.

2.7.C. If G is a non-compact simple Lie group of dimension > 6 (i.e. bigger than $SL_2\mathbb{R}$ and $SL_2\mathbb{C}$) then the lattices Γ in G appear rigid many times over. In fact by looking at the list of known algebraic properties of such Γ one could naturally conclude that no such Γ may exist at all.

Now, if the mere existence of Γ is a miracle, the existence of a deformation would be a miracle squared. This makes the rigidity theorem look rather shallow, and one expects here deeper "overrigidity" results.

2.8. Mostow rigidity theorem. *Let* G *be a non-compact simple Lie group of dimension* > 3 *(i.e. not locally isomorphic to* $SL_2\mathbb{R}$ *). Then every lattice* $\Gamma \subset G$ *is Mostow rigid.*

We shall prove this theorem for some G in §§ 3 and 4.

2.8.A. Let us formulate the Mostow theorem in terms of locally symmetric spaces.

Let V *and* V' *be complete locally symmetric spaces of non-compact type (i.e. with Ricci* < 0 *) of finite volume, such that* Vol V = Vol V' . *If* V *is irreducible of dimension* > 3 *, then every isomorphism between the fundamental groups,* $\alpha : \pi_1(V) \to \pi_1(V')$ *is induced by a unique isometry of* V *onto* V' .

2.8.B. Remarks and corollaries. (a) As we fix no reference points in V and V' the fundamental groups are not truly defined and α should be thought of as not as an actual isomorphism but as an isomorphism up to conjugations in $\pi_1(V)$ and $\pi_1(V')$.

(b) An immediate corollary to the Mostow theorem reads

The group of the exterior automorphisms of $\Gamma = \pi_1(V)$ *, that is* Aut Γ/Inn Aut Γ , *is canonically isomorphic to the isometry group of* V .

To see this we apply the theorem to the case V = V' .

Next we observe that the isometry group of V is *finite*. In fact this finiteness is a general (and well known) property which holds true for all complete Riemannian manifolds V of finite volume and negative Ricci curvature (Bochner theorem, see [B-Y]). Finally we conclude that the group Ext Aut Γ is finite as well.

(c) The equality Vol V = Vol V' is a normalization condition. We could equally use another condition, say

$$\text{Scal Curv } V = \text{Scal Curv } V'$$

(d) The Mostow theorem extends with little trouble to the *semi*-simple groups G (which correspond to reducible locally symmetric spaces V) if one takes proper care of $SL_2\mathbb{R}$-components of G .

2.8.C. References. The first instance of the Mostow rigidity was proven in 1967 for compact spaces of constant negative curvature (see $[\text{Mos}]_1$) . The general compact case appears in $[\text{Mos}]_4$ in 1973. The Mostow rigidity for certain non-compact spaces (of $\mathbb{Q}$-rank one) is due to Prasad [Pra] and the general superrigidity of Margulis (see $[\text{Mar}]_1$, $[\text{Mar}]_2$). The basic idea of Mostow consists, in the study of the so-called *ideal boundary* ∂X of a non-compact symmetric space X , but the study of this boundary is quite different for the cases of $\text{rank } X = 1$ and $\text{rank } X \geq 2$.

2.8.C'. Recall that $\text{rank } X$ is the maximal dimension of a *flat* in X that is a totally geodesic subspace isometric to a Euclidean space. Also notice that a symmetric space has rank one if and only if its sectional curvature is (strictly) negative, $K(X) < 0$.

According to the Cartan classification there are three series of non-compact rank 1 symmetric spaces and one exceptional space. They are

I. The real hyperbolic space H^n or $H^n_{\mathbb{R}}$. This is the only complete simply connected Riemannian manifold of constant sectional curvature.

II. The complex hyperbolic space $H^{2n}_{\mathbb{C}}$. This has an invariant Kähler metric of constant holomorphic curvature. One can also think of $H^{2n}_{\mathbb{C}}$ as the unit ball in $\mathbb{C}^n$ with the *Bergman* metric.

III. The quaternionic hyperbolic space $H^{4n}_{\mathbb{H}}$.

IV. The hyperbolic Cayley plane $H^{16}_{\mathbb{C}a}$.

We refer to $[\text{Mos}]_4$ and $[\text{Pan}]_2$ for a study of these spaces. Here we only mention the following important characteristic property.

The isometry group of every of the above hyperbolic spaces H *is transitive on the pairs of points* (x,y) *with a fixed distance* $\text{dist}(x,y) = d$. This is equivalent to saying that the isotropy group $K_x \subset \text{Iso } H$ of each point $x \in H$ (that is the maximal compact subgroup in Iso H) is transitive on the unit tangent sphere of $S_x(1) \subset T_x(H)$. For $H = H^n_{\mathbb{R}}$ this K_x is the orthogonal group $O(n)$ naturally acting on $\mathbb{R}^n = T_x(H)$; for $H^{2n}_{\mathbb{C}}$ this K is $U(n)$; for $H^{4n}_{\mathbb{H}}$ this is the (unitary) symplectic group $Sp(n)$ (of linear transformations of the quaternionic space $\mathbb{H}^n$) and for $H^{16}_{\mathbb{C}a}$ this is the famous group Spin 9 acting on the Cayley plane $\mathbb{C}a^2 = \mathbb{R}^{16}$.

2.9. Margulis' (super)rigidity. Consider a discrete subgroup $\Gamma \subset G$ and a homomorphism of Γ into another Lie group, say $h : \Gamma \to G'$. Margulis' property, says, roughly speaking, that h extends to a (continuous !) homomorphism $\bar{h} : G \to G'$ unless the homomorphism h was somewhat "exceptional".

Here is a precise statement for simple Lie groups G .

2.9.A. *Let* G *be a non-compact simple Lie group which is not locally isomorphic to* $O(n,1) = \text{Iso}\, H^n_{\mathbb{R}}$ *or to* $U(n,1) = \text{Iso}\, H^{2n}_{\mathbb{C}}$ *and let* Γ *be a lattice in* G *. Then every linear representation* h *of* Γ *, that is a homomorphism* $h : \Gamma \to GL(N)$ *extends to a continuous homomorphism* $\bar{h} : G \to GL(N)$ *unless the image* $h(\Gamma)$ *is precompact in* $GL(N)$.

This theorem for $\text{rank}_{\mathbb{R}} G \geq 2$ (where the underlying symmetric space has $\text{rank} \geq 2$) was proven by Margulis in 1974 (see $[\text{Mar}]_1$, $[\text{Mar}]_2$, [Zim], $[\text{Für}]_2$) using random walks on Γ and G and related ideas of ergodic theory. In the remaining cases, $G = Sp(n,1) = \text{Iso}\, H^{4n}_{\mathbb{C}}$ and $G = F_4 = \text{Iso}\, H^{16}_{\mathbb{C}a}$, this theorem was recently proven by K. Corlette (see $[\text{Cor}]_1$), who used the method of harmonic maps (see §4).

2.9.B. Let us indicate a special case of superrididity in terms of locally symmetric spaces. Let V and V' be locally symmetric spaces with fundamental groups Γ and Γ' and let $h : \Gamma \to \Gamma'$ be a homomorphism. We want to find a *geodesic* map $\hat{h} : V \to V'$ which induces h on Γ (where h is taken up to conjugation as in 2.8.B.). Recall, that a map $V \to V'$ between Riemannian manifolds is called geodesic if the graph of this map is a totally geodesic submanifold in $V \times V'$. For example, geodesic maps $\mathbb{R}^m \to \mathbb{R}^n$ are the same as affine maps.

Now we suppose $V = K \setminus G/\Gamma$ where G satisfies the assumptions of the Margulis-Corlette theorem and let us explain how to construct $\hat{h}$ for an arbitrary locally symmetric V' of negative Ricci curvature. We use the adjoint linear representation of the group $G' = \text{Iso}\, X'$ for the universal covering X' of V' and by composing this representation, say $G' \to GL(N)$ with $h : \Gamma \to \Gamma'$ and $\Gamma' \subset G = \text{Iso}\, X'$ we obtain a linear representation $\Gamma \to GL(N)$ to which the super-rigidity applies and delivers a homomorphism $\bar{h} : G \to GL(N)$. Then it is not hard to show that $\bar{h}$ sends G to G' (i.e. to the image of G' in GL(N) for the adjoint representation) and then we chose maximal compact subgroups $K \subset G$ and $K' \subset G'$ such that $\bar{h}(K) \subset K'$. Then we obtain a G-equivariant map $G/K \to G'/K'$ which moreover, can be made geodesic with an appropriate choice of $K' \supset \bar{h}(K)$ (see $[\text{Mos}]_5$) and which descends to the required map $V \to V'$. (We suggest to the reader to fill in the details).

Notice, that the above argument can be reversed. Namely, every geodesic map $V \to V'$ lifts to the universal coverings and gives a map $X \to X'$ which induces a homomorphism $G \to G'$. For example, if X is geodesically embedded into X' then Iso X embeds into Iso X' , as Iso X' is generated by geodesic symmetries of X around the points $x \in X' \subset X$.

In these lectures we do not present Margulis' proof but we explain how an arbitrary continuous map $\hat{h}_o : V \to V'$ (inducing h on Γ) can be deformed to a harmonic map $\hat{h}$ and then we shall show following Corlette that this $\hat{h}$ is, in fact, totally geodesic for the case $G = Sp(n,1)$.

2.9.C. Let us explain the origin of the non-precompactness of the $h(\Gamma)$ condition in the super-rigidity theorem. First we observe that Γ admits "many homomorphisms" (see 2.9.C_1. below) onto finite groups F which embed into some $GL(N)$. For example, F acts by translations $f(\alpha) \mapsto f(\beta\alpha)$ on the space L of *all* functions $f : \Gamma \to \mathbb{R}$, where $N = \dim L = \text{card}\, F$. It is perfectly clear that the resulting homomorphisms $h : \Gamma \to F = GL(N)$ do not usually extend to $G \supset \Gamma$. For instance, if G is simple, then every homomorphism $G \to GL(N)$ is either trivial or injective and so non-trivial non-injective homomorphisms h do not extend to G.

2.9.C_1. Let us explain where "many homomorphisms" $\Gamma \to F$ come from. The basic example is the group $\Gamma = SL_n\mathbb{Z} \subset SL_n\mathbb{R}$ which admits for every prime number p, a homomorphism $\Gamma \to F = SL_n\mathbb{F}_p$, where $\mathbb{F}_p$ denotes the finite field with p elements and $SL_n\mathbb{F}_p$ denotes the group of matrices with the entries from $\mathbb{F}_p$ and with determinant one. The homomorphisms $\Gamma \to F = SL_n\mathbb{F}_p$ is obtained by reducing the entries of integral matrices modulo p. Thus the kernel $\Gamma_p \subset \Gamma$ consists of the matrices having the diagonal entries equal $1 (\text{mod}\, p)$ and outside the diagonal they are $0 (\text{mod}\, p)$. (One can see directly that so defined Γ_p is a normal subgroup of finite index in Γ and then one constructs F as Γ/Γ_p without ever mentioning $\mathbb{F}_p$).

The above (mod p)-construction was extended by Selberg (see [Sel]) to all finitely generated subgroups in $GL(n)$. Thus Selberg has proven that Γ is *residually finite* that is *for every non-identity element* $\gamma \in \Gamma$ *there exists a homomorphism of* Γ *onto a finite group* F, *such that the image of* γ *in* F *is* $\neq \text{id}$.

2.9.C_2. Let us indicate an example of a homomorphism $\Gamma \to GL(N)$ with an *infinite* precompact image. We start with the group $\Gamma_o = \Gamma_o(q) \subset SL_n\mathbb{R}$ consisting of the matrices with the entries of the form $m+n\sqrt{q}$ where q is a fixed integer ≥ 2 and m and n run over $\mathbb{Z}$. Notice that Γ_o is a dense countable subgroup in $SL_n\mathbb{R}$. Next we observe that the transformation $\alpha : \Gamma_o \to \Gamma_o$ induced by

$$m+n\sqrt{q} \to m-n\sqrt{q} \qquad (*)$$

is an automorphism of Γ_o as (*) is an automorphism of the ring

$$\mathbb{Z}(\sqrt{q}) = \{m+n\sqrt{q}\}_{m,n\in\mathbb{Z}}.$$

Notice that the composition of α with the original embedding $i : \Gamma_o \subset SL_n\mathbb{R}$ defines another embedding, called $\bar{i} = \alpha \circ i : \Gamma_o \to SL_n\mathbb{R}$. This new embedding is (highly!) discontinuous for the topology in Γ_o induced by $i : \Gamma_o \subset SL_n\mathbb{R}$ (this is obvious) and therefore $\bar{i}$ does not extend to $SL_n\mathbb{R} \overset{i}{\supset} \Gamma_o$.

Another interesting property here is that the embedding $i \oplus \bar{i} = \Gamma_o \to SL_n\mathbb{R} \times SL_n\mathbb{R}$ has *discrete* image and this image has finite covolume in the group $SL_n\mathbb{R} \times SL_n\mathbb{R}$. (The discreteness of $i \oplus \bar{i} = \Gamma_o$ is rather obvious but the finite covolume property requires a little thought).

Now, we construct our example by taking the subgroup $\Gamma \subset \Gamma_o \subset SL_n\mathbb{R}$ which consists of the linear transformations of $\mathbb{R}^n$ preserving the quadratic form

$$f = \sum_{i=1}^{n-1} x_i^2 - \sqrt{q}\, x_n^2 .$$

We observe that

$$\Gamma = \Gamma_o \cap O(f) ,$$

where $O(f)$ is the subgroup in $SL_n\mathbb{R}$ consisting of *all real* matrices preserving f. Notice $O(f)$ is isomorphic to $SO(n,1)$ that is the group of transformations preserving

$$f_o = \sum_{i=1}^{n-1} x_i^2 - x_n^2 .$$

Then we observe that the second embedding $\bar{i}$ sends Γ onto the group

$$\Gamma' \underset{def}{=} \Gamma \cap O(\bar{f}) ,$$

where

$$\bar{f} = \sum_{i=1}^{n-1} x_i^2 + \sqrt{q}\, x_n^2 ,$$

and the pair $(i,\bar{i})$ *discretely* embeds Γ into $SO(f) \times SO(\bar{f})$. Now, we observe that $SO(\bar{f})$ is compact as it is (obviously) isomorphic to the usual special orthogonal group $SO(n)$ and so the discreteness of Γ in $SO(f) \times SO(\bar{f})$ implies the discreteness of Γ in $SO(f)$. (Projecting Γ from $SO(f) \times SO(\bar{f})$ to $SO(f)$ is just an instance of the step 3 in the construction of arithmetic groups indicated in 1.4.). It follows from the general theory of arithmetic groups that Γ is a lattice in $SO(f)$. But here, moreover,using the elementary fact that the only solution to the equation $f(x) = 0$ for $x \in \mathbb{Z}(\sqrt{q})$ is $x = 0$, one can easily show (without any general theory) that Γ is cocompact in $SO(f)$.

Thus we obtained a cocompact lattice $\Gamma \subset G = O(n-1,1)$ which admits an isomorphism on an (obviously dense) subgroup $\Gamma' \subset SO(n)$ and this isomorphism is by no means extendible to a homomorphism $SO(n-1,1) \to SO(n)$.

A judicious reader might notice that the group $O(n-1,1)$ is the one where the super-rigidity does not apply anyway. We leave to such a reader to work out a similar example for the group $O(n-2,2)$ which has rank 2 and to which the Margulis theorem applies.

2.10. Non-Archimedean superrigidity and arithmeticity. Margulis (see $[\mathrm{Mar}]_1$, [Zim]) proved his super-rigidity theorem for homomorphisms $h : \Gamma \to G'$ where G' may be a *p-adic* Lie group. Namely he has shown in this case that the image $h(\Gamma) \subset G'$ is necessarily precompact. This result, together with the ordinary (Archimedean) super-rigidity was used by Margulis to deduce his famous

Arithmeticity theorem. *Every lattice in a simple Lie group* G *of* $\mathrm{rank}_{\mathbb{R}} \geq 2$ *is arithmetic.*

If $\mathrm{rank}_{\mathbb{R}} G = 1$ then there may be non-arithmetic lattices $\Gamma \subset G$. For example every group $O(n,1)$ contains such a lattice (see [G-P]) as well as $U(2,1)$ and $U(3.1)$ (see $[\mathrm{Mos}]_6$). On the other hand, Corlette's super-rigidity proof with harmonic maps extends to the p-adic case (see [G-S]) which yields the arithmeticity theorem for $Sp(n,1)$ and F_4. The case of $U(n,1)$ for $n \geq 4$ remains open.

Finally, we notice that as on the previous occasions all of the super-rigidity and arithmeticity discussion extends to the *semi*-simple case if appropriate provisions are made for the "dangerous" $O(n,1)$ and $U(n,1)$-parts of G.

2.11. On the proofs of the rigidity theorems. The methods to prove rigidity are often more interesting and have wider scope than a particular result they may yield. Here is a list of basic methods some of which are discussed with more details later on.

I. Asymptotic geometry of G and $\Gamma \subset G$. This is sometimes called *long range, large scale* or *large distance* geometry.

The idea is to look at Γ from infinitely far away and recapture G (or at least the essential of G) from what one can see of Γ with a properly adapted eye (see §3).

II. Random walk on G and on Γ. If $\mathrm{Vol}\, G/\Gamma < \infty$ then the random walk on Γ, as the time parameter $\to \infty$, approximates the random walk on G. The idea of relating G and Γ via their random walks originates from the work of Fürstenberg (see $[\text{Für}]_1$, $[\text{Für}]_3$). The power of this idea was demonstrated by Margulis in the course of his proof of the super-rigidity theorem. An interested reader is referred to [Zim], $[\mathrm{Mar}]_3$.

III. Representation theory. One can think of the representation theory of a group as a linearization of random walks. Here again the passage from G to Γ is best possible if $\mathrm{Vol}\, G/\Gamma < \infty$ and one has an especially beautiful representation (action) of G on the space of L_2-functions on G/Γ. The most important link between G and Γ established up to-day by means of the representation theory is the equivalence of *Kazdan's* T-*property* for G and lattices in G. A quick (and hardly comprehensible) way to define T for G is by saying that the trivial representation of G is isolated (rigid) in the space of all unitary representations. Basic examples of groups having T are the simple Lie groups except O(n,1) and U(n,1). The reader can find this (and much more) in [Kaz], [Zim], [dlH-V].

IV. Geometry and combinatorics of subgroups in G of positive codimension. By intersecting a lattice $\Gamma \subset G$ with Lie subgroups $G' \subset G$ one can often produce "interesting" subgroups $\Gamma' \subset \Gamma$. Conversely, starting from certain subgroups $\Gamma' \subset \Gamma$ one can sometimes recapture the geometric pattern of the corresponding Lie subgroups and eventually, rebuild G out of Γ. This method goes back to the first paper by Selberg and then appears in Mostow's rigidity theorem for rank $r \geq 2$, where one uses Abelian subgroups of rank r (see $[\mathrm{Mos}]_4$). These ideas are also important in Margulis' super-rigidity proof.

V. Subgroups of finite index in Γ. It may be sometimes easier (and still very useful) to construct a p-adic Lie group $G_p \supset \Gamma$ rather than the original (real) Lie group $G \supset \Gamma$. For example, the group $\Gamma = SL_n\mathbb{Z}$ naturally embeds into $G_p = SL_n\mathbb{Q}_p$ as the integers embed into the field $\mathbb{Q}_p$ of the p-adic numbers. Notice that the closure $\bar{\Gamma}$ of this Γ in G_p is an *open* subgroup (consisting of the matrices with integer p-adic entries). It follows that the Lie algebra of G_p can be recovered from Γ and then one can recover G_p itself. Now, $\bar{\Gamma}$ is a compact totally disconnected group and so it is obtained by completing Γ for the topology defined by some system of subgroups $\Gamma_i \subset \Gamma$, $i \in I$ of finite index. in fact, in our case $\Gamma = SL_n\mathbb{Z}$ one gets the right Γ_i by taking the integral matrices $\equiv 1 \pmod{p^i}$.

The difficulty arising in this method is a characterization of the subgroups Γ_i in purely group theoretic terms. This problem does not appear however if Γ satisfies the *congruence subgroup property* saying that every subgroup of finite index in Γ contains a congruence subgroup distinguished by the $\equiv 1$ (mod q)-condition. Then one has a very quick proof of the super-rigidity theorem, as was pointed out to the authors by Alex Lubotzki. A reader interested in congruence subgroups is referred to [Bas].

VI. Harmonic maps. These are best explained in the language of locally symmetric spaces. If, for example, V and V' are locally symmetric spaces of non-compact type then every homomorphism between the fundamental groups, $\pi_1(V) \to \pi_1(V')$, is induced by a continuous map, say $f_o : V \to V'$, because the universal covering of V' is contractible. Next, if we assume V and V' are compact and we recall that the sectional curvature of V' is non-positive, then we are in a position to apply the Eells-Sampson theorem which provides a *harmonic* map $f : V \to V'$ homotopic to f_o (see [E-S], $[E\text{-}L]_1$, $[E\text{-}L]_2$ and §4 in these lectures). Now, we want to prove that f is, in fact, a geodesic map, which is by far stronger condition than mere harmonic. This can be sometimes done by means of Bochner type formulas as explained in §4.

§ 3 - Asymptotic geometry of symmetric spaces and the Mostow rigidity theorem.

We explain in this section how to recapture the geometry of a (symmetric) space X in terms of algebraic properties of a discrete cocompact group Γ acting on X. Then we sketch the proof of the Mostow rigidity theorem for symmetric spaces X of rank 1 (i.e. of negative sectional curvature) and indicate a sharpening of this theorem for the spaces $X = H^{4n}_{\mathbb{H}}$ and $H^{16}_{\mathbb{C}a}$.

3.1. Let a discrete group Γ isometrically act on a complete Riemannian manifold X, such that the quotient space X/Γ is compact. Then every orbit $\Gamma(x_0) \subset X$, $x_0 \in X$, is a net in X, i.e. there exists $R_0 \geq 0$ such that every ball of radius R_0 contains a point of $\Gamma(x_0)$ (compare 1.2.) . Furthermore, $\Gamma(x_0)$ is *uniformly discrete* in X, which means a uniform bound on the number of points of $\Gamma(x_0)$ in the ball : For every $x \in X$ and $R > 0$ the number of points of $\Gamma(x_0)$ in the ball $B_x(R) \subset X$ is bounded,

$$\#((\Gamma(X_0) \cap B_x(R)) \leq N = N(R).$$

To simplify the exposition we assume below that Γ acts freely on X. Then X can be identified with the universal covering of $V = X/\Gamma$ and $\Gamma(X_0)$ equals the pull-back of a point $v_0 \in V$ under the covering map $p : X \to V$.

The simplest property of X which can be expressed in terms of $\Gamma(x_0)$ (and eventually in terms of Γ itself) is the *asymptotic volume* of X. Namely, we fix a point $x \in X$ and look at the volume of the ball $B_x(R)$ as $R \to \infty$. We first observe that in our case the volume function

$$Vo_x(R) \underset{def}{=} Vol\, B_x(R)$$

grows at most exponentially. In fact, since $X/Iso\, X$ is compact, a simple argument shows that

$$Vo_x(R + R') \leq CVo_x(R) \qquad (*)$$

for some constant $C = C(R')$ depending only on R' and the geometry of $X/\mathrm{Iso}\, X$. In particular,

$$Vo_x(R) \leq (C(1))^{R+1}$$

(Notice that this bound on the volume is satisfied whenever Ricci $X \geq -\delta > -\infty$, see [G-L-P]). Then we notice that the function $Vo_x(R)$ changes by an at most bounded amount if we move the point x,

$$Vo_x(R)/Vo_{x'}(R) \leq \mathrm{const} < \infty.$$

In fact the above constant satisfies

$$\mathrm{const} \leq C(d) \quad \text{for } d = \mathrm{dist}(x, x'),$$

as

$$B_{x'}(R) \subset B_x(R + d).$$

Next we compare the volume function $Vo_x(R)$ with the number

$$Nu_x(\Lambda ; R) = \#(\Lambda \cap B_x(R))$$

for a given net $\Lambda \subset X$. Using the net properties and the fact that

$$\mathrm{Vol}\, B_x(R) \leq \mathrm{const}\, (R)$$

(which follows from the above (*)), one immediately sees that

$$Vo_x(R) \leq C\, Nu_x(\Lambda, R), \; R > 0, \qquad (+)$$

for some positive constant $C = C(X, \Lambda)$. Next we observe that the volume of the unit ball $B_x(1) \subset X$ is bounded away from zero uniformly for all $x \in X$, since $X/\mathrm{Iso}X$ is compact. Then we conclude that

$$\mathrm{Vo}_x(R) \geq C\, \mathrm{Nu}_x(\Lambda, R), \qquad (++)$$

for some C as above and for all $R \geq R_o$, where R_o is another constant depending on X and Λ.

We express (+) and (++) together by the notation

$$V_o \sim \mathrm{Nu},$$

which should convey the idea of similar asymptotic growths of the functions $\mathrm{Vo}(R)$ and $\mathrm{Nu}(R)$ for $R \to \infty$.

Now we are interested in the case $\Lambda = \Gamma(X_o)$ and we want to relate the function $\mathrm{Nu}(\Gamma(X_o), R)$ to some characteristic of the group Γ itself.

The key to such a relation is the following definition.

3.1. A. Word metric. Let $\Delta = \{\gamma_1, \dots \gamma_k\}$ be a system of generators of Γ. Then for each $\gamma \in \Gamma$ we define $|\gamma|_\Delta$ as the length of the shortest word in $\gamma_1, \dots \gamma_k$ and $\gamma_1^{-1}, \dots \gamma_k^{-1}$ representing γ. Then we define a *metric* $\mathrm{dist}_\Delta(\gamma, \gamma')$ on Γ by

$$\mathrm{dist}_\Delta(\gamma, \gamma') = |\gamma^{-1} \gamma'|_\Delta.$$

It is easy to see that this is indeed a metric which is, moreover, left invariant on Γ,

$$\text{dist}(\alpha\gamma, \alpha\gamma') = \text{dist}(\gamma, \gamma')$$

In fact dist_Δ is the *maximal* left invariant metric on Γ, satisfying

$$\text{dist}_\Delta (\gamma_i^{\pm 1}, \text{id}) \leq 1.$$

This dist_Δ is what is called the *word metric* associated to Δ.

3.1.A. Lemma. *Every two word metrics on Γ are bi-Lipschitz equivalent. That is for every two finite generating systems Δ and Δ' the ratio $\text{dist}_\Delta/\text{dist}_{\Delta'}$ is a bounded function on the set of distinct pairs of points γ and γ' in Γ and the ratio $\text{dist}_{\Delta'}/\text{dist}_\Delta$ is also bounded.*

Proof. As the system Δ' is finite the length function $|\gamma|_\Delta$ is (obviously) bounded on $\Delta' \cup (\Delta')^{-1}$. Then this bound (obviously) extends to all of Γ as Δ' generates Γ.

3.1.A' Remark . An equivalent way to express the lemma is to say that every *isomorphism* between finitely generated groups $f : \Gamma \to \Gamma'$ is bi-Lipschitz for arbitrary word metrics in Γ and Γ'. In fact, every *homomorphism* (obviously) is *Lipschitz* ,

$\text{dist}_{\Gamma'} (f(\alpha), f(\beta)) \leq \text{const}\ \text{dist}_\Gamma(\alpha, \beta)$ for all α and β in Γ and some $\text{const} \geq 0$.

Next, we need a similar lemma which relates the geometry of Γ and an orbit $\Gamma(x_o)$.

3.1.B. *Let Γ discretely freely and isometrically act on a complete Riemannian manifold X. If X/Γ is compact then the (bijective) map $\Gamma \to \Gamma(x_0)$ for $\gamma \to \gamma(x_o)$ is bi-Lipschitz. That is there exists $\varepsilon > 0$, such that*

$$\varepsilon \leq \text{dist}_\Delta(\gamma, \gamma')/\text{dist}_X(\gamma(x_o), \gamma'(x_o)) \leq \varepsilon^{-1},$$

where dist_Δ denotes some word metric on Γ, where dist_X denotes the metric on X and where (γ, γ') run over all pairs of distinct elements in Γ.

Proof. Let $\Delta = \{\gamma_1, \dots \gamma_k\}$ and let $\ell_1 \dots \ell_k$ be smooth geodesic loops in V with a given base point $v_o \in V$ which represent the generators $\gamma_1, \dots \gamma_k$ of the group Γ identified with the fundamental group $\pi_1 (V, v_o)$. Then we consider the pull-back $\tilde{\Gamma} \subset X$ of the union

$\ell_1 \cup \ldots \cup \ell_k \subset V$ under the covering map and observe that $\tilde{\Gamma}$ is a *connected graph* in X consisting of geodesic segments joining certain pairs of points in the orbit $\Gamma(x_0)$ for some point $x_0 \in X$ over v_0. Notice that these segments that are edges of the graph $\tilde{\Gamma}$ may accidentally meet at some points besides the vertices of $\tilde{\Gamma}$ which are the points $\{\gamma(x_0)\}$, $\gamma \in \Gamma$. This happens if some loop ℓ has a double point or if two loops intersect at a point $v \neq v_0$. If $\dim V \geq 3$, we can slightly perturbe ℓ_i in order to remove extra intersection. In general, we pretend these intersections are not there and treat $\tilde{\Gamma} \subset X$ as an abstract graph with the vertices $\{\gamma(x_0)\}$, $\gamma \in \Gamma$ and the edges corresponding our geodesic segments.

Then we introduce the *generalized word* metric $\widetilde{\text{dist}}$ in $\Gamma = \Gamma(x_0) \subset \tilde{\Gamma}$ as the length of the shortest path between γ and γ' in $\tilde{\Gamma}$. Notice that this reduces to the ordinary dist_Δ if length $\ell_i = 1$ for $i = 1, \ldots, k$ and that $\widetilde{\text{dist}}_\Delta$ is, obviously, Lipschitz equivalent to dist_Δ.

Now the proof of the lemma reduces to the following.

3.1.B'. Sublemma. *The metric* dist_X *is bi-Lipschitz equivalent to* $\widetilde{\text{dist}}$ on $\tilde{\Gamma} \subset X$.

Proof. The inequality

$$\text{dist}_X \leq \widetilde{\text{dist}}$$

is obvious. In fact, $\text{dist}_X(x, y)$ equals the length of the shortest path in X between the points x and y chosen in $\tilde{\Gamma} \subset X$, while the metric $\widetilde{\text{dist}}$ requires such a path to lie in $\tilde{\Gamma}$ which makes it, a priori, longer.

To prove the reverse inequality we start with the shortest path $[x, y]$ in V between x, y, which is a geodesic segment of length $d = \text{dist}_X(x, y)$. Then we subdivide $[x, y]$ by some points $x = t_0, \ldots t_m, t_{m+1} = y$ into the subsegments

$$[x, t_1], [t_1\, t_2], \ldots [t_m, y],$$

such that the length of each of these subsegments does not exceed 1 and the number of them (i.e. $m + 1$) is no more than d.

Finally, for each point t_i, $i = 1, \ldots, m$ we take the nearest point $\tilde{t}_i \subset \tilde{\Gamma}$ and observe that

$$\text{dist}_X(t_i, \tilde{t}_i) \leq C,$$

where

$$C = \sup_{x \in X} \operatorname{dist}(X, \Gamma(x_0)) < \infty,$$

as $\Gamma(x_0)$ is a net in X. It follows, by the triangle inequality that

$$\operatorname{dist}_X(\tilde{t}_i, \tilde{t}_{i+1}) \le 2C + 1$$

for all t_i, $i = 0, \ldots, m + 1$.

Now, we use the cocompactness of Γ on $\tilde{\Gamma}$ and conclude to the existence of a constant $\tilde{C}$, such that

$$\widetilde{\operatorname{dist}}(\tilde{t}_i, \tilde{t}_{i+1}) \le \tilde{C} \quad \text{for } i = 0, \ldots, m + 1$$

It follows, that

$$\widetilde{\operatorname{dist}}(x, y) \le \tilde{C}(\operatorname{dist}_X(x, y) + 1),$$

which yields the required inequality

$$\widetilde{\operatorname{dist}} \le \tilde{C}' \operatorname{dist}_X \qquad \text{on } \Gamma(x_0)$$

with a somewhat larger constant $\tilde{C}'$ as both distances are $> \varepsilon \ge 0$ on distinct points of $\Gamma(x_0)$.

3.1.C. Let us summarize what we have achieved.

I. The volume growth of X is equivalent to the growth of $\Gamma(x_0) \subset X$ for the metric dist_X,

$$\operatorname{Vo}(X ; R) \sim \operatorname{Nu}(\Gamma(x_0) ; R).$$

II. The orbit $\Gamma(x_0)$ with the metric dist_X is bi-Lipschitz to Γ with some word metric $\operatorname{dist}_\Delta$.

Now using $\operatorname{dist}_\Delta$ one can introduce the growth function $\operatorname{Nu}(\Gamma, \Delta ; R)$ that is the number of points in the $\operatorname{dist}_\Delta$-ball in Γ of radius R around $\mathrm{id} \in \Gamma$.

To say it differently, $\operatorname{Nu}(\Gamma, \Delta ; R)$ is the number of elements in Γ representable by Δ-words of length $\le R$.

Unfortunately, the bi-Lipschitz equivalence of metrics does not lead to equivalent growth functions. In fact, the distance R becomes about CR under a bi-Lipschitz equivalence but N(R) is not equivalent to N(CR) for most functions N(R). For example exp 2R is not equivalent to exp R in our sense. However, our growth functions in question have at most exponential growth. For example, the number of words of length $\leq R + 1$ does not exceed 2k (the number of words of length $\leq R$), where k is the number of generators. Thus

$$\mathrm{Nu}(R + 1) \leq 2k\, \mathrm{Nu}(R)$$

which is the required growth bound.

Using this one immediately sees that

$$\log \mathrm{Nu}(CR) \leq \mathrm{const} \log \mathrm{Nu}(R)$$

that is

$$\log \mathrm{Nu}(CR) \sim \log \mathrm{Nu}(R).$$

Also one obtains this equivalence for the bi-Lipschitz equivalent metrics and then concludes to the following.

3.1.D. Theorem. *The volume growth function* Vo(R) *of* X *and the growth function* Nu(R) *for some word metric* dist_Δ *in* Γ *satisfy*

$$\mathrm{Log}\, \mathrm{Nu}(R) \sim \log \mathrm{Vo}(R) \quad \text{for } R \to \infty,$$

that is

$$0 < \varepsilon \leq \log \mathrm{Nu}(R)/\log \mathrm{Vo}(R) \leq \varepsilon^{-1} < \infty$$

for all sufficiently large R. (we assume Γ is infinite and so Nu(R) and Vo(R) go to ∞ for $R \to \infty$)

3.1.D'. Remarks (a) This theorem was first stated and proven by A.Švarc in 1955, following a hint by Efremowitz (see [Efr], [Šv]). Then this result was rediscovered by J. Milnor in 1968. (see [Mil]$_2$). In fact, the results of Švarc and Milnor are somewhat more precise than 3.1.D. For example, they show, that if the group Γ is free Abelian of rank k, then the balls $B_x(R) \subset X$ have

$$\mathrm{Vol}\, B_x(R) \geq CR^k \quad \text{for } R \to \infty.$$

We suggest to the reader to furnish the proof which is a slight modification of the argument presented above. (Notice that

$$\log R^k \sim \log R^\ell$$

for all k, $\ell > 0$ and so 3.1.D is useless here).

(b) Theorem 3.1.D shows how, in principle, one can relate asymptotic invariants of X and Γ. But it seems infinitely far from recapturing X itself from Γ, rather than only the growth rate of Γ. Yet some ideas in the proof will serve in the Mostow theorem later on where we take into account the shapes of X and Γ as well as of their mere sizes.

3.2. Quasi-isometries of metric spaces. Let us formulate explicitely the basic relation which connects a Riemannian manifold X and a cocompact lattice $\Gamma \subset \text{Iso}\, X$.

3.2.A. Definition . Two metric spaces X and Y are called *quasi-isometric* if there exist nets $X' \subset X$ and $Y' \subset Y$ and a bi-Lipschitz equivalence $f : X' \to Y'$, where X' is given the metric dist_X and Y' comes with dist_Y.

3.2.A'. Basic example. Let X and Γ be as above and Γ is given some word metric. Then X and Γ are quasi-isometric. In fact, if Γ acts freely at some point $x_0 \in X$ one can take $\Gamma(x_0)$ as a net in X and use all of Γ as a net in itself. Then the orbit map $\gamma \mapsto \gamma(x_0)$ is bi-Lipschitz. This has been shown in the previous section for free actions and the general case requires a minor adjustment of the argument.

Remark. It is true, but not automatic, that the quasi-isometry is an equivalence relation. However, it is clear in the above situation, that if the same group Γ cocompactly and isometrically acts on two different manifolds X and Y then X and Y are quasi-isometric. For example, if two *compact* manifolds V and W have isomorphic fundamental groups, then the universal covering $\tilde{V}$ and $\tilde{W}$ are quasi-isometric. In particular, such $\tilde{V}$ and $\tilde{W}$ have similar volume growth,

$$\log \text{Vo}(\tilde{V}\,;R) \sim \log \text{Vo}(\tilde{W}\,;R)$$

3.2.B. Let us give a more flexible definition of quasi-isometry. For this we need multi-valued maps $f : X \dashrightarrow Y$, where f(x) is a subset in Y for each $x \in X$. Such a map can be represented by its graph

$$\Gamma r_f = \{x, y \mid y \in f(x)\} \subset X \times Y$$

and its convenient to allow arbitrary subsets in $X \times Y$ corresponding to *partially* defined multivalued maps $X \dashrightarrow Y$, which are called *correspondences*. . We define the (Hausdorff) distance between correspondences f, and g, denoted $|f - g|$, as the Hausdorff distance between their graphs Γr_f and Γr_g in $X \times Y$, where $X \times Y$ is given the sup-product metric

$$\operatorname{dist}((x_1, y_1), (x_2, y_2)) = \max(d^X, d^Y)$$

for $d^X = \operatorname{dist}(x_1, x_2)$ and $d^Y = \operatorname{dist}(y_1, y_2)$. Also recall that the Hausdorff distance between subsets A and B in a metric space is the infimum of those ε for which A is contained in the ε-neighbourhood of B and, conversely, B is contained in the ε-neighbourhood of A.

An important equivalence relation between maps is $|f - g| < \infty$.

<u>Example</u>. Let $X' \subset X$ be a uniformly discrete net and let $\nu(x) \subset X$ consist of the nearest to x points in X' (generically f(x) consists of a single point but for some $x \in X$, f(x) contains more the one element). Then, clearly, this ν, called the *normal projection* of X to $X' \subset X$, is equivalent to the identity map, that is $|\mathrm{id} - \nu| < \infty$.

The following definition of quasi-isometry is invariant under the above equivalence relation.

3.2.B'. A map (correspondence) $f : X \dashrightarrow Y$ is called *quasi-isometry* if it satisfies the following two conditions.

QI_1. There exist positive constants A and B such that every four points x_1 and x_2 in X and $y_1 \in f(x_1) \subset Y$ and $y_2 \in f(x_2) \subset Y$ satisfy the following relation

$$A^{-1}\operatorname{dist}(x_1, x_2) - B \leq \operatorname{dist}(y_1, y_2) \leq A\operatorname{dist}(x_1, x_2) + B, \qquad (*)$$

(where we assume x_1 and x_2 lie in the domain of the definition of f).

QI_2. The domain of definition of f is a net in X while the image of f is a net in Y. That is, the projections of $\Gamma r_f \subset X \times Y$ to X and Y are nets.

3.2.B". <u>Example</u>. Let $X' \subset X$ and $Y' \subset Y$ be uniformly discrete nets and $f' : X' \to Y'$ be a bijection. Then the related correspondence $f : X \dashrightarrow Y$ with $\Gamma r_f = \Gamma r_{f'} \subset X' \times Y' \subset X \times Y$ is a quasi-isometry if and only if f' is a bi-Lipschitz map.

There is the following converse to this statement. For an arbitrary quasi-isometry $f : X \to Y$ there exist nets $X' \subset X$ and $Y' \subset Y$ and a bi-Lipschitz map $f' : X' \to Y'$ equivalent to f. The proof is a trivial exercise on the notions of a net and a quasi-isometry.

With all these definitions we can modify our basic problem of relating the geometries of X and Γ as follows : which geometric properties of X (and Γ) are quasi-isometry invariant ? So far we have only one such invariant namely the (logarithm of the) volume growth of X. But we want much more, at least if X is a symmetric space. Namely we want to reconstruct all of the geometry of X in quasi-isometric terms. We shall see below that this is possible for some symmetric spaces X.

3.3. <u>The group of quasi-isometries</u> $\overline{\text{QIs}}(X)$. Composition of two quasi-isometries is not, in general, a quasi-isometry. In fact, if the image of $f : X \dashrightarrow Y$ misses the domain of definition of $g : Y \dashrightarrow Z$, then $g \circ f : X \dashrightarrow Z$ has empty graph in $X \times Z$ and so is *not* quasi-isometric. To remedy this situation we consider only *full quasi-isometries* $f : X \dashrightarrow Y$, where the domain of definition equals X and the image is Y. Now, there is no problem to compose these.

Composition of full quasi-isometries (obviously) is a full quasi-isometry.

Since the inverse of a full q.i. is again a full q.i. we have a group, denoted FQIs(X) for every metric space X. This group appears too big to be useful but there is a reasonable factor group, denoted $\overline{\text{QIs}}(X)$, where the relevant normal subgroup consists of the quasi-isometries equivalent to the identity map. In other words, one goes from FQIs to $\overline{\text{QIs}}$ by using the equivalence relation $|f - g| < \infty$ on FQIs.

3.3.A. There is an obvious homomorphism of Iso X to $\overline{\text{QIs}}(X)$. For example, if $X = \mathbb{R}^n$ then the kernel of this homomorphism consists of the parallel translations of $\mathbb{R}^n$ and so the image is isomorphic to $O(n)$. It means, in plain words that an isometry f of $\mathbb{R}^n$ has $|f - \text{id}| < \infty$ if and only if f is a parallel translation. To see this, one should notice that the metric $|f - g|$ on isometries is equivalent to the ordinary

$$\sup_{x \in X} \text{dist}(f(x), g(x)) .$$

3.3.A' Our next example is that of a symmetric space X with Ricci $X < 0$ (i.e. X has no compact and flat de Rham factors). Then the homomorphism

$$\text{Iso } X \to \overline{\text{QIs}}(X)$$

is injective. In fact, let X be an arbitrary complete simply connected Riemannian manifold with non-positive sectional curvature and let $f : X \to X$ be an isometry, such that $f \neq \text{id}$ and

$$\sup_{x \in X} \text{dist}(x, f(x) = s < \infty.$$

Then (see [B-C-S]) X isometrically splits.

$$X = X' \times \mathbb{R}$$

such that the isometry becomes

$$f : (X', t) \mapsto (x', t + s).$$

3.3.A". Remark. Let us interpret the above result in terms of the group $G = \text{Iso } X$. Namely, we observe that the condition $|f - \text{id}| < \infty$ for a left invariant metric in G is equivalent to

$$\text{dist}(fg, g) = |g^{-1} f^{-1} g| \leq C < \infty$$

for all $g \in G$. Intuitively, this says that f "almost commutes" with all $g \in G$. For example if G is a discrete group, then this condition implies that the centralizer of f has finite index in G. Furthermore if G is a non-compact simple group, then f must be in the center of G. We leave the proofs of all this to the reader.

3.3.B. Let us evaluate the group $\overline{\text{QIs}}$ for $X = \mathbb{R}$. We observe, that for every continuous bounded function $\varphi : \mathbb{R} \to \mathbb{R}$ such that $\inf |\varphi| > 0$, the integral $t \mapsto \int_0^t \varphi(\tau)\, d\tau$ is a quasi-isometry (in fact, a bi-Lipschitz homeomorphism) of $\mathbb{R}$. Two such quasi-isometries give the same element in $\overline{\text{QIs}}$ if and only if the corresponding function φ_1 and φ_2 have

$$\sup_{t \in \mathbb{R}} \left| \int_0^t (\varphi_1 - \varphi_1)(\tau)\, d\tau \right| < \infty .$$

This shows that the group $\overline{\text{QIs}}\, \mathbb{R}$ is infinite dimensional.

Exercice. Show that the above $\int \varphi$ represent all of the group $\overline{\text{QIs}}\mathbb{R}$.

3.4. Rigid metric spaces. A metric space X is called QIs-*rigid* if the homomorphism

$$\text{Iso } X \to \overline{\text{QIs}}X$$

is bijective (Compare [Ul].

The notion of a rigid space sharpens Mostow rigidity. In fact, every isomorphism between discrete groups Γ_1 and Γ_2 is a quasi-isometry and if Γ_1 and Γ_2 are cocompact lattices in Iso X, then the quasi-isometry $\Gamma_1 \to \Gamma_2$ defines a quasi-isometry $X \to X$. If X is rigid, this is equivalent to an isometry, as required by Mostow's rigidity. (Our definition in 2.2.A. allows different spaces X and Y and one needs an additional theorem saying that two symmetric spaces of non-compact type are isometric if and only if they are quasi-isometric. Such a theorem can be proven but we do not concern ourselves here with this aspect of rigidity).

In view of 3.3.B. the rigidity of X may appear an extremely unlikely property. In fact, one sees with 3.3.B. that the Euclidean space $\mathbb{R}^n$, $n \geq 1$, is not rigid. We shall also see in 3.11.C that the *hyperbolic spaces* $H^n_{\mathbb{R}}$ *and* $H^{2n}_{\mathbb{C}}$ *are not QIs-rigid.* However, it seems that *apart from* $H^n_{\mathbb{R}}$ *and* $H^{2n}_{\mathbb{C}}$ *all noncompact irreducible symmetric spaces are rigid.*

This was conjectured for the spaces of rank ≥ 2 by Margulis about 15 years ago (according to a private communication by G. Prasad) but the proof of Margulis'conjecture is still not available. On the other hand the QIs-rigidity for $H^{4n}_{\mathbb{H}}$ and $H^{16}_{\mathbb{C}a}$ has been proven in $[\text{Pan}]_2$ and the idea of the proof is explained below in 3.11.

3.4.A. Remark on non-rigid spaces. Even if X is non-QIs-rigid one may try to define Iso X in quasi-isometric terms. One possibility is to use *uniformly quasi-isometric families* F of maps $f : X \to X$. This means there exists a constant A depending on F, such that each $f \in F$ satisfy the q.i. inequality (*) of 3.2.B with this A and some B depending on f. One can show that the subgroup

$$\text{Iso } H^n_{\mathbb{R}} \subset \overline{\text{QIs}}\ H^n_{\mathbb{R}}\ ,\ n \geq 2$$

is a *maximal subgroup of uniform* quasi-isometries. That is, for each $\varphi \in \overline{\text{QIs}} - \text{Iso}$ the subgroup generated by Iso and this φ contains quasi-isometries with arbitrary large (Lipschitz) constant A. Probably Iso $H^n_{\mathbb{R}}$ is a *unique* (up to conjugation in $\overline{\text{QIs}}$) maximal uniformly quasi-isometric subgroup subject to some extra property (e.g. being connected or something similar), and the same should be true for $H^{2n}_{\mathbb{C}}$. However, this weakened QIs-rigidity does not seem to directly imply Mostow's rigidity.

3.5. <u>The sphere at infinity</u>. Recall that a (geodesic) *ray* in a complete Riemannian manifold X is an isometric copy of $\mathbb{R}_+$ in X. On can think of a ray as of a *subset* $R \subset X$ isometric to $\mathbb{R}_+$ or, sometimes, as an isometric map $\mathbb{R}_+ \to X$. By the definition of the Riemannian metric in X, the length of every segment $[a, b] \subset R \subset X$ equals $\mathrm{dist}_X(a, b)$ and so R is a *geodesic* ray.

Call two rays R_1 and R_2 in X *equivalent* if the Hausdorff distance between them is *bounded*. Then define the *ideal boundary* ∂X as the set of the equivalence classes of rays in X. The point $s \in \partial X$ represented by a ray R in X is called the (ideal) *end* of R.

3.5.A. <u>Remarks and examples</u> (a) There are many alternative definitions of an ideal boundary (see $[Gr]_2$, $[Gr]_3$) but they all turn out to be equivalent for symmetric spaces of rank one which are studied in this section. On the other hand, our definition may become quite ugly for some manifolds. Yet it serves well for *simply connected* manifolds with *non-positive* sectional curvature (e.g. for symmetric spaces of non-compact type).

(b) Let see what happens to $X = \mathbb{R}^n$. Here two rays are equivalent iff they are parallel and point in the same direction. Therefore, $\partial\mathbb{R}^n$ can be identified with the sphere S^{n-1} of the "directions" in $\mathbb{R}^n$ represented by the unit vectors.

(c) One may think of $X \cup \partial X$ as a *compactifiction* of X. In fact, let $x_i \in X$ be a divergent sequence of points such that dist $(x_o, x_i) \to \infty$ for $i \to \infty$. Then we consider minimizing geodesic segments $[x_o, x_i] \subset X$ between x_o and x_i and take a subsequence $[x_o, x_j]$ for which the unit tangent vectors τ_j to $[x_o, x_j]$ at x_0 converge to some vector τ in the unit tangent sphere $S_{x_o}(X)$. It is easy to see that the one -sided infinite geodesic R in X issuing from x_o in the direction of τ is a (geodesic) *ray* in X as the length of every segment $[x_o = 0, r] \subset R = \mathbb{R}_+$ equals $\mathrm{dist}_X(x_o, r)$. Thus one thinks of (the end of) R as an *ideal* (sub)-*limit* of the sequence x_i.

3.5.B. <u>The case</u> $K(X) \leq 0$. Let the sectional curvature $K(X)$ of X be non-positive and let us also assume X is simply connected as well as complete. Then by the Cartan-Hadamard theorem (see [B-G-S]) every one sided infinite geodesic is a (geodesic) ray in X. Thus if we take all geodesic rays in X issuing from a given point $x_o \in X$ we obtain a map of the unit tangent sphere $S_{x_o}(X)$ into ∂X. Now we have the following basic theorem also going back to Cartan and Hadamard.

3.5.B_1. *The above map say* $G_{x_o} : S_{x_o} \to \partial X$ *is a bijection. Moreover, the composed map* $G_{x_o}^{-1} G_{y_1} : S_{y_1} \to S_{x_o}$ *is a homeomorphism for every pair of points* x_o *and* y_1 *in* X.

Sketch of the proof. The essential property of X one uses is the following.

3.5.B_2. Convexity of dist. *The function* $(x, y) \mapsto \text{dist}(x, y)$ *is convex on* $X \times X$. *That is for every two geodesics* $g_1 : \mathbb{R} \to X$ *and* $g_2 : \mathbb{R} \to X$ *the function* $\text{dist}(g_1(t), g_2(\theta))$ *is convex on* $\mathbb{R}^2$. (See [B-G-S]).

Now one immediately sees that the map $S_{x_0} \to \partial X$ is *injective* as for any two distinct rays issuing from x_0, the distance $\text{dist}(r_1(t), r_2(t))$ goes to infinity for $t \to \infty$. In fact, every *convex* function $d(t)$ with $d(0) = 0$ goes to infinity unless it is identically zero.

Next, to see that our map G_{x_o} is surjective we take an arbitrary ray $R_1 \subset X$ issuing from $r_1 \in X$ and look at the geodesic segments $[x_o, r_1] \subset X$ for all $r_1 \in R_1$ and a fixed x_o. Then we let $r_1 \to \infty$ and we need to show that the segments $[x_o, r_1]$ converge (or at least subconverge) to a ray R_o issuing from x_o, such that $R_o \sim R_1$, which means bounded distance

$$\text{dist}(r_0, r_1) \le \text{const} < \infty$$

for all $r_o \in R_o$ and $r_1 \in R_1$ satisfying $\text{dist}(r_o, x_o) = \text{dist}(r_1, x_o)$ (This is obviously equivalent to the boundedness of Hausdorff distance).

Denote by $\tau(r_1) \in S_{x_o}(X)$ the tangent vector to $[x_o, r_1]$ at x_o and look at the resulting curve $R_1 = \mathbb{R}_+ \to S_{x_o}(X)$ for $r_1 \mapsto \tau(r_1)$.

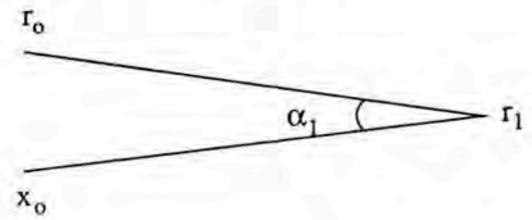
3.5.B_3. Lemma. The curve $\tau(r_1)$, $r_1 \in [0, \infty]$ has finite length in the sphere $S_{x_o}(X)$.

Proof. The radial projection of $X - \{x_0\}$ to $S_{x_o}(X)$ contracts every tangent vector $\tau_1 \in S_{x_1}(X)$ by the factor at least $(\mathrm{dist}(x_o, x_1))^{-1}$ as easily follows from the convexity of the distance function between the segment $[x_o, x_1]$ and nearby segments x_o, x_1', for infinitesimal pertubations x_1' of x_1 in the direction of τ_1.

Moreover, this contraction is even stronger for all vectors τ_1 having small angle α_1 with the segment $[x_o, x_1]$. Namely it is of the order $\alpha_1(\mathrm{dist}(x_o, x_1))^{-1}$.

For example, the vectors tangent to $[x_o, x_1]$ go to zero under the (differential of the) radial projection.

Now, the angle α_1 in our case can be again estimated using the convexity of the distance between the segments $[x_o, r_1]$ and $[y_1, r_1]$,



This convexity shows that

$$\alpha_1 < c(\mathrm{dist}\,(x_o, r_1))^{-1}$$

and so the length of our curve is bounded by the integral

$$\int_0^\infty c\left(\mathrm{dist}\,\left(x_o, r_1\right)\right)^{-2} dr_1 .$$

As $\mathrm{dist}\,(x_o, r_1) \geq r_1 - \mathrm{dist}\,(x_o, y_1)$, this integral converges for $r_1 \to \infty$ and the lemma follows.

Exercise. Show that the above curve in the sphere has length $< \pi$.

3.5.B_4. Now we see that $\tau(r_1)$ converges in $S_{x_o}(X)$ to some vector τ_o and we take the ray R_o defined by τ_o. Again, the convexity of the distances from $[x_o, r_1]$ to R_1 shows that the distance from R_o to R_1 is bounded (by dist(x_o, y_1)) and so $R_o \sim R_1$ as required. The final *homeomorphism* property follows from the principle of L.E.J. Brower (applied to $G_{x_o}^{-1} G_{y_1}$ and $G_{y_1}^{-1} G_{x_o}$) claiming that every effectively defined map is continuous (The readers who do not trust this principle may check the continuity by looking carefully at the above (effective !) construction of R_o).

3.5.C. *Conformal boundary* $\partial_\varphi X$. Let us give another definition of the ideal boundary of X using a metric on X conformal to the original metric g on X. Such a conformal metric is $g_{new} = \varphi^2 g$ where φ is a continuous positive function on X. The new length of a smooth curve $x(t), a \le t \le b$, in X is given by the integral

$$\text{length}_{new}\, x(t) = \int_a^b \varphi(x(t))\, d\ell_g ,$$

where ℓ_g is the g-length parameter on the curve. Then the new distance between two points is defined as the infimum of such integrals for all curves joining the points in question.

3.5.C_1. Definition (Floyd, see [Fl]). The boundary $\partial_\varphi X$ is the set of the "new" points of the *metric* completion $\bar{X}_\varphi \supset X$, that is

$$\partial_\varphi X = \bar{X}_\varphi - X.$$

Remarks (a) Unlike the geodesic boundary ∂X which is just a set, the conformal boundary $\partial_\varphi X$ comes with a structure of a *metric* space, for the metric induced from $\bar{X}_\varphi$. This metric on $\partial_\varphi X$ (and on $\bar{X}_\varphi$) will be called the φ-*metric*.

(b) As we want $\partial_\varphi X$ to be non-empty we must choose φ, such that the metric space $(X, \varphi g)$ is *non-complete* . For this we need $\varphi(x)$ which decays sufficiently fast for $x \to \infty$, such that for some curve $x(t), t \in [0, \infty]$, in X with $x(t) \to \infty$, the integral $\int_a^\infty \varphi(x(t))\, d\ell_g$ converges and so $x(t)$ converges in the new metric to some point $x_\infty \in \partial_\varphi X$. On the other hand, we want more than a single point for $\partial_\varphi X$ and so we need a *positive* lower bound for the diameters in the φ-metric of the exteriors of the g-balls $B(r) \subset X$ for $x \to \infty$. Thus $\varphi(x)$ should not decay too fast. We shall see soon that a proper balance can be achieved for manifolds X with $K \leq -\kappa^2 < 0$.

3.5.C_2. If two functions φ and φ' are *equivalent* on X in the sense that the ratios φ/φ' and φ'/φ are bounded on X, then the identity map $(X, \varphi^2 g) \leftrightarrow (X, \varphi'^2 g)$ is bi-Lipschitz. Therefore it extends by continuity to a (bi-Lipschitz) homeomorphism of the completions

$$\bar{X}_\varphi \leftrightarrow \bar{X}_{\varphi'} .$$

In particular, $\partial_\varphi X = \partial_\varphi X'$.

More generally, let $f : X \to Y$ be a Lipschitz map and $\varphi(x) = \psi(f(x))$ for some positive function ψ on Y then f is also Lipschitz for the conformal metrics in X and Y with the multipliers φ and ψ. Hence, f extends to a Lipschitz map $\bar{X}_\varphi \to \bar{X}_\psi$ which sends $\partial_\varphi X$ to $\partial_\psi X$ whenever f is *proper*, i.e. $x \to \infty$ implies $f(x) \to \infty$. The same remains true if we replace φ by an equivalent function φ' on X. Furthermore, if f is a bi-Lipschitz homeomorphism the resulting boundary map $\partial_{\varphi'} X \to \partial_\psi Y$ is a homeomorphism.

3.5.C_2'. Example. Let $\psi(y) = (1 + \mathrm{dist}(y_0, y))^\alpha$ for a fixed point $y_0 \in Y$ and some $\alpha \in \mathbb{R}$, and let $\varphi'(x)$ be a similar function with the same exponent α. Then for every bi-Lipschitz map $f : X \to Y$ the composed function $\varphi = \psi \circ f$ obviously is equivalent to φ' and so we obtain the boundary homeomorphism $\partial_{\varphi'} X \to \partial_\psi Y$.

3.5.C_2''. Remark. If X is complete then the boundary $\partial_\varphi X$ is non-empty for the above φ if and only if $\alpha < -1$. Indeed, for $\alpha < -1$ the integral of φ over each geodesic ray in X converges ; on the contrary if $\alpha \geq -1$ one gets divergent integrals over all infinite curves in X.

3.5.D. Let us construct the boundary map $\partial_\varphi X \to \partial_\psi Y$ for an arbitrary quasi-isometry $f : X \dashrightarrow Y$. We assume here that X and Y are complete manifolds and the functions $\varphi(x)$ and $\psi(y)$ decay on X and Y for x and y going to infinity. Furthermore, we assume that $\log \varphi$ and $\log \psi$ are *Lipschitz* which signifies the bound on $\varphi(x_1)/\varphi(x_2)$ in terms of dist

(x_1, x_2) and a similar bound for ψ. Finally, we suppose that φ is equivalent to $\psi \circ f$, which means a uniform bound on $\varphi(x)/\psi(y)$ and on $\psi(y)/\varphi(x)$ for all $(x, y) \in X \times Y$ in the graph of f. Notice that all these conditions are satisfied for φ and ψ of the form $(1 + \text{dist})^\alpha$, $\alpha < 0$, as in $3.5.C_2'$.

3.5.D'. Proposition. *Under the above assumptions* f *induces a boundary homeomorphism*

$$\partial_\varphi X \to \partial_\psi Y$$

which is, moreover, bi-Lipschitz.

Proof. A straightforward argument shows that our f satisfies the following version of the Lipschitz condition for the metrics $\varphi^2 g$ and $\psi^2 h$ which is somewhat better that the condition (*) in 3.2.B'.

There exists a (Lipschitz) constant $A = A(f, \varphi, \psi)$ *and a function* $b(x_1, x_2)$ *which decays for* $x_1, x_2 \to \infty$ *and* $\text{dist}_{\varphi^2 g}(x_1, x_2) \to 0$ *such that*

$$A^{-1}\text{dist}_{\varphi^2 g}(x_1, x_2) - b \le \text{dist}_{\psi^2 h}(y_1, y_2) \le A\ \text{dist}_{\varphi^2 g}(x_1, x_2) + b \qquad (+)$$

for all (x_1, y_1) *and* (x_2, y_2) *in the graph* $\Gamma r_f \subset X \times Y$.

Notice that the improvement here compared to (*) in 3.2.B' is the decay of b, that is

$$b \le \delta(r, \varepsilon)$$

for $r = \text{dist}_g(x_1, x_o)$ and $\varepsilon = \text{dist}_{\varphi^2 g}(x_1, x_2)$, where $\delta \to 0$ for $r \to \infty$ and $\varepsilon \to 0$. This improvement is mainly due to the decay of φ and ψ.

Now one sees with (+) that f matches the $\varphi^2 g$-Cauchy convergent sequences in X with $\psi^2 h$-Cauchy convergent ones in Y in-so-far as these sequences diverge in the original complete metrics. This matching is exactly what we want,

$$\partial_\varphi X \leftrightarrow \partial_\psi Y .$$

(We suggest to the reader to go through the details more carefully than we have done here).

$3.5.D_1'$. Example. If φ and ψ are of the type $(1 + \text{dist})^\alpha$ as in $3.4.C_2'$ then an arbitrary quasi-isometry $f: X \to Y$ induces a homeomorphism $\partial_\varphi X \geq \partial_\psi Y$.

3.5.E. <u>The case</u> $K \leq -\kappa^2 < 0$.

Our discussion on ∂_φ was so far quite general and rather trivial. Now, we assume once again that X is a complete simply connected manifold with $K(X) \leq 0$. We fix a point $x_0 \in X$ and we denote by P_r the *radial projection* of the *exterior* $E(r) \subset X$ of the r-ball, that is

$$E(r) = \{x \in X \mid \mathrm{dist}(x, x_0) \geq r\},$$

to the boundary sphere

$$S(r) = \partial E(r) = \{x \in X \mid \mathrm{dist}\ (x, x_0 \geq r\},$$

Recall that P_r equals the unique intersection point of the geodesic segment $[x_0, x] \subset X$ with $S(r)$.

The convexity of the distance function between (points on) rays issuing from x_0 implies that P_r is a contracting map, that is the differential D of P_r has $\|D\| \leq 1$.

At this point we make a stronger assumption on X, namely, we assume the sectional curvature $K(X)$ is *strictly negative* that is $K(X) \leq -\kappa^2$ where $\kappa > 0$.

For example, X may be a non-compact symmetric space of rank one.

The only feature of $K \leq -\kappa^2$ we need is the following basic bound on the norm of the above D at all points $x \in X$ *outside the sphere* S(r)

$$\|D_x\| \leq \exp - \kappa(r_x - r) \qquad \boxtimes$$

where $r_x = \mathrm{dist}\ (x, x_0)$ (see [Ch-Eb] for the proof).

<u>Remarks</u> (a) The exponential contraction is usually expressed as the *exponential growth* of the Jacobi fields along the rays issuing from x_0.

(b) We shall later need $\boxtimes$ only for a single value r, say for $r = 1$.

(c) Instead of P we could equally use the projection G to the *tangent* sphere at x_0 as we did in $3.5.B_1$. We prefer here P as it is more geometric and admits a generalization to *singular* spaces of negative curvature (compare [Gr]4).

3.5.E_1. We want to compare the boundary ∂X constructed with geodesic rays (see 3.5) and the conformal boundary $\partial_\varphi X$. If $\varphi(X) \leq r_x^{-1-\varepsilon}$, $\varepsilon > 0$ for large r_x, then every geodesic ray R has a finite $\varphi^2 g$-length, that is $\int_R \varphi(r)\, dr < \infty$, and so it defines a point s at $\partial_\varphi X$. If, furthermore, $\|\mathrm{grad}\ \varphi(x)\| \to 0$ for $x \to \infty$, then this s does not change if we take an (equivalent) ray lying within finite Hausdorff distance from R. In this case we obtain a natural map $\partial X \to \partial_\varphi X$ which is clearly, continuous. But unfortunately, this map may easily be constant which is not, of course, what we want.

Yet for $K \leq -\kappa^2 < 0$, we are presented with the following surprise

3.5.E_1'. Lemma (See [Fl]). *If* $\varphi(x)$ *does not decay faster than* $\exp - \kappa r_x$ *but yet decays at least as* $r_x^{-1-\varepsilon}$, *then the map* $\partial X \to \partial_\varphi X$ *is a homeomorphism.*

Proof. Every point $s \in \partial_\varphi X$ can be arrived at by an infinite curve $x(t)$ in X, for $t \in [0, \infty)$, of finite $\varphi^2 g$-length. This means $L_\varphi = \int_0^\infty \varphi(x(t))\, d\ell_g < \infty$ and $\mathrm{dist}_{\varphi^2 g}(s, x(t)) \to 0$ for $t \to \infty$. Given such $x(t)$ we consider the geodesic segments $[x_0, x(t)] \subset X$ and we observe that the decay bound

$$\varphi(x) \geq \mathrm{const} \exp - \kappa r_x$$

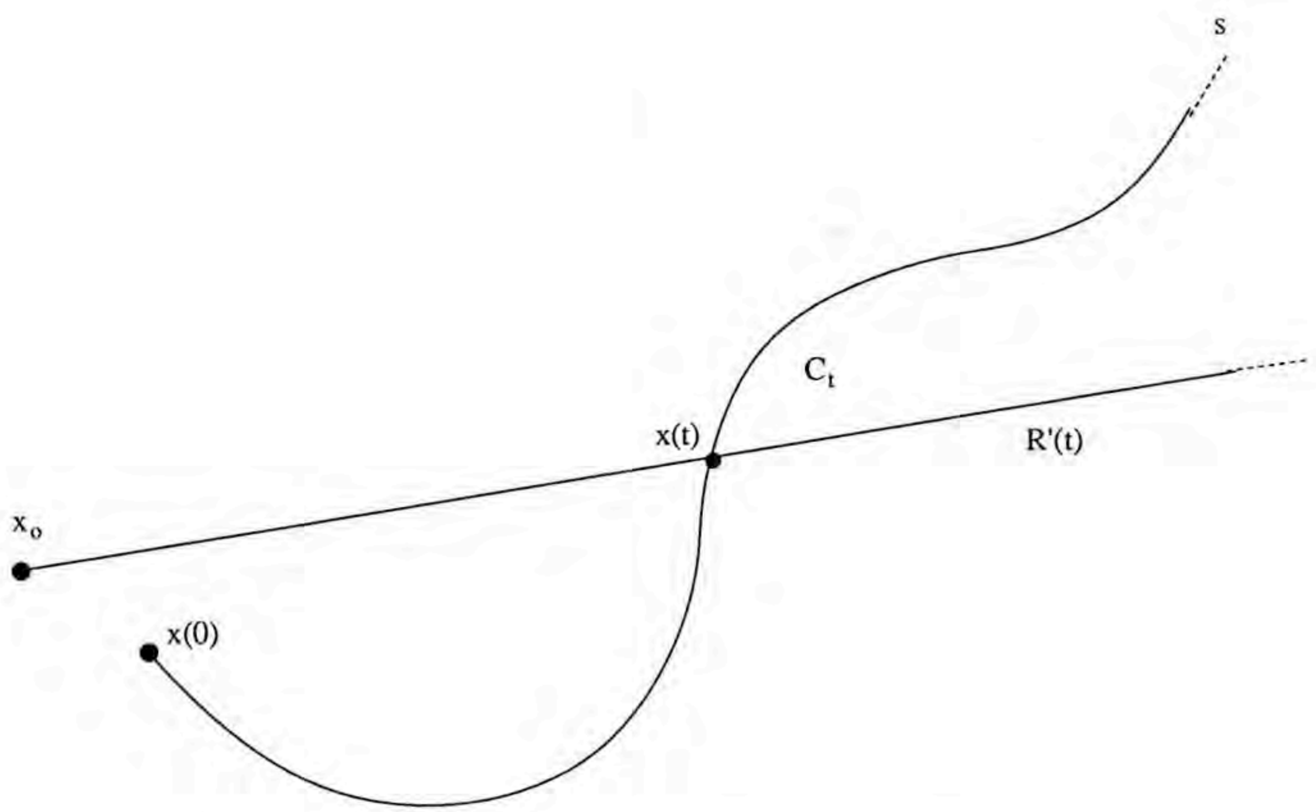
gives the bound on the length of the radial projection $P(t)$ of $x(t)$ to the unit sphere $S(1) \subset X$ by

$$\mathrm{length}\ P[0, \infty) \leq \mathrm{const}\ L_\varphi < \infty, \qquad (*)$$

where we assume without loss of generality that $x(t)$ lies in the exterior $E(1) \subset X$ of $S(1)$, which insures the contraction with the rate $\exp - \kappa\, \mathrm{dist}(x_0, x(t))$. The inequality $(*)$ implies the convergence of $P(t)$ to some point in $S(1)$ for $t \to \infty$ and thus the convergence of the rays R_t extending $[x_0, x(t)]$ to a ray R issuing from x_0.

Each R_t has finite $\varphi^2 g$-length and thus defines a point $s_t \in \partial_\varphi X$. This point is joined with s by a double infinite curve C_t which consists of the subray $R_t' \subset R_t$ starting from the point

$x(t) \in R_t$ and the part $x[t, \infty)$ of the curve $x[0, \infty) \subset X$,



Clearly, $\text{length}_{\varphi^2 g} C_t \to 0$ for $t \to \infty$ and so the limit ray R defines the point $s \in \partial_\varphi X$ we have started with. Therefore the map $\partial X \to \partial_\varphi X$ is surjective.

In order to prove that this map cannot bring together two points represented by different rays R_1 and R_2 issuing from x_o we must show that the $\varphi^2 g$-distance $\text{dist}_{\varphi^2 g}(r_1, r_2)$ is bounded away from zero for the points $r_1 \in R_1$ and $r_2 \in R_2$ going to infinity. In fact, let C be an arbitrary curve in X between r_1 and r_2. If C meets the unit sphere S(1) at x_o, then clearly $\text{length}_{\varphi^2 g} C \geq \text{dist}_{\varphi^2 g}(S(1), S(2)) > 0$, for r_1 and r_2 outside S(2). But if C lies outside S(1) the projection of C to S(1) has

$$\text{length } P(C) \geq \text{dist}(p_1, p_2) > 0$$

for $p_1 = S(1) \cap R_1$ and $p_2 = S(1) \cap R_2$, and so

$$\text{length } C \geq (\text{const})^{-1} \text{ length } P(C) > 0$$

as well (compare (*) above).

This concludes the proof of the bijectivity of the map of the geodesic ray boundary ∂X to the conformal boundary $\partial_\varphi X$. Finally, to prove this map is a homeomorphism, one can either invoke Brouwer's principle or check carefully all steps of the proof.

3.5.E_2. Corollary. *Let* X *and* Y *be complete simply connected manifolds with* $K \le -\kappa^2 < 0$. *Then every quasiisometry* $X \dashrightarrow Y$ *entends to a homeomorphism between the geodesic ray boundaries,* $\partial X \to \partial Y$.

Proof. Choose appropriate functions φ and ψ on X and Y, for example $(1 + \text{dist})^\alpha$ with $\alpha = -2$ for the distance functions from fixed points in X and in Y (compare 3.5. C_2'). Then, by the above lemma $\partial_\varphi X = \partial X$ and $\partial_\psi Y = Y$ and then 3.5. D_1'. applies.

Remarks. (a) The above proof is essentially due to Floyd, but the existence of the boundary homeomorphism was discovered earlier by Efremowitz and Tihomirova (see [Ef - Ti]). This homeomorphism also appears in $[\text{Mos}]_1$ and $[\text{Mar}]_3$ where it is used is the first step in the proof of the Mostow rigidity for lattices $\Gamma \subset \text{Iso}\, X$.

(b) An immediate consequence of the boundary homeomorphism is the invariance of the dimension under quasi-isometries. That is dim X = dim Y as the boundaries ∂X and ∂Y are spheres of the dimensions $\dim X - 1$ and $\dim Y - 1$ respectively. Notice that in general (without the assumptions $K < 0$ etc) quasi-isometries may easily change dimension. For example if K is an arbitrary compact manifold, then $X \times K$ is quasi-isometric to X.

(c) The topology of ∂X is not *the only* quasi-isometry invariant of X. For example the real and the complex hyperbolic space $H^{2n}_{\mathbb{R}}$ and $H^{2n}_{\mathbb{C}}$ have homeomorphic boundaries, namely S^{2n-1}, but they are *not quasi-isometric* (compare 3.6.).

(d) The notion of the geodesic ideal boundary ∂X makes sense if X has ordinary boundary, provided this boundary is convex and so the pairs of points in X can be joined by geodesic segments. In fact most of our discussion extends to this case. However, the boundary ∂X in the general case is not homeomorphic to a sphere but may be an arbitrarily complicated finite dimensional compact space. (It is homeomorphic to the projective limit of the spheres $S(i) \subset X$, $i \to \infty$ for the normal projections $S(i + 1) \to S(i)$). Moreover, the theory of $K < 0$ extends to a large class of (singular) hyperbolic spaces (see $[Gr]_4$).

3.5.F. The action of the group $\overline{QIs}$ on ∂X. It is immediate with the *proof* of 3.5.D' *and* 3.5.E_2 that the group $\overline{QIs}$ acts on ∂X by homeomorphisms. To see this we must check that the homeomorphism on ∂X induced by a quasi-isometry $f : X \dashrightarrow X$ depends only on the equivalence class of f under the relation $|f - f'| < \infty$. But this is obvious as $\partial X = \partial_\varphi X$ with a function $\varphi(x) \to 0$ for $x \to 0$ and so a bounded perturbation f' of f becomes asymptotically zero in the conformal metric $\varphi^2 g$. (Of course all this is built into the proof of 3.5.D'). One also must check that the resulting map

$$\overline{QIs}\ X \to \text{Homeo}\ \partial X$$

is a homomorphism, but this is again clear from the functionality of the extension of f to $\partial_\varphi X$ (see 3.5.D').

A somewhat less obvious fact here reads.

3.5.F'. *The homomorphism* $\overline{QIs}\, X \to$ Homeo ∂X *is injective*.

Proof. We need here the following remarkable property of complete simply connected spaces X with $K(X) \leq -\kappa^2 < 0$.

3.5.F_1'. Morse Lemma. *Let* $Y \subset X$ *be a subset which is quasi-isometric to* $\mathbb{R}$. *Then there is a unique geodesic* Y' *in* X *(which is isometric to* $\mathbb{R}$*!), such that the Hausdorff distance between* X *and* Y' *is finite. Moreover, this distance* δ *is bounded in terms of the constants (A and B in 3.2.B'.) involved in the definition of quasiisometry between* Y *and* $\mathbb{R}$.

The proof can be derived by a direct geometric argument from the contraction property in 3.5.E, but this is not completely trivial. Various versions of the proof can be found in [Mors], [Bus], [Klin], $[Mos]_4$, $[Gr]_4$ etc ...

3.5.F_2'. Corollary. *Let* f *be a quasi-isometry of* X *which fixes the ideal ends* y_+ *and* y_- *in* ∂X *of a geodesic* Y_0 *in X. Then all points* $y \in Y_0$ *satisfy*

$$\text{dist}(f(y), Y_0) \leq \delta < \infty$$

where δ *is bounded in terms of the quasi-isometry constants* (A and B) *of* f.

Proof. The image $f(Y_0)$ is quasiisometric to $Y_0 = \mathbb{R}$ and so there exists a unique geodesic Y' within finite Hausdorff distance δ from $f(Y_0)$. Since f (extended to ∂X) fixes y_+ and y_-, this Y' has y_+ and y_-, as its own ends. It follows (by the convexity of the distance between Y_0 and Y') that Y_0 and Y' have finite Hausdorff distance and by the uniqueness claim of $3.5.F_1'$ the two are, in fact, equal. Q.E.D.

Now we can prove the injectivity of $\overline{QIs} \to$ Homeo. Take two mutually orthogonal geodesics Y_0 and Y_1 passing through a given point $x \in X$. If a quasi-isometry f fixes ∂X it fixes, in particular, the four ends of our geodesics and so the f-image of x is contained in the intersection I of the δ-neighbourhoods of Y_0 and Y_1. It is easy to see that I is contained in the 2δ-ball around x, and so

$$\text{dist}(x, f(x)) \le 2\delta$$

for all $x \in X$. That is, $f \sim \text{Id}$ as we claimed.

3.5.G. Remark and exercices. (a) The Morse Lemma remains valid for subsets Y quasi-isometric to segments [a, b] and also to $\mathbb{R}_+$ but of course, there is no uniqueness in the case of $\mathbb{R}_+$.

(b) The Morse Lemma for rays gives another proof of the extension of a quasi-isometry $f : X \to Y$ to $\partial X \to \partial Y$. In fact the f-image of every ray R can be δ-approximated by a ray in Y thus giving the corresponding point in ∂Y.

(c) The convergence of $f(R) \subset Y$ to some point in ∂Y can be seen directly without Morse Lemma. For example, if f is a bi-Lipschitz map, then the radial projection of f(R) to the unit sphere S(1) in Y around a fixed point $y_0 \in Y$ has finite length. This immediately implies, (as we have seen already several times) the convergence of this projection to some point $y_1 \in S(1)$ and thus the convergence of f(R) to the point $y_\infty \in \partial X$ represented by the ray extending the segment $[y_0, y_1] \subset Y$.

Notice that the above arguments work for maps more general than quasi-isometries. For example if f is Lipschitz, than the quasi-isometry bound

$$\text{dist}(y_0, f(x)) \ge \text{const dist}(x_0, x)$$

can be easily relaxed to

$$\text{dist}(y_0, f(x)) \ge \text{const dist}(x_0, x)^\alpha, \ \alpha > 0.$$

3.6. <u>Carnot-Mostow metric on the sphere at infinity</u>. Let again X be a complete simply connected manifold with $K \leq 0$ and let S denote the unit tangent sphere at some point $x_0 \in X$.

We consider the radial map of S onto the sphere $S(R) \subset X$ of radius R and we denote by g_R the Riemannian metric on S induced from X. Recall that if $K < -\kappa < 0$, then g_R exponentially grows for $R \to \infty$. In particular

$$\mathrm{Diam}(S, g_R) \geq C^R,$$

for some $C > 1$ and all $R \geq 1$.

Now we denote by di_R the distance function on S corresponding to g_R and normalized by the condition

$$\mathrm{Diam}(S, di_R) = \pi.$$

That is $di_R(s_1, s_2)$ is the g_R length of the shortest path in S between s_1 and s_2 divided by $\pi^{-1}\,\mathrm{Diam}_R\,(S, g_R)$.

In general, the family of the functions $di_R : S \times S \to \mathbb{R}$ may diverge for $R \to +\infty$. However one knows that di_R converges in the case where X is a symmetric space. For example, if rank$X \geq 2$, then the limit metric di_∞ at the sphere $S = \partial X$ equals a positive constant times the Tits metric on ∂X, see [B-G-S]. The Tits metric on ∂X does not depend on the choice of the reference point x^0 and so the action of Iso X on ∂X is di_∞-isometric for rank$X \geq 2$. The simplest case is that of $X = \mathbb{R}^n$, $n \geq 2$ where (S, di_∞) equals the unit Euclidean sphere S^{n-1} and the group Iso $\mathbb{R}^n$ acts on S^{n-1} via the homomorphism Iso $\mathbb{R}^n \to O(n-1)$.

Now let us describe di_∞ for the symmetric spaces of rank 1.

3.6.A. <u>The space $H^n_{\mathbb{R}}$ of constant curvature</u> -1. Notice that the isotropy subgroup $\mathrm{Iso}_{x_0} \subset$ Iso $\mathbb{H}^n_{\mathbb{R}}$ equals $O(n-1)$ which isometrically acts on the spheres $S(R)$. It follows that $g_R = c(R)\, g_0$ for the unit spherical metric g_0 on S. Therefore di_R equals the distance di_{g_0} associated to g_0 for all $R > 0$ and so $di_\infty = di_{g_0}$ as well. Notice that the conformal factor c(R) equals

$$(\sinh R)^2 = \left(\tfrac{1}{2}(e^R + e^{-R})\right)^2$$

If we change the reference point x_0 then the metric di_∞ also changes and so the action of Iso $\mathbb{H}^n_{\mathbb{R}}$ on (S, di_∞) is *not isometric* . In fact, since this action is faithful (this follows from 3.3.4') and the group $\mathbb{H}^n_{\mathbb{R}} = PO(n, 1)$ is non-compact the boundary action of Iso $\mathbb{H}^n_{\mathbb{R}}$ can not be isometric for any metric on $S = \partial\, \mathbb{H}^n_{\mathbb{R}}$. On the other hand Iso $\mathbb{H}^n_{\mathbb{R}}$ acts on (S, di_∞) by *conformal* transformations. In fact, as everybody knows,

$$\mathrm{Iso}H^n_{\mathbb{R}} = PO(n, 1) = \mathrm{Conf}\, S^{n-1}.$$

3.6.A_1. <u>Comparison between the Carnot-Mostow metric and the φ-metric</u>. Let φ be a positive function on $X = H^n_{\mathbb{R}}$ which depends only on dist (x, x_0) (e.g. $\varphi = (\mathrm{dist}(x, x_0) + 1)^{-2}$ as in 3.5.D_1'.) and let us look at the corresponding φ-metric d_φ on ∂X (compare 3.5.C_1.).

Then the $O(n-1)$-symmetry shows that d_φ has the same balls in $\partial X = S^{n-1}$ as the standard metric di_{g_0} . In other words

$$d_\varphi = \Phi(di_{g_0})$$

for some function $\Phi(d)$ depending on φ. If we normalize d_φ by the condition diam $(\partial X, d_\varphi) = \pi$, then, necessarily

$$d_\varphi \le (di_{g_0})$$

because di_{g_0} (unlike d_φ) is a *length* metric associated to a length function on curves. The length property implies (in fact is equivalent to the $O(n-1)$-invariant situation as ours) that di_{g_0} is the *maximal* metric on S^{n-1} which is $O(n-1)$-invariant and having diam = π.

If the function φ in question has polynomial decay (i.e. $\mathrm{dist}(x_0, x)^\alpha$) then one can show that Φ (d) is of order $|\log d|^{-1}$.

In this situation the Lipschitz property of maps for the φ-metric (established in

3.5.D'.) does not tell much about the di-metric. On the other hand if φ has exponential decay (exp -2 dist) then d_φ and di are Hölder equivalent,

$$d_\varphi \sim (\mathrm{di})^\beta, \ \beta \leq 1$$

as one can prove with a little effort.

We shall explain later on (following Margulis [Mar]$_3$) how the Morse lemma (see 3.5. F_1'.) yields di-*quasiconformality* of the boundary map on ∂X induced by a quasi-isometry of X (alternatively, one could first prove the d_φ-quasiconformality for $\varphi \sim \exp{-\lambda}$ dist and then derive the di-quasiconformality). This quasi-conformality plays the crucial role in Mostow's approach to the rigidity for $H^n_{\mathbb{R}}$ as well as for the other hyperbolic spaces $H^{2n}_{\mathbb{C}}$, $H^{4n}_{\mathbb{H}}$ and $H^{16n}_{\mathbb{C}a}$.

3.6.B. The complex hyperbolic space $H^{2n}_{\mathbb{C}}$. The ideal boundary here is S^{2n-1} which we think of as the unit sphere in $\mathbb{C}^n$. Then there is a natural orthogonal splitting of the tangent bundle.

$$T(S^{2n-1}) = T' \oplus T''$$

where T' is tangent to the Hopf fibers for the S^1-bundle $S^{2n-1} \to \mathbb{C}P^{n-1}$ and T'' is the codimension one subbundle defined with the (standard) complex structure J on the tangent bundle T of $\mathbb{C}^n$ restricted to $S^{2n-1} \subset \mathbb{C}^n$ as follows :

$$T'' = T(S^{2n-1}) \cap J\, T(S^{2n-1}),$$

where $T(S^{2n-1})$ of S^{2n-1} is embedded into T in the obvious way. In other words T'' is the maximal *complex* subbundle in $T(S^{2n-1}) \subset T$.

The splitting $T' \oplus T''$ induces a splitting of the Riemannian metric g_o on S^{2n-1}, that is

$$g_o = g_o' + g_o'',$$

where g_o' lives on T' (i.e. vanishes on T'') and g_o'' lives on T''. Now we consider the family $g_\lambda = \lambda g_o' + g_o''$ and try to see what happens for $\lambda \to \infty$.

3.6.B_1. Proposition. *The family of distance* function di_{g_λ} *converges to a certain metric* di *on* S^{2n-1}.

Proof. As metrics di_{g_λ} increase in λ one only has to check they remain bounded, i.e. that

$$\mathrm{Diam}(S^{2n-1}, \mathrm{di}_{g_\lambda}) \leq \mathrm{const}$$

for $\lambda \to \infty$. To prove this it suffices to show that every two points in S^{2n-1} can be joined by a curve C *tangent to* T" (i.e. T(C) ⊂ T") and having length C ≤ const. In fact, such a curve has $\mathrm{length}_{g_\lambda} C = \mathrm{length}_{g_o} C$ which implies the bound on Diam.

Now an elementary argument shows that every two points in S^{2n-1} can be joined by a "broken geodesic", that is a piece-wise smooth curve consisting of at most three circular arcs tangent to T". This concludes the proof of the proposition.

3.6.B_1. Remarks (a) It is not hard to see that the limit distance di equals the length of the shortest path tangent to T" between two given points. Metrics of this kind have been discovered by Carnot and Caratheodory (in the framework of the thermodynamical formalism) and have been intensively studied under various names (see [Str]).

(b) The above metric di on S^{2n-1} is highly symmetric. Namely, the natural action of U(n) is di-isometric. Furthermore the boundary action of $U(n, 1) = \mathrm{Iso}(H^{2n}_{\mathbb{C}})$ on $S^{2n-1} = \partial_\infty H^{2n}_{\mathbb{C}}$ is di-conformal as we shall see later on.

3.6.B_2. Now let us look at the spheres $S(R) \subset H^{2n}_{\mathbb{C}}$ around some point $x_o \in H^{2n}_{\mathbb{C}}$ where the metric in $H^{2n}_{\mathbb{C}}$ is normalized in such a way that the sectional curvature K has inf K = −1. Then the induced metric g_R on $S(R) = S^{2n-1} \subset T_{x_o}(H^{2n}_{\mathbb{C}}) = \mathbb{C}^n$ is given by the formula

$$g_R = e'(R)\, g'_o + e''(R)\, g''_o,$$

where

$$e'(R) = (\sinh r)^2$$

$$e''(R) = (2\sinh r/2)^2$$

To prove this formula for e' we observe that the exponential image of every complex line in

$\mathbb{C}^n \supset S^{2n-1}$ (whose intersection with S^{2n-1} is the Hopf fiber where g_o' lives) is a totally geodesic plane (complex geodesic) in $H^{2n}_{\mathbb{C}}$ with constant curvature -1 (this is easy to prove, compare ([Mos]$_4$) and the e'-formula follows from the corresponding formula for H^2 of curvature -1.

Now, to understand e", we look at the 2–planes τ in $\mathbb{C}^n = T_{x_o}(H^{2n}_{\mathbb{C}})$ which are *normal* to the Hopf fibers, i.e. $\tau \cap S^{2n-1}$ is tangent to T". Notice that every plane in the real locus $\mathbb{R}^n \subset \mathbb{C}^n$ has this property and, on the other hand, every τ may be brought to $\mathbb{R}^n$ by a unitary transformation of $\mathbb{C}^n$. What is relevant for us in this picture is the existence of τ containing a given vector in $T'' \subset T(S^{2n-1})$. (This follows, for example, from the transitivity of the U(n)-action on such vectors).

Next we recall that the exponential image of $\mathbb{R}^n \subset \mathbb{C}^n = T_{x_o}(H^{2n}_{\mathbb{C}})$ is a totally geodesic submanifold isometric to H^n with curvature $-\frac{1}{4}$ (this is also easy and can be found in [Mos]$_4$). It follows that the metric grows along each τ according to the e"-formula and this formula applies to T" as all vectors there are covered by such planes τ.

3.6.B_1'. Corollary. *The limit (Carnot-Mostow) metric*

$$di_\infty =_{def} \lim \pi\, di_R / \text{Diam}\, di_R$$

equals the above metric di *on* S^{2n-1} *times the normalizing constant* $= \pi/\text{Diam}\, di$.

3.6.C. The spaces $H^{4n}_{\mathbb{H}}$ and $H^{16}_{\mathbb{C}a}$. We use again the Hopf fibration of the sphere at infinity that are $S^{4n-1} \to \mathbb{H}P^{n-1}$ and $S^{15} \to S^8 = \mathbb{C}aP^1$ and we use the *normal* splitting of the tangent bundle of the sphere as in the complex space, that is $T' \oplus T''$ where T' is tangent to the Hopf fibers. Then we blow up the T'-component of the standard metric in the sphere and arrive as earlier at the metric di. Then it is easy to identify di with the (renormalized) metric di_∞ at the sphere at infinity.

To get some feeling about $di = di_\infty$ we mention some properties of this metric.

(a) The Hausdorff dimension of the sphere S with di satisfies (see [Pan]$_2$).

$$\text{Dim}_{\text{Haus}} = \dim_{\text{Top}} S + \text{codim}\, T''.$$

Thus

$$\dim_{Haus} \partial_\infty(H^n_R) = n - 1$$

$$\dim_{Haus} \partial_\infty(H^{2n}_{\mathbb{C}}) = 2n$$

$$\dim_{Haus} \partial_\infty(H^{4n}_{\mathbb{C}}) = 4n + 2$$

$$\dim_{Haus} \partial_\infty(H^{16}_{\mathbb{C}}) = 22$$

(b) Let $f: S \to S$ be a C^1-smooth di-Lipschitz map. Then Df maps T" to T" as only the curves *tangent* to T" may have *finite* di-length. This gives no restriction on f for $H^n_{\mathbb{R}}$. In the case of $H^{2n}_{\mathbb{C}}$ the maps of S^{2n-1} preserving the (codimension one) subbundle T" are known under the name of *contact* maps. Such maps form an infinite dimensional Lie group and every contact map is (obviously) di-bi-Lipschitz.

The situation drastically changes for codim T" ≥ 3. Here an old theorem of Cartan says that every *sufficiently smooth* map $S \to S$ preserving T" comes from an isometry of the corresponding hyperbolic space ($H^4_{\mathbb{H}}$ or $H^{16}_{\mathbb{C}a}$). In particular the automorphism group of (S, T") is a (finite dimensional) Lie group. We shall see later on that this conclusion remains valid for all (not necessary smooth) Lipschitz maps and even for di-*quasiconformal* maps.

3.7. <u>Conformal and quasi-conformal maps between metric spaces</u>. Intuitively, a map $f: A \to B$ is *conformal* if it sends "infinitely small balls" in A to such balls in B. To make it precise we introduce the following notion of *asphericity* of a family $\{\mathcal{U}\}$ of neigbourhoods $\mathcal{U} \subset A$ of a fixed point $a \in A$. We denote by inrad $(\mathcal{U}, a)$ the infimum of the distance function dist (a, u) over $u \in \mathcal{U}$ and let outrad denote the supremum of dist(a, u). Then we set

$$\operatorname{asph}(\mathcal{U}, a) = \text{outrad}/\text{inrad}$$

and

$$\operatorname{asph}\{\mathcal{U}\} = \limsup_{\operatorname{Diam} \mathcal{U} \to 0} \operatorname{asph}(\mathcal{U}, a).$$

For example if all $\mathcal{U}$ are balls around a, then

$$\operatorname{asph}\{\mathcal{U}\} = 1.$$

(The number 1 here is the neutral element in the group $\mathbb{R}_+^\times$ which can be turned into 0 by taking log).

Next, for a continuous map $f : A \to B$ we define the (non)-conformality of f at a by

$$\mathrm{conf}_a f = \mathrm{asph}\,\{f(B(r))\}$$

for the balls $B(r) \subset A$ around a for $r \in (0, \infty)$. We call f *conformal* at a if it is conformal at all $a \in A$.

3.7.A. Basic example. Let S be the round sphere with the subbundle $T'' \subset T(S)$ as earlier and g'' be the T''-component of the spherical metric. With this g'' we have our Carnot-Caratheodory-Mostow metric di on S defined by the length of curves tangent to T''.

3.7.A'. Proposition. *A diffeomorphism* $f : S \to S$ *is di-conformal if and only if the differential* Df *sends* T'' *to* T'' *and the metric* g'' *on* T'' *goes to a (conformal) metric of the form* $\varphi^2 g''$ *on* T'' *for some positive funcion* φ *on* S.

Proof. This is obvious (and known to everybody) in the classical case codim $T'' = 0$ (i.e. $T'' = T(S)$ and so g'' is a Riemannian metric). In the general case the proof is equally easy once one understands the geometry of di-balls in S of radii $r \to 0$. A little thought shows that such a ball at $s \in S$ looks roughly as a small thickening of a d-dimensional (round) ball $\mathcal{D}(r) \subset S$ for $d = \mathrm{rank}\, T''$ which is tangent at its center to the space T''_s. In other words $\mathcal{D}(r)$ is obtained by exponentiation (for the spherical metric in S) of the r-ball in T''_s. Now, the small thickening $\mathcal{D}_+(r)$ of $\mathcal{D}(r)$ is the ε-neighbourhood of $\mathcal{D}(r)$ for $\varepsilon = d^2$. One can show (we suggest it to the reader) that $\mathrm{asph}\, \mathcal{D}_+(r) \to 1$ for $r \to 0$ and then the Proposition follows as in the classical case.

Remark. The above Proposition serves only as an illustration. In our applications of (quasi)-conformal maps to the rigidity we shall use a sligtly different family of quasi-spherical neighbourhoods (see below).

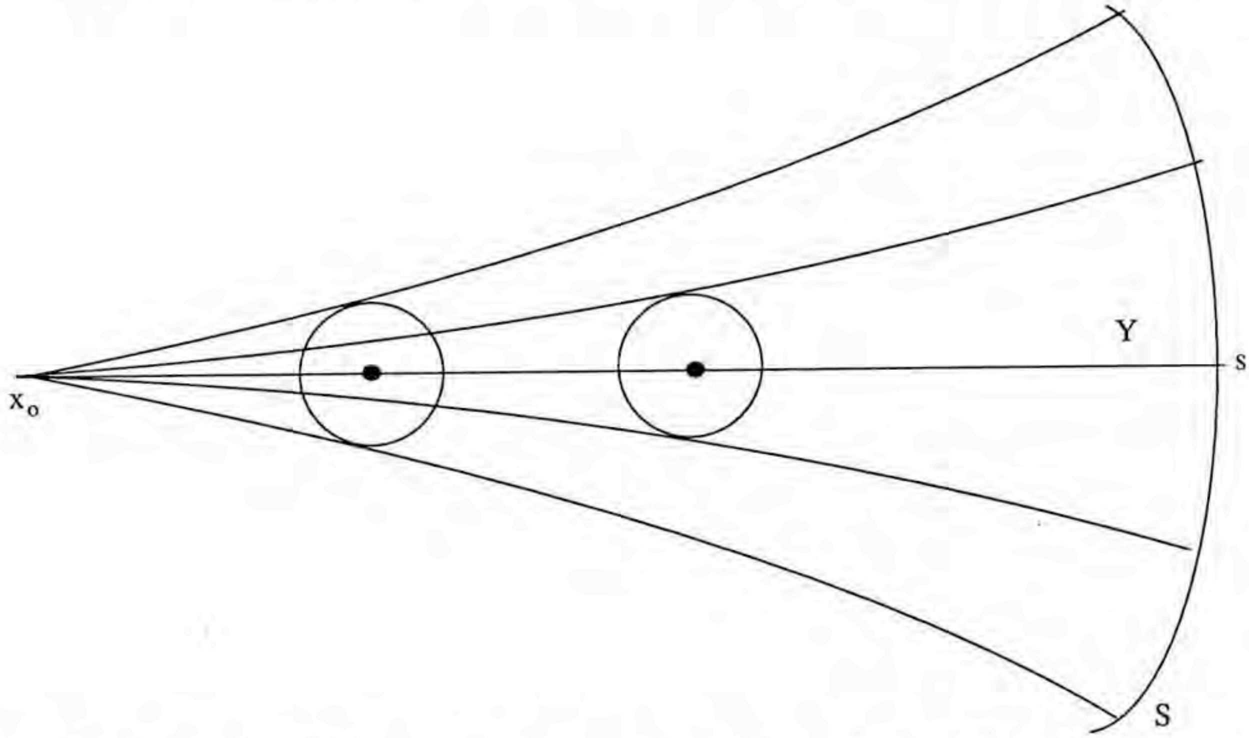
3.7.B. Quasiconformal maps. A continuous map $f : A \to B$ is called *quasi-conformal* if

$$\mathrm{Conf}_a f \leq \lambda < \infty$$

at all points $a \in A$. Sometimes one says f is *λ-quasiconformal* for $\lambda = \sup_{a \in A} \mathrm{conf}_a f$.

3.7.B' <u>Example</u>. Let us look again at maps $f : S \to S$. If such a map is a diffeomorphism then it is quasi-conformal if and only if T'' goes into itself. In fact, the constant λ equals the conformality of the differential Df on T'', i.e. $\sup_{s \in S} \mathrm{Conf}(Df|T''_a)$. The proof follows from that of Proposition 3.7.A'.

3.8. <u>(Quasi) conformal structure on the sphere at infinity</u>. Let us give a different definition of the (quasi)-conformal structure at $S = \partial_\infty X$ for an arbitrary complete simply connected manifold X with $K \le -\kappa < 0$. We fix a point $x_0 \in X$ and some real number $R_0 > 0$. Then we take a ray Y in X joining x_0 with some point $s \in S$ and we consider the R_0-balls $B(R_0) \subset X$ with centers $y \in Y$. Then we radially project these balls from x_0 to S.

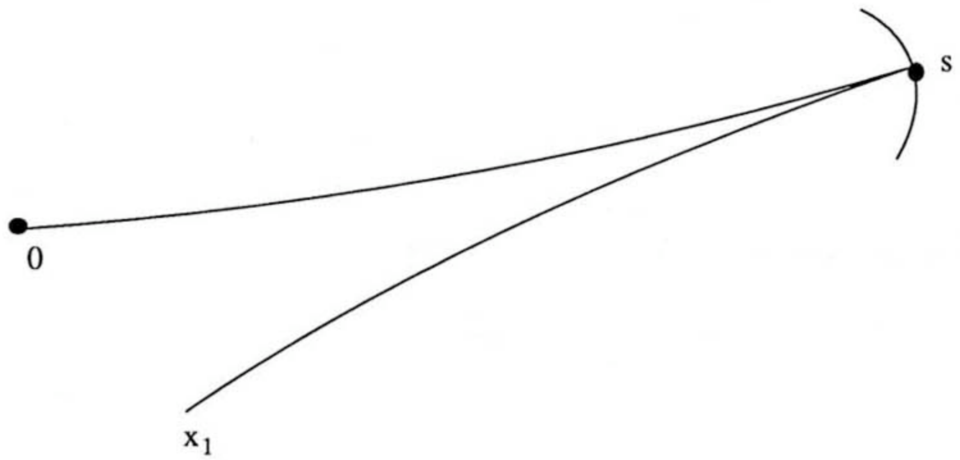


Following Margulis (see [Mar]3) we *call* these projections *balls* in S . More precisely, a Ma-r-*ball* , $r \in (0, 1)$ in S around s is defined as the projection of R_0-ball in X around the point $y \in Y$ with $\mathrm{dist}(x_0, y) = -\log r$. Using these Ma-r-balls one defines in an obvious way the asphericity of (small) neighbourhoods $U \subset S$ and then the rest of the conformal definitions follows.

Examples. If $X = H^n_{\mathbb{R}}$ then the projected balls *are* ordinary spherical balls in $\partial X = S^{n-1}$, but the spherical radius of a Ma-r-ball may be different from r. However,

one can easily see that the spherical radius of a Ma-r-ball is asymptotic (for $r \to 0$) to $C_0 r^\alpha$, where α depends on the curvature $K(H^n)$ (if the metric is normalized by $K = -1$, then $\alpha = 1$) and C_0 depends on the number R_0 in the Margulis definition. It immediately follows, that these Mar-balls lead to the same (quasi)-conformal geometry : a map $S \to S$ is (quasi) conformal in the usual sense if and only if it is such in Margulis' sense.

In principle, Margulis' structure may depend on x_0 and R_0. However , it does not depend on x_0 for R_0 being kept fixed : the identity map $S \to S$ is conformal for the Mar-conformal structures associated to different points x_0 and x_1 in X, provided the curvature of X satisfies $K(X) \le -\kappa < 0$. This follows from the *exponential convergence* at v infinity of the geodesic rays joining x_0 and x_1 with a point $s \in S$.



In fact, there is a correspondence between the points of rays, say $y_0 \in Y_0$ and $y_1 \in Y_1$, such that $\mathrm{dist}(y_0, y_1)$ (exponentially) decays as these points go to infinity (as they converge to s). Then the balls $B(R_0)$ around y_0 and y_1 become closer and closer at infinity. It means their projections to S, that are Ma-r-balls with $r \to 0$ also become very close. Namely the asphericity of the family of the Ma-r-balls projected from x_0 with the respect to (the conformal structure defined by) the balls projected from x_1 converges to 1 for $r \to 0$. This is immediate once the definitions are recalled.

Now, if we also change R_0 we may change the conformal structure, but the *quasi-conformal* structure does not change. That is the identity map $S \to S$ is quasi-conformal for the structures defined with two different R_0 and R_1. This immediately follows from the exponential divergence of rays.

A more profound quasi-conformality result (of Mostow and Margulis see [Mos]4, [Mae]3) concerns the boundary correspondence of a quasi-isometry $X_1 \to X_2$. Here both manifolds X_1 and X_2 are assumed complete simply connected with pinched negative curvature, $0 > -\kappa_1 \geq K \geq -\kappa_2 > -\infty$.

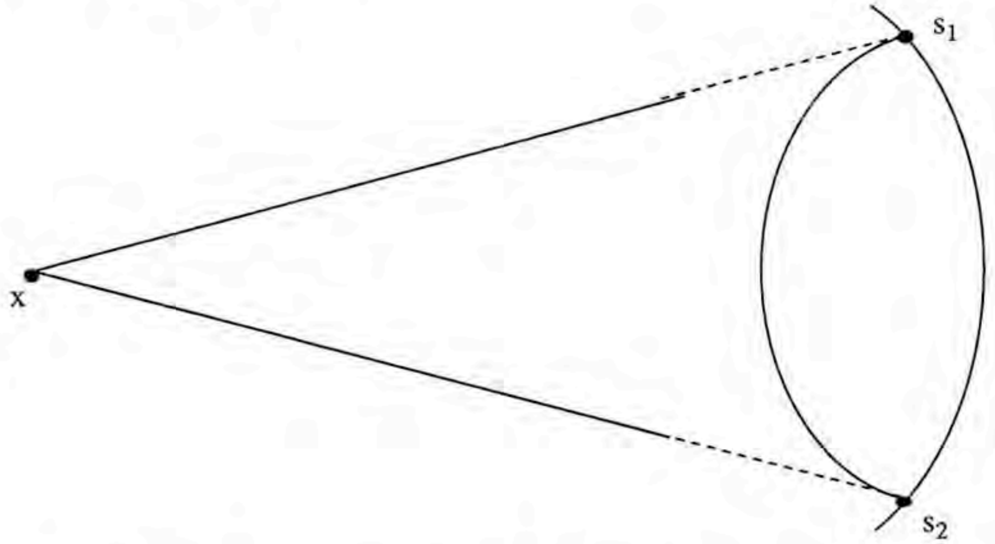
3.8.A. <u>Quasi-conformality theorem</u>. *The boundary map $\partial_\infty X_1 \to \partial_\infty X_2$ induced by a quasi-isometry $f : X_1 \to X_2$ (see 3.2.B'.) is quasiconformal for the Margulis conformal structures in X_1 and X_2.*

<u>Sketch of the proof</u>. We may assume the quasi-isometry f in question is full (see 3.3.). Then the images of large R-balls in X have bounded asphericity. Namely if $R > 10B$ for the constant B in the quasi-isometry definition (see(*) in 3.2.B'.). Then the asphericity of f(B) is at most 2A (for A from (*) in 3.2.B'.).

Now we use the Margulis structures with large R_0, we take a small Mar-ball B_1 in $\partial_\infty X_1$, which is the projection of an R_0-ball $B(R_0)$ in X lying far from $x_0 \in X$, and we want to estimate the asphericity of the image $\bar{B}_1 \subset \partial_\infty X_2$ under the boundary map. To do that we invoke the Morse lemma (see 3.5.F. and the discussion which follows) and observe that $\bar{B}_1$ is obtained by a *quasi-radial* projection of $f(B(R_0)) \subset X_2$ to $\partial_\infty X_2$, where "quasi-radial" refers to *quasi-rays* in X_2 which are subsets quasi-isometric to $\mathbb{R}_+$ and which are in our case the f-images of the rays in X_1 projecting $B(R_0)$ to $\partial_\infty X_1$. Denote by $\bar{B}_1'$ the (usual) radial projection of $f(B(R_0))$ to $\partial_\infty X_2$. Now we have *three* structures in $\partial_\infty X_2$. The first one is the quasiconformal structure defined with the Mar-balls B_2 in $\partial_\infty X_2$. Then we have the structure defined with the images $\bar{B}_1$ of the Mar-balls B_1 in $\partial_\infty X$. The third structure is given by the balls $\bar{B}_1'$. The quasi-conformal equivalence between B_2 and $\bar{B}_1'$ structures is obvious with the above asphericity bound for $f(B(R_0)) \subset X_2$. The structures defined with the "balls" $\bar{B}_1'$ and $\bar{B}_1$ are also quasi conformally equivalent but this is less obvious. In fact, it easily follows from the Morse Lemma for quasi-rays. Thus we have the equivalence of B_2-structure to the $\bar{B}_1$-structure which is the image of the Margulis structure in $\partial_\infty X_1$. Q.E.D.

Remark. The above argument is due to Margulis who generalized and simplified the original proof by Mostow in $[Mos]_1$. A detailed treatment of the classical hyperbolic spaces is given in $[Mos]_4$.

3.8.B. Let us indicate another more global point of view on the quasi-conformal structure on $\partial_\infty X$ which was much emphasized by D. Sullivan. We use the fact (see [B-G-S]) that every two points s_1 and s_2 in the sphere at infinity $\partial_\infty X$ can be joined by a unique geodesic in X denoted $(s_1, s_2) \subset X$. (This follows, for example, from Morse lemma as the union of two rays $(s_1 x]$ and $[x, s_2)$ for any $x \in X$ is quasi-geodesic (i.e. quasi-isometric to $\mathbb{R}$ and so can be approximated by a geodesic.



Our conformal structure on $S = \partial_\infty X$ is given by the following function Δ on $S \times S \times S \times S$, $\Delta(s_1, s_2 ; s_3, s_4) = \mathrm{dist}((s_1, s_2), (s_3, s_4))$ where the distance between subsets in X refers to the infimum of the distances between their points. Then a map $S \to S'$ is called quasi-conformal if there exist positive constants a and b such that the functions Δ and Δ' applied to all quadruples of corresponding points s_i and s_i', $i = 1, 2, 3, 4$, satisfy

$$a^{-1}\Delta - b \leq \Delta' \leq a\Delta + b \qquad (+)$$

It is obvious with the Morse lemma that the boundary map of every quasi-isometry is quasiconformal in (+) sense. Conversely, one can show that every (+)-quasi conformal map of the boundary is induced by a quasi-isometry of the underlying space. In other words one can reconstruct X up to quasi isometry from its boundary S. To do that one needs first of all a description of points $x \in X$ in terms of S. The idea is to reverse the map $x \mapsto T_x \subset S \times S$ relating to x the set of T_x the pairs of ends of the geodesic in X going through x. Before doing this we notice the following two properties of $T_x \subset S \times S$.

(i) the projection of T_x on the both components of the Cartesian product $S \times S$ are onto.

(ii) $\Delta(s_1, s_2 ; s_3, s_4) = 0$ whenever (s_1, s_2) and (s_3, s_4) lie in T_x.

Now, for a given number $\rho > 0$ we consider subsets $T \subset S \times S$ which satisfy (i) and the following modified version of (ii),

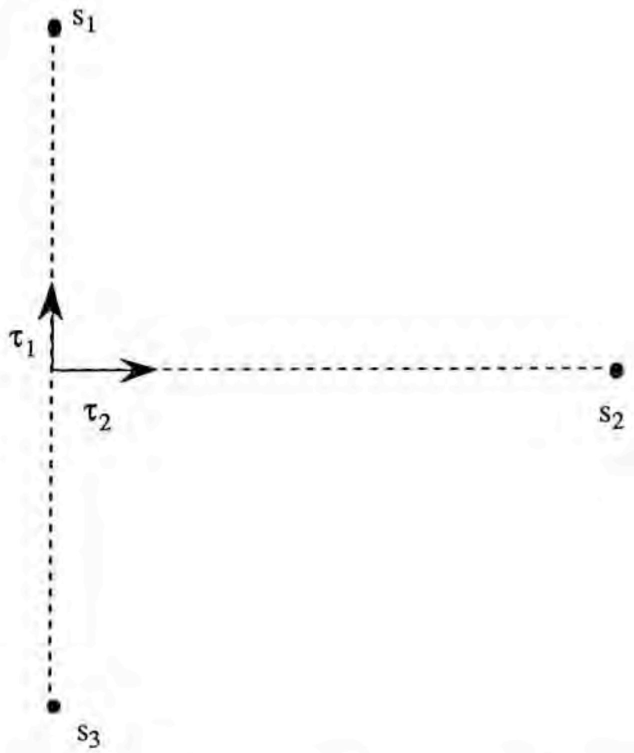
(ii)' $$\Delta(s_1, s_2 ; s_3', s_4') \leq \rho$$

whenever (s_1, s_2) and (s_3, s_4) lie in T. We call such subsets ρ-*points* and denote by X_ρ the set of all ρ-points $T \subset S \times S$. The distance between two ρ-points T and T' is defined by

$$\sup \Delta(s_1, s_2, s_1', s_2')$$

over all $(s_1, s_2) \in T$ and $(s_1', s_2') \in T'$. It is easy to see that X_ρ with this distance is quasi-isometric to X and that every Δ-quasiconformal map is quasi-isometric on X_ρ if ρ is large compared with the implied constants a and b.

Remark. There is an alternative way to reconstruct X from S based on Ahlfors-Cheeger homeomorphism between the Stiefel bundle Y of the pairs of orthonormal vectors in X and the space of triples of distinct points in X. This homeomorphism maps the pair (τ_1, τ_2) to the ends of the rays defined by τ_1, τ_2 and $-\tau_1$.



A non-trivial (but not difficult either) point here is bijectivity of this map. Using this we can immediately recapture Y from S and then we observe that the projection $Y \to X$ is a quasi-isometry for the metric structure in Y defined with Δ on S.

$3.8.B_1$. The quasi-conformal structure of Margulis can be seen using Δ as follows. Take a point s_o and fix two auxilary points s_1 and s_2 such that $s_o \neq s_1 \neq s_2 \neq s_o$.

Now let

$$\delta(s) = \Delta(s_o, s ; \ s_1, \ s_2),$$

and define Ma-r-balls for $r < 1$ by the condition

$$\{ s \in S \mid \delta(s) \geq - \log r \}$$

One can show with no effort at all that the quasiconformal type of this structure at x_o does not depend on x_1 and x_2. In fact, one can easily see that this structure is equivalent to the Margulis structure.

What looks non-trivial however is the following

Open Problem. *Let* S *be the ideal boundary of a complete simply connected manifold* X *with the curvature pinched between two negative constants.When can one reconstruct the* Δ-*quasiconformal structure from the Margulis structure ?. Equivalently when a Mar-quasi-conformal map* $S \to S$ *is induced by a quasi-isometry of* X ?

It was been known for many ($\approx$ 50) years that Mar $\Rightarrow \Delta$ for $X = H^n$ and $n \geq 3$. This is equivalent to the existence of *global* quasiconformal invariants (moduli) of subsets in S^{n-1} for $n - 1 \geq 2$. On the other hand the local quasiconformal structure on $S^1 = \partial_\infty H^2$ is not strong enough to reconstruct Δ but here one has a refined structure called *quasi-symmetric* (see below 3.10.C_3).

The next known case is that of $S = \partial_\infty H^{2n}_{\mathbb{C}}$ $n \geq 2$ where the global structure can be reconstructed from the local one by the work of M. Reimann [Rei] . A similar (even stronger) conclusion holds for the spaces $H^{4n}_{\mathbb{C}}$ and $H^{16}_{\mathbb{C}}$ (see [Pan]$_2$ and 3.11.A below) but for a somewhat different reason.

3.9. Idea of Mostow's proof of the rigidity for $H^n_{\mathbb{R}}$. We think here of $H^n_{\mathbb{R}}$ as open (Poincaré) ball

$$B^n \subset S^n \subset \mathbb{R}^{n+1}$$

which is formed by the rays in $\mathbb{R}^{n+1}$ on which the quadratic form $h = x_o^2 - \sum_{i=1}^{n} x_i^{\,2}$ is positive. Then the group O(n, 1) acts on B^n by conformal transformations (for the usual conformal structure in B^n). Since the isotropy subgroup Iso_x of each point $x \in B^n$ is compact (it is O(n)) there exists an O(n, 1) invariant metric g on B^n. This metric is unique up to a scaling constant since the action of Iso_x on $T_x(B^n)$ is irreducible. In fact the metric g turns B^n into the hyperbolic space $H^n_{\mathbb{R}}$ with constant negative curvature whose isometry group equals

$$PO(n, 1) = O(n, 1)/\{\pm 1\},$$

which is the same as the group of conformal automorphisms of the boundary S^{n-1} of B^n. (The ball B naturally embeds into $\mathbb{R}^{n+1}$ as each h-positive ray $b \subset \mathbb{R}^{n+1}$ contains a unique point x, such that $h(x, x) = 1$. The form h is positive on the tangent bundle of B embedded into $\mathbb{R}^{n+1}$ by $b \mapsto x$ and $h|T(B)$ can be taken for our g).

Notice (see the first example in 3.8) that the ordinary conformal structure on S^{n-1} is the same as the Margulis structure. In particular, every quasi-isometry $H^n_{\mathbb{R}} \to H^n_{\mathbb{R}}$ extends to a unique quasiconformal map $S^{n-1} \to S^{n-1}$ for the usual conformal structure of S^n. The basic property of such maps is the following.

3.9.A. Regularity Theorem. *Every quasi-conformal map* $f : S^{n-1} \to S^{n-1}$, $n - 1 \geq 2$, *is almost everywhere differentiable and the differential* Df *is quasiconformal in the following sense. There exists a constant* $\lambda > 0$ *(depending on* f*) such that for almost all* $s \in S^{n-1}$ *and all pairs of unit tangent vectors* τ_1 *and* τ_1 *in* $T_s(S)$ *the norms of their* Df *images satisfy*

$$\lambda^{-1}\|Df(\tau_1)\| \leq \|Df(\tau_2)\| \leq \lambda\|Df(\tau_1)\| .$$

The proof of that can be found in [Vai]. Also notice that the regularity fails for $n - 1 = 1$.

Another ingredient of Mostow's proof is the following special case of Mautner's ergodicity theorem (see [Mau]).

3.9.B. *Let* Γ *be a lattice in* $O(n, 1)$. *Then the action of* Γ *on the projectivised tangent space* $PT(S^{n-1})$ *is ergodic. That is every invariant measurable subset has measure zero or full measure.*

3.9.B'. Corollary. *Let* $f: S^{n-1} \to S^{n-1}$ *be a quasi-conformal map such that the action* $s \mapsto f\gamma f^{-1}(s)$ *of* Γ *on* S^{n-1} *is conformal. Then* f *is conformal.*

Proof : Define

$$\varphi(t) = \|Df(\tau)\| \,/\, \|Df \mid T_s(S^{n-1})\|$$

for all unit vectors $\tau \in T_s(S)$ and all $s \in S$. As $\varphi(\tau) = \varphi(-\tau)$ this defines a function, also called φ, on $PT(S^{n-1})$. This function clearly is Γ-invariant and by the ergodicity it is a.e. constant. It follows that Df is a.e. conformal, which implies (as one knows, see [Resh], [Geh]) that Df everywhere exists and is conformal. This means f is conformal. Q.E.D.

Now Mostow's proof runs as follows. We consider two compact n-dimensional manifolds V and V' with constant negative curvature and let $\Gamma \to \Gamma'$ be an isomorphism between their fundamental groups. This induces a quasi-isometry between their universal coverings $\tilde{V} \to \tilde{V}'$, both of which are isometric to $H^n_{\mathbb{R}}$, and this quasi-isometry induces a *quasi-conformal* (see 3.8.A.) homeomorphism f between the ideal boundaries of $\tilde{V}$ and $\tilde{V}'$ both of which are conformal to S^{n-1}. The functoriality of the boundary map implies that f carries over the action of Γ on $S^{n-1} = \partial_\infty \tilde{V}$ to the action of Γ' on $S^{n-1} = \partial_\infty \tilde{V}$. Since both actions are conformal, f is conformal and so it extends to a unique isometry between $\tilde{V}$ and $\tilde{V}'$. Since f agrees with the Γ and Γ' actions, same is true for our isometry which shows this isometry comes from an isometry between underlying compact manifolds V and V'. Q.E.D.

3.10. Rigity for $H^{2n}_{\mathbb{C}}$. One can identify $H^{2n}_{\mathbb{C}}$ with the open ball $B^{2n} \subset \mathbb{C}P^n$ which consists of those lines in $\mathbb{C}^{n+1}$ on which the Hermitian form $z_o \bar{z}_o - \sum_{i=1}^{n} z_i \bar{z}_i$ is positive. Then we have the natural holomorphic action of the group $PU(n, 1)$ on B^{2n} and this action admits a unique up to scale invariant Riemannian metric g which turns B^{2n} into $H^{2n}_{\mathbb{C}}$. The sectional curvatures of g between $-\frac{1}{4}\kappa$ and $-\kappa$ for some $\kappa > 0$ and we normalize g to have $\kappa = 1$.

Now we recall (compare 3.6.B.) the following totally geodesic subspaces in $H^{2n}_{\mathbb{C}}$.

(a) complex geodesics that are traces of projective lines in $\mathbb{C}P^n \supset B^{2n} = H^{2n}_{\mathbb{C}}$. These are isometric to $H^2_{\mathbb{R}}$ with curvature -1.

(b) Isotropic planes, that are $U(n, 1)$-translates of

$$H^2_{\mathbb{R}} \subset H^n_{\mathbb{R}} \subset H^{2n}_{\mathbb{C}} \text{ for } H^n_{\mathbb{R}} = \mathbb{R}P^n \cap B^{2n} = H^{2n}_{\mathbb{C}}$$

for the standard embedding $\mathbb{R}P^n \subset \mathbb{C}P^n$. These planes have constant curvature $-\frac{1}{4}$. The boundary sphere $S^{n-1} = \partial B^{2n} = \partial_\infty H^{2n}_{\mathbb{C}}$ admits a unique $U(n, 1)$-invariant subbundle $T'' \subset T(S^{n-1})$ of codimension one which is the maximal complex subbundle of the real hypersurface $\partial B^{2n} \subset \mathbb{C}P^n$. Furthermore, the *Levi form* of ∂B^{2n} defines a conformal structure (i.e. a positive definite quadratic form up to a multiplication by a positive function on ∂B^{2n}) in T".

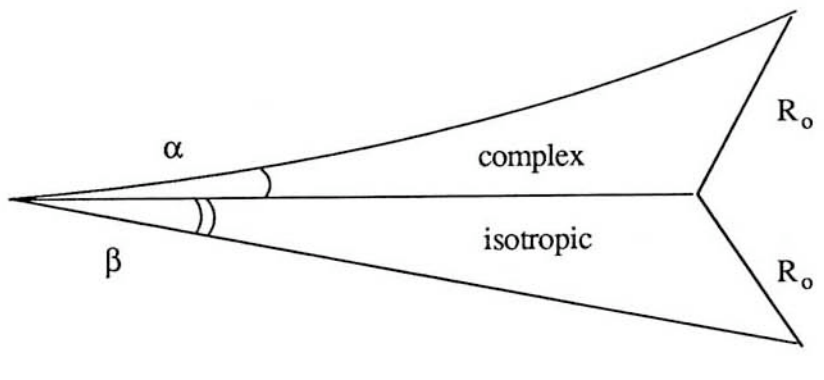
Then the group PU(n, 1) acts on $S^{2n-1} = \partial B^{2n}$ by *conformal transformations* as this group preserves the complex structure as well as the hypersurface ∂B^{2n} and, hence, the conformal class of Levi's form). (Notice that the word "conformal" includes the contact property, that is the preservation of T" as well as the conformality on T"). Conversely, every conformal transformation of S^{2n-1} is induced by a unique isometry of $H_{\mathbb{C}}^{2n}$, as a (relatively) simple argument (similar to the real case) shows. Notice that the isometry group Iso $H_{\mathbb{C}}^{2n}$ consists of two connected components where the component of Id equals PU(n, 1).

Now, the main step of Mostow rigidity proof is as follows.

3.10.A. *Every quasi-isometry* $H_{\mathbb{C}}^{2n} \to H_{\mathbb{C}}^{2n}$ *induces a Mar-quasiconformal homeomorphism* f *of the boundary* S^{2n-1}.

Here we explain the geometric significance of Mar-quasiconformality. For this we notice that every small Ma-r-ball in S^{2n-1} looks like a small thickening of a $(2n - 2)$-dimensional disk tangent at the center to T". (Compare 3.6.B.) In fact such a ball is obtained by a radial projection of a ball $B(r_0) \subset H_{\mathbb{C}}^{2n}$ to $S^{2n-1} = \partial_\infty H_{\mathbb{C}}^{2n}$. To see that we identify S^{2n-1} with the unit tangent sphere in $T_{x_0}(H_{\mathbb{C}}^{2n})$ and observe that the exponential divergence rate of rays in the complex geodesics passing through x_0 is "twice" as fast as in the isotropic planes. In fact in a complex geodesic (which has curvature -1) the distance between two nearby rays grows as a function of the length parameter t as $\sinh t \sim e^t$ and in the isotropic planes it is $\sim e^{t/2} = \sqrt{e^t}$. It follows, the projection of the R_0 ball to S^{2n-1} is much more squeezed in the complex direction than in the isotropic direction.

$$R_0 \approx \alpha \exp t \approx \beta \exp \frac{1}{2} t.$$



Thus $\alpha \approx \sqrt{\beta}$ which gives us a picture of Ma-r-balls as in 3.6.B. This shows the quasi-conformality of the differential of *smooth* Mar-quasiconformal maps, (which amounts in the smooth case to preservation of T"), but the real issue is to prove this without the smoothness assumption. We need an analogue of Theorem 3.9.A for Mar-quasiconformal mappings. It turns out that even the definition of differentiability has to be adapted to the anisotropic character of the sphere at infinity S of the complex hyperbolic space $H^{2n}_{\mathbb{C}}$. In a sense made precise by Metivier and Mitchell, the "tangent cone" of S at any of its points is a copy of the Heisenberg group. A notion of differentiability can be concocted in such a way that differentials become Heisenberg group automorphisms instead of mere linear maps.

3.10.B. <u>The sphere at infinity of rank one symmetric spaces is modelled on Heisenberg groups</u> (see [Mos]4). We first explain this for $S = \partial X$, $X = H^{2n}_{\mathbb{C}}$. Fix two points s_∞ and s_0 in S (note that the isometry group of X is transitive on the pairs of points of S, so the choice makes no difference). The complement $S - \{s_\infty\}$ should be viewed as a copy of the Heisenberg group $N^{2n-1}_{\mathbb{C}}$.

The Heisenberg group $N^{2n-1}_{\mathbb{C}}$ is the simply connected Lie group with Lie algebra $\mathfrak{N}$ split as

$$\mathfrak{N} = \mathfrak{N}'' \oplus \mathfrak{N}',$$

where

- $\mathfrak{N}'$ is the center, and $\dim \mathfrak{N}' = 1$.
- $\dim \mathfrak{N}'' = 2n-2$ and the Lie bracket

$$[\ ,\ ] : \Lambda^2 \mathfrak{N}'' \to \mathfrak{N}'$$

is a symplectic form.

These properties define a unique group $N^{2n-1}_{\mathbb{C}}$ (up to isomorphism). It possesses a one parameter group of automorphsms δ_t, $t \in \mathbb{R}_+$, defined by

$$\delta_t v = tv \quad \text{for } v \in \mathfrak{N}'' ,$$

$$\delta_t v = t^2 v \quad \text{for } v \in \mathfrak{N}' .$$

Fix some complex structure J on $\mathfrak{N}''$ compatible with the symplectic structure [,]. J determines a metric on $\mathfrak{N}''$. Use left translations to propagate $\mathfrak{N}''$ and its metric into a plane distribution on $N = N_{\mathbb{C}}^{2n-1}$. These data determine a Carnot-Caratheodory metric $\widetilde{di}$ on N (compare 3.6.B_1). This metric is invariant under the left translations, the U(n − 1) "rotations", and is multiplied by t under δ_t.

There exists (and unique up to U(n − 1)) an isometric action of $N_{\mathbb{C}}^{2n-1}$ on $H_{\mathbb{C}}^{2n}$ which fixes s_∞, and has the following properties :

- The map $a : N \to S - \{s_\infty\}$, $n \mapsto n.s_0$ is a diffeomorphism ;
- a takes the left invariant plane field $\mathfrak{N}''$ to T'' ;
- a takes J to the natural complex structure on T'' ;
- a conjugates the group δ_t into a group of isometries of $H_{\mathbb{C}}^{2n}$ which fixes s_∞ and s_0;
- under a, the sphere at infinity of a complex geodesic (compare 3.10) through s_∞ corresponds to an orbit of the action (by translation) of the center ;
- a takes the Carnot-Caratheodory metric $\widetilde{di}$ to a metric conformal, either to the Levi-form (3.10) or the metric di arising as a normalized limit of induced metrics on large spheres (3.6.B_1).

As a consequence, a takes Mar-quasiconformal mappings of S fixing s_∞ to $\widetilde{di}$-quasiconformal mappings of N.

A similar picture exists for the other rank one symmetric spaces. The group $N_{\mathbb{H}}^{4n-1}$ (resp. $N_{\mathbb{C}a}^{15}$) has Lie algebra

$$\mathfrak{N} = \mathfrak{N}'' \oplus \mathfrak{N}'$$

where $\mathfrak{N}''$ identifies with quaternionic space $\mathbb{H}^{n-1}$ (resp. with the Cayley line $\mathbb{C}a$), $\mathfrak{N}'$ with imaginary quaternions Im$\mathbb{H}$ (resp. imaginary Cayley numbers Im$\mathbb{C}$a), and the Lie bracket

$$[\ ,\ ] : \Lambda^2 \mathfrak{N}'' \to \mathfrak{N}'$$

is given by

$$\left[(q_1, \dots, q_{n-1}), (q_1', \dots, q_{n-1}')\right] = \mathrm{Im}\left(\sum_{i=1}^{n-1} q_i \overline{q_i'}\right).$$

3.10.C. Differentiability with respect to a group with dilations.

One should think here of the sphere at infinity of the real hyperbolic n-space as modelled on (i.e. locally conformally isomorphic to) $\mathbb{R}^{n-1}$ equipped with its usual dilations. Theorem 3.9.A is concerned with the existence of a differential Df in the usual obvious sense, i.e., if f is a self map of $\mathbb{R}^{n-1}$ and $f(0) = 0$,

$$Df_o = \lim_{t \to 0} \frac{1}{t} f(th) .$$

This theorem generalizes to groups N equipped with dilations δ_t as we have encountered in 3.10.B.

3.10.C_1. Definition : Let N denote a Heisenberg group over $\mathbb{C}$, $\mathbb{H}$ or $\mathbb{C}a$, and δ_t denote the natural one parameter group of (dilating) automorphisms.

Let f be a map $N \to N$ such that $f(e) = e$. We say that f is *δ-differentiable* at e *with differential* Df_e if, as $t \to 0$, the maps

$$f_t = \delta_{1/t} \circ f \circ \delta_t$$

converge uniformly on compact subsets to Df_e. Using left translations, we define δ-differentiability at any point.

δ-differentiability is somewhat weaker than usual differentiability. Indeed, it essentially means

- f is differentiable in the $\mathfrak{N}''$-directions, and Df takes $\mathfrak{N}''$ into $\mathfrak{N}''$;
- the $\mathfrak{N}'$ component of f is differentiable in the $\mathfrak{N}'$ direction ;
- the $\mathfrak{N}'$ component of f behaves like $\sqrt{u}$ under an increment of u in the $\mathfrak{N}''$ direction.

Thus a map f which is differentiable in the ordinary sense *and preserves* $\mathfrak{N}''$ is not far from being δ-differentiable.

Our main technical result is

3.10.C_2. Regularity theorem : *Every* Mar-*quasiconformal mapping of the sphere at infinity of a rank one symmetric space (locally viewed as a Heisenberg group with dilations) is absolutely* *continuous, and admits almost everywhere a δ-differential which is a group automorphism commuting with* $\{\delta_t\}$.

The strength of the theorem comes from the assertion that the differential is a group morphism. Indeed, in the complex cases, it states that the derivative in the $\mathfrak{N}''$ (or T") direction exists and preserves the natural symplectic form up to a scale. The symplectic form can be viewed as follows : let α be a 1-form on S whose kernel is T" then $d\alpha|_{\mathfrak{N}''}$ only depends on $\mathfrak{N}''$ (up to a scale). If f is a C^1 diffeomorphism of S which preserves T", then

$$f^*\alpha = \alpha \quad \text{up to a scale.}$$

The proof that $f^*d\alpha_{|T''} = d\alpha_{|T''}$ up to a scale requires that f is twice differentiable. Thus theorem $3.10.C_2$ to some extent asserts existence of some second derivative.

$3.10.C_3$. There is an unpleasant point here as the existing proofs of theorem $3.10.C_2$, in fact, require stronger assumptions on the mapping. Let Y be a metric space. Call the difference set

$$A = B(x, r_2) - B(x, r_1)$$

between concentric balls in Y with ratio of the radii $\frac{r_2}{r_1} = a$ an a-*annulus*. We say that a map between metric spaces $f : Y \to Z$ is *quasisymmetric* if f takes every small enough a-annulus into a $\eta(a)$-annulus (where η is some function on $\mathbb{R}_+$).

Quasisymmetry obviously implies quasiconformality and is the only reasonable definition for quasiconformality in one dimension.

The boundary extension of a quasiisometry of $H^{2n}_{\mathbb{C}}$ is easily seen to be quasisymmetric as well as its inverse ; this saves our approach to Mostow rigidity.

Once the a.e. existence and quasiconformality of Df is established the proof is concluded as in the real case with the Mautner lemma which now yields the ergodicity of the Γ-action on PT".

3.11. QIs-*rigidity of* $H^{4n}_{\mathbb{H}}$ *and* $H^{16}_{\mathbb{C}a}$. Recall (from 3.4) that a metric space X is called QI_s-*rigid* if the homomorphism

$$\text{Iso}(X) \to \overline{\text{QIs}}(X)$$

is bijective.

Let us collect the facts relevant to the QIs-rigidity previously established for X a simply connected Riemannian manifold with curvature $K \leq -\kappa < 0$.

a) The natural map $\mathrm{Iso}\,(X) \to \overline{\mathrm{QIs}}\,(X)$ is injective (see 3.3. A').

b) The group $\overline{\mathrm{QIs}}\,(X)$ embeds as a subgroup of homeomorphisms of the sphere at infinity ∂X of X (see 3.5. F'). Furthermore, these homeomorphisms are quasiconformal with respect to the natural Mar-conformal structure on ∂X.

We see that in the context of negative curvature , the QIs-rigidity problem translates into a problem about Mar-quasiconformal mappings of ∂X.

3.11.A. Theorem. : *Let* X *be either quaternionic hyperbolic space* $H^{4n}_{\mathbb{H}}$, $n \geq 2$, *or Cayley hyperbolic plane* $H^{16}_{\mathbb{C}a}$. *Every Mar-quasiconformal mapping of* ∂X *extends to an isometry of* X.

As a consequence, X is $\overline{\mathrm{QIs}}$ -rigid.

3.11.B. The proof follows the lines of Mostow rigidity, as described in 3.9. According to $3.10.C_2$ Mar-quasiconformal (in fact, quasisymmetric, see $3.10.C_3$) mappings of ∂X have almost everywhere differentials which are Heisenberg automorphisms.

The extra rigidity of the quaternionic and Cayley spaces follows from the fact that the corresponding Heisenberg groups have few automorphisms.

Whereas, in the complex case, potential differentials identify with $(2n-2) \times (2n-2)$ matrices that are scalar multiples of symplectic matrices, in the other cases we have

Lemma : *Let* $Z^3 \subset \Lambda^2 H^{n-1}$ (respectively $Z^7 \subset \Lambda^2\, \mathbb{C}a$) *be the* 3-*dimensional subspace generated by the components of*

$$\mathrm{Im}\left(\Sigma\, q_i \bar{q}_i'\right)$$

Let A *be a* $\mathbb{R}$-*linear map of* H^{n-1} *(resp.* $\mathbb{C}a$*) with det* $A = 1$. *If* A *fixes* Z^3 (resp. Z^7) *then* A *is* $\mathbb{H}$-*linear and orthogonal, i.e.,*

$$A \in Sp(n-1),$$

$$(\textit{resp. } A \in \mathrm{Spin}(7) \subset \mathrm{Sl}(8, \mathbb{R})).$$

Combining 3.11.C_2 and this lemma, we see that every Mar-quasiconformal mapping of the sphere at infinity of $H^{4n}_{\mathbb{H}}$, $n \geq 2$, or H^{16}_{Ca} is "almost everywhere conformal". The proof is concluded as in the real or complex case.

3.11.C. $H^n_{\mathbb{R}}$ and $H^{2n}_{\mathbb{C}}$ are not $\overline{\text{QIs}}$-rigid.

A selfmapping f ∂X for $X = H^n_{\mathbb{R}}$ or $H^{2n}_{\mathbb{C}}$, has a natural extension $\tilde{f}$ to X in polar coordinates. Indeed, ∂X identifies with the sphere of angular coordinates θ, and one sets

$$\tilde{f}(r, \theta) = (r, f(\theta)).$$

It is readily checked that if f is bilipschitz (case $H^n_{\mathbb{R}}$) and preserves the hyperplane field T" (case $H^{2n}_{\mathbb{C}}$), then $\tilde{f}$ is bilipschitz. Since $\overline{\text{QIs}}$ injects into Homeo (∂X), one finds in this way many non trivial elements in $\overline{\text{QIs}}$ (X) for $X = H^n_{\mathbb{R}}$ or $H^{2n}_{\mathbb{C}}$. In fact, one has the following

3.11.C_1. Theorem (Tukia, Reimann [Tuk]$_2$, [Rei]). *When* $X = H^n_{\mathbb{R}}$ or $H^{2n}_{\mathbb{C}}$, $\overline{\text{QIs}}$ (X) *identifies with the group* QC(∂X) *of* Mar-*quasiconformal (resp. quasisymmetric) homeomorphisms of* $S = \partial X$.

The first step in the proof is the fact, alluded to in 3.8.B_1, that any Mar-quasiconformal mapping of S is Δ-quasiconformal in the sense of 3.8. In fact, the function Δ can be reconstructed in terms of *moduli of curve families* in S or *capacities*, and the point is to show that the relevant capacities are non zero. This is an inequality of Sobolev type.

The second step is simple and general : Δ quasiconformal mappings of $S = \partial X$ extend to quasiisometries of X almost by definition (see 3.8.B.).

3.11.C_2. Recapturing O(n, 1) or U(n, 1) from a lattice.

Here is a way to attach a "continuous" group to a discrete group, which in the case of cocompact lattices in G = O(n,1) or U(n, 1) produces G. This method is not functorial under quasiisometries, and thus does not fulfill the program of 3.4.A, but it is functorial under group isomorphisms, and thus implies Mostow-rigidity.

Let Γ be a finitely generated group with a word metric. Say an element f of $\overline{\text{QIs}}$ (Γ)

is a 1-quasiisometry of Γ if the subgroup generated by Γ in $\overline{\mathrm{QIs}}(\Gamma)$ forms a *uniformly* quasiisometric family (i.e. every transformation φ from this subgroup satisfies the inequality (*) of 3.2.B'. with a constant A independent of φ). We obtain a group $1-\overline{\mathrm{QIs}}(\Gamma)$ of 1-quasiisometries.

For Γ a cocompact lattice in O(n, 1), $\overline{\mathrm{QIs}}(\Gamma) = \mathrm{QC}(S^{n-1})$ identifies with quasiconformal mappings of the standard sphere and a 1-quasiisometry is a mapping of S^{n-1} which generates with Γ a uniformly quasiconformal group. Elaborating on an idea of Mostow, Tukia-Väisälä prove that a uniformly quasiconformal group containing enough conformal transformations is conformal, see [Tuk-Väi].

The argument extends to $H^{2n}_{\mathbb{C}}$, thanks to Reimann's result.

3.12. Topological proof of Mostow rigidity for $H^n_{\mathbb{R}}$.

We adopt the point of view of locally homogeneous spaces explained in 1.5. Then discrete subgroups of the Lie group O(n, 1) correspond to Riemannian manifolds (or orbifolds) with constant curvature -1. Mostow rigidity for cocompact latices in O(n, 1) states.

3.12.A. Theorem : *Let* V, V' *be compact manifolds (orbifolds) with constant curvature* -1. *Then every homotopy equivalence* $V \to V'$ *can be deformed to an isometry.*

We shall again use the fact that homotopy equivalences h between compact manifolds give rise to homeomorphsims f_h of S, the standard sphere viewed as the sphere at infinity of $H^n_{\mathbb{R}}$. The point is to show that f_h has to be a conformal mapping of S (compare 3.9). We shall use a characterization of conformal mappings of S in terms of the regular ideal simplices.

3.12.B. Ideal simplices. Let us identify $H^n_{\mathbb{R}}$ with the ball $B^n \subset S^n$ (see 3.9) and observe that the group O(n, 1) acts on B^n by *projective* transformations which send arcs of great circles in B^n again to such arcs. It follows, that these arcs also serve as the geodesic segments in $H^n_{\mathbb{R}} = B^n$ with an O(n, 1)-equivariant metric g (compare 3.9), and so the projective structure in $H^n_{\mathbb{R}}$ defined by the geodesics is isomorphic to the ordinary projective structure in the ball B^n (which can be equally thought of as a round ball in $\mathbb{R}^n$ rather than in S^n). In particular, one may speak of (geodesically) convex subsets in $H^n_{\mathbb{R}}$ (corresponding to convex subsets in B^n) and of *geodesic simplices* in $H^n_{\mathbb{R}}$ which are *convex hulls* of systems of points in $H^n_{\mathbb{R}}$ in general position. In fact, for every k-tuple of points $x_o, \dots x_k$ in $H^n_{\mathbb{R}}$ one can construct a canonical map of the standard k-simplex Δ^k into $H^n_{\mathbb{R}}$ with the image the convex hull of $\{x_{11}, \dots x_k\}$. Namely, one sends the point with barycentric coordinates $(m_o, \dots, m_k)$ in Δ^k to the *Riemannian center of mass* of the points x_i in $H^n_{\mathbb{R}}$ with the weight m_i assigned to each x_i, $i = 0, 1, \dots, k$. This is the (necessarily unique for $K \le 0$) point $y \in H^n_{\mathbb{R}}$ which minimizes the weighted sum

$$\sum_{i=0}^{k} m_i \operatorname{dist}^2(y, x_i).$$

This map, say $\sigma : \Delta^k \to H^n_{\mathbb{R}}$, is called the *straight* (singular) simplex *spanned by* $x_o, \dots, x_k$.

Now, let some of the points among $x_o, \dots, x_k$ lie on the (ideal) boundary $S = H^n_{\mathbb{R}}$ identified with the ordinary boundary of the ball $B = H^n_{\mathbb{R}}$. Then one can still speak of convex hull of points in $H^n_{\mathbb{R}} \cup S = B \cup \partial B$ and then intersect these hulls with $H^n_{\mathbb{R}} \subset H^n_{\mathbb{R}} \cup S$. For example, the convex hull of two distinct points in $S = \partial H^n_{\mathbb{R}}$ is a double infinite geodesic in $H^n_{\mathbb{R}}$. Furthermore, the map $\sigma = \sigma(x_o, \dots, x_k) : \Delta^k \to H^n_{\mathbb{R}} \cup S$ can also be defined in this case unless among the points $x_o, \dots x_k$ there are at most two distinct ones. In fact, one defines this σ as the limit (whose existence is easy to prove) of the corresponding maps $\sigma(x_i')$ for $x_i' \in H^n_{\mathbb{R}}$ as these points approximate the points x_i, $i = 0, \dots, k$. A particularly interesting case is where we have all points $x_1, \dots, x_k$ on the ideal boundary and then the corresponding map σ is called the *ideal* (straight singular) simplex spanned by these points. Notice that this σ maps Δ^k-{the set of vertices} $\to H^n_{\mathbb{R}}$.

An ideal simplex is called *regular* if it has *maximal symmetry.* This means every permutation of vertices $x_o, \dots x_k$ in $S = H^n_{\mathbb{R}}$ is induced by some isometry of $H^n_{\mathbb{R}}$, i.e. by a conformal transformation of the sphere $S = S^{n-1}$. Every ideal 2-simplex with distinct vertices is (obviously) regular but it is not at all so for $k \geq 3$. In fact for $n \geq 3$ one has the following elementary

3.12.B_1. Lemma. *A homeomorphism of* S *is conformal if and only if it takes every regular ideal* n*-simplex (here, it is just an* $(n+1)$*-tuple of points in* S) *to another regular simplex.*

Idea of the proof. The case of a general homeomorphism h reduces to that *fixing* some regular $(n+1)$-tuple of points. Then one shows that the preservation of the regularity implies that $h = \mathrm{Id}$. The details are left to the reader.

The following geometric characterization of the regular simplices is also quite elementary but not so easy.

3.12.B_2. Theorem. (Haagerup-Munkholm). *Among all straight* n*-simplices, the regular ideal ones, and only them, have maximal hyperbolic volume.*

See [Haa-Mun] for the proof.

3.12.C. Scheme of the proof of Mostow's rigidity.

Fix an orientation of $H^n_{\mathbb{R}}$ and assume that the manifolds V and V' in question are oriented and the covering maps of $H^n_{\mathbb{R}}$ onto V and V' are orientation preserving. (This can always be achieved by taking the oriented double coverings of V and V' if necessary). The covering maps $H^n_{\mathbb{R}} \to V$ and $H^n_{\mathbb{R}} \to V'$ send singular simplices of $H^n_{\mathbb{R}}$ to those in V and V' and then one naturally defines *straight* simplices in V and V' as well a *straight ideal* and *ideal regular* simplices in V and V'. Then a straight n-simplex (in $H^n_{\mathbb{R}}$, V or in V') is called *positive*, or *positively oriented* if the implied map σ of Δ^n has positive Jacobian for the standard oriented volume element on Δ^n.

Denote by $\mathfrak{R}$ the set of positive ideal simplices in $H^n_{\mathbb{R}}$ and the group of orientation preserving isometries of $H^n_{\mathbb{R}}$ acts transitively on $\mathfrak{R}$. Furthermore if $H^n_{\mathbb{R}}$ is identified with the universal covering $\tilde{V}$ of V then the group $\Gamma = \pi_1(V)$ acts on $\mathfrak{R}$ and this action is cocompact whenever V is compact. As $\mathfrak{R}$ carries a natural $\mathrm{Iso}_+ H^n_{\mathbb{R}}$ -invariant measure, one has a natural measure on $\mathfrak{R}/\Gamma$ and in the cocompact case one can average over this measure.

In other words one can average Γ-invariant functions on $\mathfrak{R}$. For example if $f : S \to S$ is the boundary map induced by a given homotopy equivalence $h : V \to V'$ via the covering map $\tilde{h} : \tilde{V} = H^n_{\mathbb{R}} \to H^n_{\mathbb{R}} = \tilde{V}'$, then the function

$$\text{vol}(f\sigma) = \text{volume of } f(\sigma) \subset H^n_{\mathbb{R}}$$

is Γ-equivariant on the set $\mathfrak{R}$ of regular ideal simplices σ in $H^n_{\mathbb{R}}$.

In the next section, we shall explain the following formula

$$(*) \qquad \text{average}(\text{vol}(f\sigma)) = \frac{\text{vol}V}{\text{vol}V'} \text{ average } (\text{vol}(\sigma)).$$

We use theorem $3.12.B_2$: for each $\sigma \in \mathfrak{R}$, $f\sigma$ is an ideal simplex, not necessarily regular, so

$$\text{vol}(f\sigma) \leq \text{vol}(\sigma).$$

We obtain that

$$\text{vol}(V) \leq \text{vol}(V')$$

Reversing h and f, we obtain $\text{vol}(V) = \text{vol}(V')$, thus, for almost every regular ideal simplex σ,

$$\text{vol}(f\sigma) = \text{vol}(\sigma),$$

that is, $f\sigma$ is regular. By continuity, every $f\sigma$ is regular. Lemma 3.12.B, implies that f is conformal. f extends to a Γ-equivariant isometry of $H^n_{\mathbb{R}}$, that is, an isometry of V to V' homotopic to h.

3.12.D. Proof of the formula (*) for average volumes. It stems from a cohomological interpretation of the average. This requires a slight modification of singular homology, in order to admit chains which are compactly supported measures on the space of C^1-singular simplices. Let us admit this modification. Let σ be a singular simplex in $X = H^n_{\mathbb{R}}$ which admits a miror symmetry τ. Let $\pi : X \to V$ be the universal covering map. We define a (generalized) singular chain α in V by

$$a = \int_{\Gamma\backslash G} \text{sign}(g)\, \pi(g\sigma) \frac{dg}{\text{vol}\, \Gamma\backslash G}$$

where here G is the group of all isometries of X, where $\pm$ sign(g) refers to preservation or reversal of the orientation by $g : X \to X$.

For short, we will denote $\dfrac{dg}{\mathrm{Vol}(\Gamma\backslash G)}$ by $d_\Gamma g$.

3.12.D. <u>Claim</u> : α is a cycle, i.e. $\partial\alpha = 0$

<u>Proof</u> : Recall that, for a simplex σ,

$$\partial\sigma = \sum_{i=0}^{n} \partial_i \sigma$$

where $\partial_i\sigma$ are the (suitably oriented) faces of σ.

We arrange so that the mirror symmetry exchanges the i-th and the (n – i)-th faces :

$$\tau(\partial_i\sigma) = -\,\partial_{n-i}\sigma.$$

Let τ_i denote the mirror symmetry with respect to $\partial_i\sigma$ then

$$\tau_i(\partial_i\sigma) = -\,\partial_i\sigma.$$

We compute, using the biinvariance of the Haar measure on $G = \mathrm{Iso}H^n_{\mathbb{R}}$,

$$\begin{aligned}
\partial\alpha &= \sum_{i=0}^{n} \int_{\Gamma\backslash G} \mathrm{sign}\,(g)\, \pi\, \partial_i\big(g\sigma\big)\, d_\Gamma\, g \\
&= \sum \int \mathrm{sign}\,(g)\, \pi\, \partial_i\big(g\tau\, \tau_i\sigma\big)\, d_\Gamma\, g \\
&= \sum \int \mathrm{sign}\,(g)\;\; \pi\, g\tau\, \tau_i\big(\partial_i\, \sigma\big)\, d_\Gamma\, g \\
&= \sum \int_{\Gamma\backslash G} \mathrm{sign}\,(g)\, \big(-\,\pi g \partial_{n-i}\, \sigma\big)\, d_\Gamma\, g \\
&= -\,\partial\alpha
\end{aligned}$$

so that $\partial\alpha = 0$.

Thus α defines a homology class $[\alpha]$, which we evaluate against the volume form Ω of V,

$$\langle [\alpha], \Omega \rangle = \int_{\Gamma\backslash G} \mathrm{sign}(g) \left(\int_{\pi(g\sigma)} \mathrm{sign}(g)\ \Omega \right) d_\Gamma g$$

$$= \int_{\Gamma\backslash G} \mathrm{volume}(g\sigma)\ d_\Gamma g$$

$$= \mathrm{volume}(\sigma).$$

Let $h = V \to V'$ be an orientation preserving homotopy equivalence. Let $\pi' : X \to V'$ be the covering map. The image cycle is

$$h_* \alpha = \int_{G\backslash\Gamma} \mathrm{sign}(\sigma)\ \pi' \circ h \circ g \circ \sigma\ d_\Gamma g$$

evaluated against the volume form Ω' of V', it gives

$$\langle h_* \alpha, \Omega' \rangle = \int_{\Gamma\backslash G} \mathrm{volume}(h \circ g \circ \sigma)\ d_\Gamma g.$$

Since $[h^* \Omega'] = \dfrac{\mathrm{vol}\ (V')}{\mathrm{vol}\ (V)} [\Omega]$ in cohomology, we conclude that

$$\int_{\Gamma\backslash G} \mathrm{vol}(h \circ g \circ \sigma)\ d_\Gamma g = \frac{\mathrm{vol}\ V'}{\mathrm{vol}\ V} \int_{\Gamma\backslash G} \mathrm{vol}(g\ \sigma)\ d_\Gamma g.$$

3.12.D_2 Straightening

Assume now that σ is a finite straight simplex in V. We want to replace the curved simplex $h \circ \sigma$ by a straight simplex with the same vertices. This is the straightening operation. This is clearly well defined in $X = H^n_{\mathbb{R}}$ where it is functorial under isometries, and has a functorial chain homotopy

$$1 - \mathrm{straight} = \partial B + B\partial.$$

Then the straightening of simplices in V' covered by X is defined as follows. Given σ in V', take any lift $\tilde{\sigma}$ to X, and set

$$S'(\sigma) = \pi' \mathrm{straight}\ (\tilde{\sigma})$$

Clearly the operator S' induces the identity in homology.

Then we have

$$h_* \alpha = \int_{\Gamma \backslash G} \operatorname{sign}(\sigma) \; S'(h \circ g \circ \sigma) \; dg$$

and thus

$$\int_{G \backslash \Gamma} \operatorname{vol} S'(h \circ g \circ \sigma) \; d_\Gamma g = \frac{\operatorname{vol} V'}{\operatorname{vol} V} \int_{\Gamma \backslash G} \operatorname{vol}(g \sigma) \; d_\Gamma g .$$

Finally, we let σ converge to a positively oriented regular ideal simplex. By continuity, we get formula (*). This finishes the topological proof of Mostow rigidity.

3.12.E. Generalizations.

W. Thurston has extended the above argument to maps $V \to V'$ which are not necessarily homotopy equivalences.

Theorem (W. Thurston). *Let* V, V' *be complete oriented* n*-manifolds of constant curvature* -1 *and let* $h : V \to V'$ *be a proper continuous map of non-zero degree* d. *If* $n \geq 3$ *and* $\operatorname{Vol} V \leq |d| \operatorname{Vol} V' < \infty$ *then* h *is homotopic to a locally isometric* $|d|$*-sheeted covering* $V \to V'$.

The new feature here is a (possible) lack of a *continuous* extension of $\tilde{h} : \tilde{V} \to \tilde{V}'$ to the sphere at infinity. But Thurston (see [Th]) constructs a *measurable* extension using the brownian motion in $\tilde{V}'$.

§ 4 - Rigidity via Bochner formulas for harmonic maps

Harmonic maps f between Riemannian manifolds are the critical points of the *energy functional* E(f) which is the integral of the *energy density* e(f) defined as follows.

4.1. Energy . We use here the norm of linear maps $D : \mathbb{R}^n \to \mathbb{R}^q$ defined by

$$\|D\| = (\text{trace } D^*D)^{1/2}$$

Geometrically, we look at the ellipsoid in $\mathbb{R}^q$ which is the image of the in $\mathbb{R}^n$. This ellipsoid has principal semiaxes of certain lengths $\lambda_1 \geq \lambda_2 > \ldots \geq \lambda_q \geq 0$ and

$$\|D\|^2 = \sum_{i=1}^{q} \lambda_i^2 .$$

Another equivalent definition is

$$\|D\|^2 = c_n \int_{S^{n-1}} \|D(s)\|^2 \, ds ,$$

where S^{n-1} denotes the unit sphere in $\mathbb{R}^n$ and

$$c_n = n\,(\text{Vol } S^{n-1})^{-1} .$$

Next for a C^1-map between Riemannian manifolds, say $f : X \to Y$ we denote by $Df : T(X) \to T(Y)$ the differential and by $e(f) = e(f)(x)$ half the squared norm of Df on $T_x(X)$,

$$e(f) = \frac{1}{2} \|Df\|^2$$

This is also called the *pointwise energy* or the *energy density* of f. Then the (global) energy is defined by integration over X,

$$E(f) = \int_X e(f)(x)\, dx$$

One can also obtain E(f) by integrating over the unit tangent bundle S(V) as follows

$$E(f) = c_n \int_S \frac{1}{2}\|Df(s)\|^2 \; ds$$

for the above constant c_n.

Example. If Y is the real line then $E(f) = \int \frac{1}{2}\|\text{grad } f\|^2$ and if $Y = \mathbb{R}^q$ and f is given by the components $f_1, \ldots, f_q$ then

$$E(f) = \sum_{i=1}^{q} \int \frac{1}{2}\|\text{grad } f_i\|^2 \; .$$

4.2. Non-linear Laplacian. This is a (non-linear differential) operator which assigns to every C^2-map $f : X \to Y$ a *vector field* in Y *along* $f(X)$ which is a loose name for a section of the induced bundle $f^*(T(Y))$ over X. This field is denoted by Δf (it is sometimes called the *tension of* f and denoted by $\tau(f)$) and is defined in several ways as follows.

4.2.A_1 Euclidean way. Let first $Y = \mathbb{R}^q$ and $f = (f_1 \ldots f_q)$. Then $\Delta f(x)$ is the vector in $T_Y(\mathbb{R}^q) = \mathbb{R}^q$ for $y = f(x)$ with components $\Delta f_1, \Delta f_2, \ldots, \Delta f_q$ for the ordinary *Laplace-Beltrami* operator on functions. Then the case of a non-Euclidean Y is reduced to the Euclidean one by using the exponential maps $\exp : T_y(Y) \to Y$ at the points $y = f(x)$ and thus identifying a small neighbourhood of each point $y \in Y$ with $\mathbb{R}^q = T_y(Y)$ for $q = \dim Y$. More precisely, we define $\Delta f(x)$ by first composing f with the inverse exponential at y which gives us a map of (a small neighbourhood of x in) X into $T_y(Y)$. Then we take the above Euclidean Laplacian of this composed map and bring it back to Y with the tautological isomorphism

$$T_o(T_y(Y)) = T_y(Y) \, .$$

To complete this discussion we recall that the Laplace-Beltrami operator on functions equals the ordinary Laplacian in the *exponential* coordinates,

$$\Delta f(x) = \sum_{i=1}^{n} \partial_i^2 f$$

where ∂_i are the images of coordinate vector fields in $T_x(X) = \mathbb{R}^n$ under the differential of the exponential map. An alternative definition of the Laplace-Beltrami Δ is $\Delta = -d^* d$ where d

is the (exterior) differential (thought of as the operator from functions to 1-forms) and d^* is the adjoint operator.

4.2.A_2 Laplacian and Hessian. First we recall the second quadratic (fundamental) form II of a submanifold X in a Riemannian manifold Z. This is a symmetric bilinear form on $T(X) \subset T(Z)$ with values in the normal bundle $N(X) \subset T(Z)|X$ defined by

$$II(\tau_1, \tau_2) = P_N \nabla_{\tau_1} \tau_2$$

where τ_1 and τ_2 are vector fields tangent to X, where ∇ denotes the covariant derivative in Z and P_N is the projection of $T(Z)|X$ to N. One verifies easily that II is indeed a form,

$$II(\rho_1 \tau_1, \rho_2 \tau_2) = \rho_1 \rho_2 II(\tau_1, \tau_2),$$

for arbitrary functions ρ_1 and ρ_2 on X and it is symmetric, i.e.

$$II(\tau_1, \tau_2) = II(\tau_2, \tau_1).$$

Notice that $X \subset Z$ is *totally geodesic* if and only if $II = 0$.

Next we consider the graph $\Gamma_f \subset W \times Y$ of a C^2-map $f: X \to Y$ and denote by $P: T(X) \to T(\Gamma_f)$ and $Q: N(\Gamma_f) \to f^*(T(Y))$ the obvious maps, where Γ_f is identified with X by

$$(x, f(x)) \leftrightarrow x.$$

Then we define the *Hessian* Hess f by

$$\text{Hess} = Q \circ II \circ P \qquad \text{for } II = II(\Gamma_f).$$

Thus the Hessian is a quadratic form on X with values in $f^*(T(Y))$. Notice that

$$\text{Hess } f = 0 \Leftrightarrow II(\Gamma_f) = 0$$

and maps f with Hess = 0 are called *geodesic*

Now we set

$$\Delta f = \text{Trace Hess } f.$$

4.2.A_2'. Let us give another description of the Hessian. We denote by α the differential of f viewed as a linear form on X with values in $f^*(T(Y))$. Then, using the covariant derivative ∇^X in T(X) and ∇^* in $f^*(T(Y))$ pulled back from Y, one obtains the covariant derivative $\beta = \nabla\alpha$ which is a symmetric $f^*(T(Y))$-valued bilinear form on T(X) defined by the identity

$$\nabla^*_{\tau_2} \alpha(\tau_1) = \beta(\tau_2, \tau_1) + \alpha(\nabla^X_{\tau_2} \tau_1)$$

for all tangent fields τ_1 and τ_2 on X. In other words,

$$\text{Hess } f \underset{\text{def}}{=} \nabla\alpha \underset{\text{def}}{=} \nabla Df$$

and

$$\Delta f = \text{Trace } \nabla \text{ } Df$$

as it should be. (We leave to the reader to check the correctness and the equivalence of the both definitions of Hess).

4.2.A_3. We conclude with a more geometric definition of Δf using the geodesics through $x \in X$. Every such geodesic γ_s defined with a unit tangent vector $s \in T_x(X)$ goes under f into some curve $\tilde{\gamma}_s$ in Y which we parametrize by the length parameter t in X, such that $\tilde{\gamma}_s(0) = y = f(x)$. Then we take the second covariant derivative of $\tilde{\gamma}(t)$ in Y and obtain Δf by averaging on the unit tangent sphere $S^{n-1} \subset T_x(X)$. Namely

$$\Delta f(x) = c_n' \int_{S^{n-1}} \frac{d^2}{dt^2} \tilde{\gamma}_s(0)\, ds,$$

for $c_n' = n/\text{Vol } S^{n-1}$. We suggest the reader to prove the equivalence of this to the previous definition in order to obtain certain insight into the meaning of ΔF.

4.2.B. <u>Δf as minus the gradient of E</u>. Let us view the energy $E = E(f)$ as a function on the space of maps $X \to Y$ thought of as an infinite dimensional manifold. The tangent space to this manifold at every f equals the space of fields in Y along f(X) (that are sections $X \to f^*(T(Y))$ and it has a natural scalar product

$$\langle \tau_1, \tau_2 \rangle = \int_X \langle \tau_1, \tau_2 \rangle_Y dx .$$

Then the *Euler-Lagrange equations* for the energy E gives us the gradient of E with respect to the above scalar product. In fact a simple computation shows that :

$$\text{grad } E(f) = -\Delta f, \quad (*)$$

provided X is a compact manifold without boundary. In plain words (*) gives the following formula for the variation of E along one parameter families of maps $f_t : X \to Y$ that are maps $X \times [0, 1] \to Y$.

$$\frac{dE(f_t)}{dt} = -\langle \partial_t f_t, \Delta f_t \rangle, \qquad (**)$$

where $\partial_t f_t = D(\frac{\partial}{\partial t})$ for the differential D of the map $X \times [0, 1] \to Y$ and $\frac{\partial}{\partial t}$ is the coordinate t-field in $X \times [0, 1]$. In fact, (**) is an integral identity

$$\frac{\partial}{\partial t} \int_X e(f_t)(x)\, dx = \int_X \langle \partial_t (f_t)(x), \Delta f_t(x) \rangle_Y dx$$

whose (standard) proof consists in showing that $\delta = \frac{\partial}{\partial t} e(f_t) - <.,.>_Y$ is a *divergence* term, that is the divergence of certain function on X.

An important consequence of (*) reads

4.2.B$_1$. *If a* C^2*-smooth map* f *is a critical point for* E *then* $\Delta f = 0$.

Recall that C^2-maps f with $\Delta f = 0$ are called *harmonic* . The above proposition suggests the following approach to the existence of such maps. Start with an arbitrary map f_o : $X \to Y$ and minimize the energy in the homotopy class $[f_o]$ of f_o. Then, the minimizing map $f \in [f_o]$, provided it exists and is C^2-smooth, is harmonic. We shall see in the following section that this approach is well justified if the receiving manifold Y has non-positive sectional curvature.

4.3. Energy and Laplacian for $K(Y) \le 0$. Let Y be complete simply connected with $K \le 0$. Then every two points y_o and y_1 in Y are joined by a unique geodesic segment $[y_o, y_1] \subset$

Y of points in Y. Notice that this operation for $K(Y) \neq 0$ does not abide the usual rules of commutativity and associativity if the points in question do not lie on a single geodesic. For example, in general

$$\frac{2}{3}\left(\frac{1}{2}x_1+\frac{1}{2}x_2\right)+\frac{1}{3}x_3 \neq \frac{2}{3}\left(\frac{1}{2}x_2+\frac{1}{2}x_3\right)+\frac{1}{3}x_1$$

Next, for every two maps f_0 and f_1 of X to Y one defines the *geodesic* homotopy f_t by

$$f_t(x) = (1-t)\, f_0(x) + t\, f_1(x).$$

The convexity of the distance function in Y (insured by $K \leq 0$, see [B-G-S]) implies that the length of each tangent vector $\tau \in T(X)$ in Y is a convex function in t. That is

$$\|Df_t(\tau)\| \leq (1-t)\|Df_0(\tau)\| + t\|Df_1(\tau)\|$$

for all $\tau \in T(X)$ and $t \in [0, 1]$. It follows that the pointwise energy $e(f) = \frac{1}{2}\|Df\|^2$ is convex in t and so the global energy

$$E(f) = \int_X e(f)(x)\,dx$$

is also convex,

$$E((1-t)f_0 + tf_1) \leq (1-t)E(f_0) + tE(f_1) \ .$$

4.3.A. Equivariant maps . Suppose, we are given isometric actions of a group Γ on X and Y and we look at equivariant maps $f : X \to Y$. For example, X and Y may be Galois coverings of compact manifolds V and W and f is a lift of some map $\bar{f} : V \to W$. In general, we assume the action of Γ on X is discrete but me make no such assumption on Y. Since the point-wise energy (obviously) is Γ-invariant on X it descends to a function, also called e on $V = X/\Gamma$ (This may be a singular space if the action is non-free) and we define the Γ-energy by

$$E_\Gamma(f) = \int_V e(v)\ dv\ .$$

In particular, for the above covering example, $E_\Gamma(f)$ equals the energy of the underlying map $\bar{f} : V \to W$.

4.3.A_1. *The* Γ*-energy is convex for geodesic* Γ*-equivariant homotopics of maps.*

This is obvious with the convexity of e(f).

Remarks (a). The Γ-energy is most important in our study of the rigidity of groups Γ. Notice that this enery is finite if X/Γ is compact while the ordinary energy (obtained by the integration over X rather than X/Γ) is usually infinite for Γ-invariant maps.

(b) Let us look again at Γ-equivariant maps $X \to Y$ covering some maps $V \to W$. In this case $\pi_1(W) = \Gamma$ and Y is the universal covering of W. Then every continuous map $V \to W$ induces a Γ-covering X of V along with a Γ-invariant map $X \to Y$. This Γ-space X depends only on the homotopy class of the map $V \to W$ and homotopies of maps $V \to W$ correspond to Γ-equivariant homotopies of maps $X \to Y$. Notice that there is a slight ambiguity in there as the equality $\pi_1(W) = \Gamma$ needs a reference point $w_0 \in W$ but this causes no problem.

(c) If the action of Γ on Y is non-discrete we cannot go to the quotient space $W = Y/\Gamma$. Yet, if the action on X is free and discrete we can work with $V = X/\Gamma$ and the fibration $\bar{Z} \to V$, where $\bar{Z} = X \times Y/\Gamma$ for the diagonal action of Γ on $X \times Y$. This is the fibration with the fiber Y associated to the principal Γ-fibration (covering) $X \to V$. Now, Γ-equivariant maps correspond to sections $V \to \bar{Z}$. Every section $\bar{f} : V \to \bar{Z}$ has its point-wise energy $e(\bar{f})$ since, locally, $\bar{f}$ defines a map $V \to Y$ corresponding to the projection $X \times Y \to Y$. Then we have the global energy

$$E(\bar{f}) = \int_V e(\bar{f}).$$

which clearly equals the Γ-energy of the corresponding map $f : X \to Y$.

4.3.A_2. The basic relation between harmonicity and energy remain valid for equivariant maps. Namely, if *some smooth* Γ*-equivariant map* $f : X \to Y$ *gives minimum to* E_Γ *then* f *is harmonic*.

The proof is the same as in the absolute case. Notice that for the covering example the above harmonic maps corresponds to harmonic maps $V \to W$ minimizing the energy in a given homotopy class. In the general case we have harmonic sections $V \to \bar{Z}$ which are sometimes called *twisted harmonic maps.*

4.4. Uniqueness of harmonic maps. The convexity of the energy shows that the map f giving the minimum to E(f) is essentially unique. Before giving a precise statement we look at several examples.

4.4.A. let $V = S^1$. Then harmonic maps $S^1 \to W$ are just parametrized closed geodesics in W. If $K(W) < 0$ then there is a unique geodesic in every (non-trivial) homotopy class. (For the trivial class the harmonic maps are constant map $S^1 \to w \in W$. There is no uniqueness here as w may be any point in W). The only non-uniqueness here comes from the action of S^1 on this geodesic.

If $K(W) \leq 0$, then the above uniqueness may fail. For example, if $W = W_0 \times S^1$, then the geodesic maps homotopic to $w_0 \times S^1$ are parametrized by the points $w \in W$. Namely, there is a (unique) harmonic map $S^1 \to W$ in this homotopy class sending a given point $s \in S^1$ to w.

More generally, it may happen that W contains a totally geodesic submanifold isometric to $W_0 \times S^1$. Then one has a family of geodesics in W parametrized by this submanifold.

Now for any V, we may start with a harmonic map f of V into some W_1 and then we have many harmonic maps into $W_0 \times W_1$, namely the maps $f_w(v) = (w, f(v))$ for each point $w \in W_0$. Furthermore, we may have such a product inside of a larger manifold W and then the same non-uniqueness persists for maps $V \to W$.

4.4.B. The following theorem shows that the above examples essentially exhaust all possibilities of non-unique harmonic maps in the case where $V = X/\Gamma$ is compact. To simplify the matter, we assume the manifold Y is real analytic and suppose there exist two Γ-homotopic Γ-equivariant harmonic maps f_0 and $f_1 : X \to Y$, where X is connected and X/Γ is compact.

4.4.B'. Theorem. *The images of the maps* f_0 *and* f_1 *lie in a complete totally geodesic submanifold* $Y' \subset Y$, *such that* Y' *isometrically splits by* $Y' = Y_0 \times \mathbb{R}$ *and* $f_1 : X \to Y'$ *is obtained from* f_0 *by an* $\mathbb{R}$*-translation of* Y' *that is the map* $(y_0, t) \mapsto (y_0, t + t_0)$ *for some* $t_0 \in \mathbb{R}$.

This result is well known and easily follows from the elementary geometry of $K \leq 0$ (see [B-G-S]).

Remark. Notice that we do not assume our maps are energy minimizing. In fact one can easily show that harmonic maps are necessarily energy minimizing if $V = X/\Gamma$ is compact.

4.5. On the existence of harmonic maps. A natural approach of constructing a harmonic maps $f : V \to W$ in a given homotopy class $\mathfrak{F}$ consists in taking an energy minimizing sequence of maps $f_i \in \mathfrak{F}$ and taking a sublimit of f_i for f. (Recall that "minimizing" means $E(f_i) \to \inf_{f \in \mathfrak{F}} E(f)$ and "sublimit" is the limit of a subsequence). Let us indicate some examples showing what can go wrong.

4.5.A_1. Let $V = W = S^n$ and $f_t : W \to W$ be the north pole south pole tranformations. Namely f_t fix the north pole $s_+ \in S^n$ and act on $\mathbb{R}^n = S - \{s_+\}$ by $s \mapsto e^t s$, where $\mathbb{R}^n$ is identified with $S^n - \{s_+\}$ by the stereographic projection. The maps f_t are *conformal* and a trivial computation shows that

(1) if $n = 1$ then $E(f_t) \to \infty$ for $t \to \infty$.

(2) if $n = 2$ then $E(f_t) = E(f_o)$ for all t

(3) if $n = 3$ then $E(f_t) \to 0$ if $t \to \infty$.

This example indicates invalidity of the variational approach as the maps f_t for $t \to +\infty$ converge almost everywhere (in fact, everywhere except at the south pole $= 0 \in \mathbb{R}^n$) to the constant map $S^n \to s_+ \in S^n$, which is (though harmonic) not homotopic to f_t. (One sees a more instructive picture by looking at the graphs of f_t that are maps $S^n \to S^n \times S^n$ which are harmonic for $n = 2$).

Also notice that the family is pointwise C^1-*divergent* as the energy density $e(f_t) = \frac{1}{2} \|Df_t\|^2$ blows up at the south pole.

4.5.A_2. The discontinuity phenomenon of the above example can not happen for $K(W) \leq 0$ (see below) but one can have another kind of degeneration, where the minimizing maps run to infinity all-to-gether. To see that let $W = H^2_{\mathbb{R}}/\mathbb{Z}$ be the standard *cusp* , that is the group $\mathbb{Z} \subset$

Iso $H^2_{\mathbb{R}}$ consists of *parabolic* elements. This means, there is a *single* point $s \in S_\infty(H^2_{\mathbb{R}})$ fixed under this group. Topologically, $W = S^1 \times \mathbb{R}$ and the metric is $e^{-t}ds^2 + dt^2$.

Now we look at the maps $f_t : S^1 \to W$ given by $f_t(s) = (s, t)$ and observe that $E(f_t) \to 0$ for $t \to \infty$. In fact, even the pointwise energy $e(f_t) = \frac{1}{2} \|Df_t\|^2$ exponentially decays as $t \to \infty$. In this case no limit (or sublimit) of f_t exists and there is no harmonic map in the homotopy class of f_t .

4.5.B. Smoothing . In order to prevent a blow up of $e(f)$ (as in 4.5 A_1) one may try to regularize a given minimizing sequence f_i by applying some smoothing operators to f_i. The smoothed maps f_i^{sm} must have roughly the same energy in order to have

$$\lim_{i \to \infty} E(f_i^{sm}) = \lim_{i \to \infty} E(f_i),$$

and the energy density should be uniformly bounded,

$$\sup_{v \in V} e\,(f_i^{sm})\,(v) \le \text{const.}$$

Let us explain why the condition $K(W) \le 0$ makes the existence of a smoothing possible.

4.5.B_1. The center of mass. Let Y be a complete simply connected manifold with $K(Y) \le 0$ (it will be the universal covering of W) and μ be a finite measure on Y. Then we define the *center* of μ, denoted $\bar{\mu}$ or $\int_Y d\mu$ as the (unique !) minimum point of the function d^2_μ, $x \in Y$, obtained by integrating away the second argument in $\text{dist}^2(x, y)$ with respect to μ. Notice, that because of $K \le 0$ each function $d^2_y(x) = \text{dist}^2(x, y)$ is strictly (geodesically) convex on y and so the same is true for d^2_μ. This insures the uniqueness of the minimum point of d^2_μ (x).

Example. Let μ be supported at two points y_o and y_1 in Y with some weights p_o and $p_1 = 1 - p_o$. Then $\bar{\mu}$ equals the convex combination $p_o y_o + p_1 y_1$ as defined in 4.3.

4.5.B_2. Averaging families of maps. Let $f_t : X \to Y$, be a family of maps where t runs over some space T with a given finite measure ν on T. Then for every $x \in X$ we take the push-forward measure μ_x for the map $t \mapsto (f_t)\, f_t(x) \in Y$ and then take the center $\bar{\mu}_x \in Y$ for every $x \in X$. Thus we get the averaged map $f = \bar{f}_t : X \to Y$.

The convexity properties of Y with $K(Y) \leq 0$ (which we have met already several times) show that the energy density $e(f)$ at each point $x \in X$ is bounded by the μ-average of the energies of f_t at x. This allows the smoothing of maps into Y in the following.

Flat example. Let V be the flat n-dimensional torus and μ_v be the Lebesgue measure of a ball $B_v(\varepsilon) \subset V$ around $v \in V$ of a small radius $\varepsilon > 0$ such that this $B_v(\varepsilon)$ is isometric to the Euclidean ball $B_0(\varepsilon) \subset \mathbb{R}^n$. Then for every map $f : V \to W$ we have the *smoothed map* $f^\varepsilon : V \to W$ defined by the averaging the push forward of these measures, that is

$$f^\varepsilon(v) = \overline{f_*(\mu_v)},$$

where the averaging is obtained by first lifting f to the universal covering $Y \to W$ and then projecting back to Y. Let us show this smoothing does what we want. Namely

(A) $E(f^\varepsilon) \leq E(f)$

(B) $\sup_{v \in V} e(f^\varepsilon(v)) \leq \text{const}_\varepsilon\ E(f)$

Proof. Let $B(\varepsilon) \subset V = T^n$ be the ε-ball around the identity element in the torus thought of as an additive group. Then we consider the maps $f_t(v) = f(v + t)$ for $v \in V$ and $t \in B(\varepsilon)$ and let μ be the Lebesgue measures on $B_0(\varepsilon)$. Clearly,

$$\bar{f}_t = f^\varepsilon$$

and then (A) and (B) immediatly fall through.

4.5.C. The heat flow. The idea of the extension of the above to non-flat manifolds V consists in taking an "infinitely small" ε and performing the smoothing procedure "infinitely many" times. To keep track of what happens it is easier to replace the measures on the ε-balls by the heat flow on V. Namely, we denote by $\mu(v, t)$ the result of the diffusion of the δ-measure $\delta(v)$ on V at the moment t. In other words $\mu(v, t)$ is the measure on V whose density $\varphi(v')$ is the heat kernel $h(v, v', t)$ that is the solution of the heat equation on V,

$$\Delta_{v'} h = \frac{dh}{dt}$$

with the initial condition

$$h(v', 0) = \delta(v).$$

For example if $V = \mathbb{R}^n$ then

$$h(v, v', t) = (4\pi t)^{-\frac{n}{2}} e^{-\frac{|v-v'|^2}{2t}}.$$

Recall that the *heat flow* defined by h has the following *semigroup property* schematically expressed by

$$h(t_1) * h(t_2) = h(t_1 + t_2)$$

and saying in effect that the measure $\mu(v, t_1 + t_2)$ is obtained by t_2-diffusing each point $v \in V$ with the weight $\mu(v, t_1)$. Namely

$$h\left(v, v', t_1 + t_2\right) = \int h\left(v, v'', t_1\right) h\left(v'', v', t_2\right) dv''$$

Now for maps $f : V \to W$ we define the smoothing operator $H_\varepsilon f$ by

$$H_\varepsilon f(v) = \bar{f}_*(\mu(v, \varepsilon))$$

and then for a given t we apply this operator n-times for $\varepsilon = t/n$, and get

$$H_\varepsilon^n f = H_\varepsilon(H_\varepsilon \dots (H_\varepsilon f) \dots).$$

(Here, as earlier, the averaging refers to the universal covering Y of W). The following result is due to Eells and Sampson (see [Eel-Sam]).

4.5.C_1. Theorem. *If* V *is compact and* $K(W) \leq 0$, *then for every smooth map* f *the sequence* $H_\varepsilon^n f$ *converges to a smooth map, denoted*

$$f(v, t) = \mathcal{H}_t(f),$$

and called the (non-linear) heat flow or diffusion of $f = f(v, 0)$. *Furthermore, this flow satisfies the heat equation*

$$\frac{df(v,t)}{dt} = \Delta f(v, t)$$

and the initial condition

$$f(v, 0) = f(v)$$

4.5.C_1'. Remarks (a) As we know (*) says that $f(v, t)$ is the minus gradient flow for the energy and so $E(f(v, t))$ goes down with $t \to \infty$.

(b) *The theorem remains valid for a* Γ*-equivariant map* $f : X \to Y$, *provided* X/Γ *is compact.*

4.5.D. Bochner's formula and apriori estimation. A standard method of proving the existence of regular solutions of P.D.E. is by first establishing a conditional result giving a bound on certain derivatives of solutions *assuming* a solution exists. Let us indicate such a bound for harmonic maps (and later on for the heat flow) $V \to W$. The basic tool here is the following Bochner formula of Eells and Sampson which holds true for all C^2-smooth harmonic maps $f : V \to W$,

$$\Delta e(f) = \|\mathrm{Hess} f\|^2 + Q_V - Q_W \qquad (*)$$

where Q_V and Q_W are the following expressions involving Df and the curvatures of V and W. First

$$Q_V = \mathrm{Ricci}(Df, Df),$$

where the Ricci tensor in V is thought of as a quadratic form on $T(V)$ which naturally defines with the Riemannian metric in W a quadratic form on the bundle $\mathrm{Hom}(T(V), f^*T(W))$.

Then this extended Ricci applies to $Df \in \mathrm{Hom}(T(V), f^*T(W))$ and defines Q_V.

Now, to define Q_W we take a frame of orthonormal vectors $\tau_1 \dots \tau_n$ at every point v in V and we set

$$Q_W(v) = \sum_{i,j} \left\langle R\left(D\tau_1, D\tau_2\right) D\tau_3, D\tau_4 \right\rangle_W,$$

where R denotes the curvature tensor of W at $w = f(v)$ and $D = Df$

Notice that $Q_W \le 0$ if $K(W) \le 0$.

We don't prove (∗) (see [Eel-Sam]) but rather explain its meaning. First of all, if V and W are flat, then the formula reduces to

$$\Delta e(f) = \|\mathrm{Hess} f\|^2 \qquad (*_0)$$

which is a well known elementary (Euclidean) identity. The major consequence of this is the inequality

$$\Delta e(f) \geq 0 \qquad (*+)$$

which says that the energy density is a subharmonic function on V.

Another important corollary of $(*_0)$ for *closed* manifold V reads

$$\int_V \|\mathrm{Hess}\, f\|^2 = 0 \qquad (*i)$$

Hence $\mathrm{Hess}\, f = 0$ and so *every harmonic map is geodesic*.

Now if V is flat, then the condition $K(W) \leq 0$ reinforces $(* +)$ as the term $-Q_W$ in (∗) is positive. The same happen to (∗ i) and so we see that *every harmonic map of a compact flat manifold* V *into* W *with* $K(W) \leq 0$ *is geodesic.*

This can be also seen with the averaging on V which works for all flat manifolds as well as for tori (see 4.5.B_2). The averaged map satisfies $E(f^\varepsilon) < E(f)$ unless f is a geodesic map, as a simple convexity argument shows (compare 4.4.B'). Thus every energy minimizing (harmonic) map must be geodesic.

Now we allow non-flat V and observe that $K(W) \leq 0$ implies via (*) the following inequality on $e(f) = \|Df\|^2$,

$$\Delta\, e(v) \geq Q_V \geq -C(v)\, e(v) \qquad (+)$$

where $-C(v)$ is the lower bound for Ricci_V at v. If V is compact (+) implies the bound

$$\sup_{v \in V} e(v) \leq \mathrm{const}\, E(f) = \int_V e(v)\, dv\ .$$

Then this point-wise bound on $\|Df\|$ can be extended with little extra effort to the higher order derivatives (see [Eel-Sam]).

4.5.D_1. Bochner for the heat flow. Now let f(v,t) be the solution to the heat equation $\frac{df}{dt} = \Delta f$. Then the Bochner formula generalizes to

$$\Delta e(v) = \|\mathrm{Hess} f\|^2 + Q_V - Q_W + \frac{de(f)}{dt} \qquad (**)$$

where f is viewed as function in v with t as a parameter.

This implies, for $K(W) \leq 0$ in (+) that

$$\frac{de}{dt} \leq \Delta e + C(v)e$$

which leads by a simple argument to the following important conclusion.

4.5.D_1'. *If* V *is compact then the map* f(v, t) *satisfies for every* $t \geq 1$ *the following point-wise bound on* $e(f) = \frac{1}{2}\|Df\|^2$,

$$e(v, t) \leq \text{const } Ef(v, 0),$$

where

$$E f(v,0) = \int_V e(v,0)\, dv,$$

for $e = e(f)$. *Furthermore, a similar bound remains valid for* Γ*-invariant maps* $X \to Y$ *if* $V = X/\Gamma$ *is compact.*

This shows that the diffusion (heat flow) operator $f \mapsto \mathcal{H}_t(f)$ (compare 4.5.C_1) for each $t \geq 1$ does smooth maps $f : V \to W$. Moreover, one can prove that all derivatives of $f(v, t) = \mathcal{H}_t(f)$ are controlled by E(f) and so $\mathcal{H}_t$ provides us with a perfect smoothing. Besides $\mathcal{H}$ *decreases* the energy of maps (see 4.5.C_1') and so it is well adapted to the minimization process. In fact one can use $\mathcal{H}_t(f) = f(v, t)$ as a minimizing family. Namely, the energy $E\mathcal{H}_t(f)$ converges to Inf E(f) over all f in a given homotopy class and $\Delta \mathcal{H}_t(f) \to 0$ for $t \to \infty$. This is shown with yet another Bochner formula, this time for

$$k = \|\Delta f\|^2 = \left\|\frac{df}{dt}\right\|^2$$

satisfied by $f(v, t) = \mathcal{H}_t f(v)$.

This formula shows that (see [Eel-Lem]$_1$, p. 24)

$$\frac{dk}{dt} \le \Delta k \qquad (\square)$$

which immediately implies for compact V the asymptotic bound

$$\lim_{t \to \infty} k(v, t) \le K(t_0) = \int_V k(v, t_0)\, dv ,$$

for every fixed t_0. Then by integrating ($\square$) over W we see that K(t) is decreasing in t. On the other hand

$$K(t) = -\frac{dEf(v,t)}{dt},$$

since $\Delta f = -\operatorname{grad} Ef$. It follows that $K(t) \to 0$ for $t \to \infty$, because Ef remains positive for all t, and so $k(v, t) \to 0$ as well. Thus $\Delta f(v, t) \to 0$ for $t \to \infty$. Since all derivatives of f(v, t) are bounded for $t \to \infty$ there are only two alternatives .

(A) The maps f(v, t) subconverge for $t \to \infty$ to a harmonic map f in the homotopy class of f(v, 0)

(B) The maps f move $f(V) \subset W$ away to infinity for $t \to \infty$. That is

$$\operatorname*{Inf}_{v \in V} \operatorname{dist} (w_0, f(v, t)) \underset{t \to \infty}{\longrightarrow} \infty ,$$

for a fixed point $w \in W$. (Compare example 4.5.A_2).

Notice that in the course of this movement the norm $\|Df\|$ remains uniformly bounded. This implies the uniform bound on the length of loops representing the image $f_*(\pi_1(V)) \subset \pi_1(W)$ as these loops move to infinity together with $f(V) \subset W$.

<u>Example</u>. (See [Don]). Let $K(W) \le -\kappa < 0$. Then in the case (B) the group $f_*(\pi_1(V)) \subset \pi_1(W)$ acting on the universal covering Y of W is *parabolic* : it fixes a unique point on the ideal boundary of W. In particular if

$$-\infty < -\kappa' \le K(W) \le -\kappa < 0,$$

then the group $f_*(\pi_1(V))$ contains a nilpotent subgroup of finite index. This follows from the above remark about the loops and basic geometry of $K < 0$ (see [B-G-S]).

Another (more important) case where (B) can be ruled out was pointed out by Corlette (see $[\mathrm{Cor}]_2$).

4.5.D_2. *Let* $W = Y/\Gamma$ *where* Y *is a symmetric space. If the Zariski closure of the subgroup* $f_*(\pi_1(V))$ *in the real algebraic group* Iso Y *(containing* $\pi_1(W) \supset f_*(\pi_1(V))$ *is semisimple then* (B) *is impossible and so* $f(v, 0)$ *is homotopic to a harmonic map. Furthermore this remains valid for the general* Γ*-invariant case where* X/Γ *is compact and* Γ *the Zariski closure of* Γ in Iso Y *is semisimple* (see $[\mathrm{Cor}]_1$).

Remark. The existence theorem for compact W is due to Eells and Sampson. Although the proof of the non-compact generalization uses the same analytic techniques, it provides us with a by far more powerful tool for the (super) rigidity problem.

4.6. Special Bochner formulas

4.6.A. We come back to the Riemannian geometric formulation of the superrigidity problem, alluded to in 2.9.A.

In this problem, representations of lattices into compact groups are neglected, as reflected in the following definition.

Definition : Let G, G' be semi simple Lie groups, Γ a subgroup of G, $\rho : \Gamma \to G'$ a homomorphism.

A virtual extension of ρ to G is the following data :

- an extension $1 \to K_1 \to G_1 \xrightarrow{\pi} G \to 1$ of G by a compact Lie group K_1
- an embedding $i : \Gamma \to G_1$ such that $\pi \circ i = \text{identity}$
- a homomorphism $h : G_1 \to G'$ such that

$$\rho = h \circ i .$$

Note that, since the group of outer automorphisms of K_1 is discrete, an extension G_1 as above is a direct product, up to finite groups.

Proposition : *Let* X, X' *be symmetric spaces with negative Ricci curvature. Let* G, G' *be their isometry groups,* $\Gamma \subset G$ *a subgroup,* $\rho : \Gamma \to G'$ *a homomorphism. Then* ρ *admits a virtual extension to* G *if and only if there exists a totally geodesic, Γ-equivariant map* $X \to X'$.

Proof.

1) *The virtual extension arising from a totally geodesic map.* Let $f : X \to X'$ be totally geodesic. If X is irreducible, then f is a homothety. Indeed, the pulled back metric

$$f^* g_{X'}$$

is parallel on X. If X is not irreducible, it splits as a riemannian product $X = X_1 \times X_2$, f factors through the projection $pr_1 : X \to X_1$ and a totally geodesic embedding $f' : X_1 \to X'$. The isometry group of $Y = f'(x_1)$ is a direct factor G_f in $G = G_f \times H$.

Let $G_2 \subset G'$ be the subgroup that fixes Y globally. Then G_2 acts by isometries on Y, which gives a compact extension

$$I \to K_1 \to G_2 \to G_f \to 1$$

then $G_1 = G_2 \times H$ is a compact extension of G with a homomorphism

$$h : G_1 \xrightarrow{pr_1} G_2 \subset G'.$$

If f is Γ-equivariant, then $\rho(\Gamma) \subset G_2$. Let $pr_2 : \Gamma \to G = G_f \times H \to H$ be the projection to the H factor then

$$i = \rho \times pr_2 = \Gamma \to G_2 \times H = G_1$$

embeds Γ into G_1, $\pi \circ i = \text{identity}$, $\rho = h \circ i$ so that h is a virtual extension of ρ.

2) *The totally geodesic orbit of a virtual extension.* Let $h : G_1 \to G'$ be a virtual extension of ρ. Then

$$G_2 = h(G_1)$$

is a reductive subgroup of G'. We use a theorem of G.D. Mostow ([Mos]$_5$, theorem 6) : G_2 is left invariant by some Cartan involution θ of G'. The involution θ acts on X' as the geodesic symmetry through a point y. Let Y be the orbit of y under G_2, and II its second fundamental form. We see that

$$\theta^* II_y = - II_{\theta(y)}$$

and so $II_y = 0$, $II = 0$ everywhere by homogeneity, and Y is totally geodesic.

Let $\pi = G_1 \to G$ be the extension. Fix an origin x_o in X, with stabilizer K. Then $h(\pi^{-1}K)$ is a compact subgroup of isometries of Y. Since Y is a symmetric space with $K \leq 0$, $h(\pi^{-1}K)$ fixes some point y_o in Y. As a consequence, h descends to a homomorphism $\ell : G \to \mathrm{Isom}(Y)$ and an equivariant map $f = X \to Y$ can be defined by

$$f(gx_o) = \ell(g)y_o .$$

The quadratic form $f^* g_y$ on X is G-invariant thus parallel with respect to g_x, i.e., f is totally geodesic.

Finally we check that f is Γ-equivariant :

for $\quad g \in G, \gamma \in \Gamma, \phi(\gamma g x_o) = \ell(\gamma)\, \ell(g)\, y_o$

$$= h(i(\gamma))\, f(x)$$

$$= \rho(\gamma)\, f(x) .$$

4.6.B. In view of 4.6.A. and 4.5, the superrigidity problem translates into the following terms :

Let f be a (twisted) harmonic map $V \to W$ where V is compact. Assume the curvature of W is sufficiently negative (e.g. W is a locally symmetric space of non compact type). For special V (e.g. cerain types of locally symmetric spaces), can one conclude that f satisfies some extra equations ? Ultimately, that f is totally geodesic ?

Let us introduce suggestive notations. Let $E = f^* TW$ denote the pull back of the tangent bundle of W . It is a vector bundle on V, equipped with a metric and a connection. We view

$$\alpha = df \in C^\infty(T^*V \otimes E)$$

as a vector valued 1-form on V. The connection on V and E determine a connection ∇ on $T^*V \otimes E$. Then the Hessian of f is $\nabla\alpha$, a vector valued bilinear form on V, compare $4.2.A_2'$.

Bilinear forms on V splits as

$$T^*V \otimes T^*V = \Lambda^2 \oplus \mathbb{R}g_V \oplus S_0^2$$

into skew symmetric forms, the metric, and tracefree symmetric forms.

Accordingly, the Hessian $\nabla\alpha$ splits as

$$\nabla\alpha = d\alpha + d^*\alpha + (\nabla\alpha)^{S_0^2}$$

and

$\alpha = df$ implies $d\alpha = 0$

f is harmonic if and only if $d^*\alpha = 0$

f is totally geodesic if and only if $\nabla\alpha = 0$.

Why should $d\alpha = 0$ and $d^*\alpha = 0$ imply that the rest of ∇a vanishes ? This is where compactness of V and Bochner formulas play a role.

$4.6.B_1$. Bochner's vanishing theorem for Riemannian manifolds with nonnegative Ricci curvature.

We illustrate the method of Bochner formulas of an example. Let V be a compact Riemannian manifold, and α a harmonic 1-form on V. Then the Bochner formula states that , if $e = \frac{1}{2}\|\alpha\|^2$, then

$$\Delta e = \|\nabla\alpha\|^2 + \mathrm{Ricci}(\alpha, \alpha)$$

where the Ricci tensor is thought of as a quadratic form on T^*V, see [Boc-Yan].

If we assume that the Ricci curvature of V is nonnegative, we conclude that $\nabla\alpha = 0$ everywhere. Indeed, on a compact manifold, the Laplacian of a function integrates to zero.

The above argument can be generalized to include vector valued forms. The extra data is a vector bundle E over V equipped with a metric and a connection. For an E-valued 1-form α, the Bochner formula becomes

$$\Delta e = \|\nabla\alpha\|^2 + \mathrm{Ricci}(\alpha, \alpha) - \langle R^E, \alpha \wedge \alpha\rangle$$

where the extra term involves the curvature of E,

$$R^E \in \Lambda^2 T^*V \otimes \mathrm{End}\, E$$

contracted with $\alpha \wedge \alpha \in \Lambda^2 T^*V \otimes E \otimes E$.

For flat bundles, the formula is unchanged.

When f is a map $V \to W$ and $E = f^* TW$, $\alpha = df$, we recover the Eells-Sampson formula used in 4.5.D. Indeed, if $R^W \in \Lambda^2 T^*W \otimes \mathrm{End}\, TW$ is the curvature tensor of W, the curvature of E is

$$R^2 = f^* R^W = (\Lambda^2 \alpha)^* R^W$$

and

$$\langle R^E, \alpha \wedge \alpha\rangle = Q_W$$

(in the notation of 4.5.D) is nonnegative provided W has nonpositive sectional curvature. We infer that every harmonic map from V to W is totally geodesic provided :

W has nonpositive sectional curvature ;

V has nonnegative Ricci curvature.

This theorem is of no use when V is locally symmetric of non compact type, since then $\mathrm{Ricci}_V < 0$. The point of the special formulas we explain next is that the sign of the curvature of V does not interfere. What is required instead is that V have special holonomy.

4.6.B_2 Kähler manifolds and pluriharmonic maps.

A Kähler manifold of real dimension 2n has holonomy a subgroup of U(n). Under U(n), trace free symmetric bilinear forms split as

$$S^2_o = S^{2,0+0,2} \oplus S^{1,1}_o$$

where the summands are the eigenspaces of the involution

$$b(u, v) \mapsto b(Ju, Jv)$$

(J denotes the complex structure).

Thus, on a Kähler manifold, there is a new equation, i.e., a new linear condition that one can impose on the second derivatives of a map.

4.6.B_3. Definition : Let V be a Kähler manifold, W a Riemannian manifold. Say a map $f : V \to W$ is *pluriharmonic* if its Hessian $\nabla\alpha$ is of type $S^{2,0+0,2}$, i.e., its components on Λ^2 and $S^{1,1}$ vanish.

When V has real dimension 2, pluriharmonic means harmonic. In higher dimensions, f is pluriharmonic if and only if its restriction to every (germ of) holomorphic curve in V is harmonic. As a consequence, pluriharmonicity does not depend on the particular Kähler metric on V, only on the complex structure of V. In fact, a real valued function on V is pluriharmonic if and only if it is locally the real part of a holomorphic function. Thus the familiar Hodge therorem

$$H^1(V, \mathbb{C}) = H^{1,0} \oplus H^{0,1}$$

says that, on a compact Kähler manifold, every harmonic 1-form is pluriharmonic.

The main step in Y.T.Siu's 1980 theorem is an extension to vector valued 1-forms of this property.

4.6.C. Negative curvature operator.

We describe now the kind of negativity required on W.

Definition : The curvature tensor of a Riemannian manifold can be viewed as a quadratic form on exterior 2-forms. It is then called the *curvature operator* $\check{R}$.

Example : Symmetric spaces of non compact type have $\check{R} \leq 0$.

Proof : Let $X = G/K$ be symmetric. Let $\underline{p}$ be the orthogonal complement of $\underline{k}$ in $\underline{g}$ with respect to the Killing form B. B is positive definite on $\underline{p}$, it gives rise to the Riemannian metric on X. Consider the Lie bracket as a map

$$[\ ,\ ] = \Lambda^2 \underline{p} \to \underline{k} .$$

then the curvature operator of X - a quadratic form on $\Lambda^2 \underline{p}$ - is the pull back of $B_{|\underline{k}}$ by $[\ ,\ ]$. As a consequence, it is non positive, and its kernel coincide with kernel of $[\ ,\ ]$.

4.6.D. A vanishing theorem for harmonic maps of manifolds with special holonomy

In 1980, Y.T. Siu found a generalization of the Hodge theorem $H^1(V, \mathbb{C}) = H^{1,0} \oplus H^{0,1}$ for harmonic maps of Kähler manifolds to hermitian locally symmetric spaces, and gave striking consequences, including a new proof of Mostow rigidity for hermitian locally symmetric spaces (see below 4.6.F_2). Later, J. Sampson observed that Siu's method could be extended to the case where the range is not Kähler. Recently, K. Corlette has included Siu's theorem into the wider framework of harmonic maps from manifolds with special holonomy.

4.6.D_1. Theorem. [Siu], [Sam], $[\mathrm{Cor}]_1$. *Let* V *be a compact Riemannian manifold that admits a parallel form* ω. *Let* W *be a Riemannian manifold with non positive curvature operator. Then every (twisted) harmonic map* $f : V \to W$ *satisfies*

$$d^*(df \wedge \omega) = 0 .$$

This equation in a generalisation of the notion of pluriharmonicity. Indeed, let us compute $d^*(df \wedge \omega)$. This is a linear expression in $\nabla\alpha$, $\alpha = df$, and we can assume α is scalar, V is flat. Let e_i be an orthonormal frame

$$d^*(\alpha \wedge \omega) = -\sum_i e_i^* \lrcorner \nabla_{e_i}(\alpha \wedge \omega)$$

$$= -\sum_i e_i^* \lrcorner \left(\nabla_{e_i}\alpha\right) \wedge \omega$$

$$= \left(-\sum_i e_i^* \lrcorner \nabla_{e_i}\alpha\right) \wedge \omega - \sum_i \left(\nabla_{e_i}\alpha\right) \wedge \left(e_i^* \lrcorner \omega\right)$$

$$= \left(d^*\alpha\right)\omega - \mathrm{ad}_{\nabla\alpha}(\omega);$$

where we view $\nabla\alpha$ as an endomorphism of TV and the natural action of $\nabla\alpha$ on TV is extended as a derivation of the exterior algebra.

When V is a Kähler manifold and ω its Kähler form, we identify 2-forms with endomorphisms via the metric. Then ω becomes the complex structure J, and $\mathrm{ad}_{\nabla\alpha}(\omega)$ becomes

$$\nabla\alpha \circ J - J \circ \nabla\alpha, \text{ i.e. } \mathrm{ad}_{\nabla\alpha}(\omega) = (\nabla\alpha)^{1,1}.$$

Thus, for a harmonic 1-form α,

$$d^*(\alpha \wedge \omega) = 0 \Leftrightarrow \nabla\alpha \in S^{2,0+0,2}$$

$$\Leftrightarrow \alpha \text{ is pluriharmonic}.$$

Proof of vanishing theorem $4.6.D_1$

The idea is that exterior multiplication with a parallel form commutes with the Laplacian, see[Che]. It is convenient to use first Clifford multiplication instead of exterior multiplication.

$4.6.D_3$. The Clifford Formalism (see [Law-Mic]).

Clifford multiplication is an associative algebra structure on $\Lambda^* T^* V$, the space of exterior forms on a Euclidean vector space. Recall that interior multiplication of exterior forms α and β is defined by the following identity for all exterior forms γ,

$$<\alpha \lrcorner \beta, \gamma> = <\beta, \alpha \wedge \gamma> ;$$

i.e., interior multiplication by α is the adjoint of exterior multiplication by α.

For a 1-form α and p-form ω, one defines the Clifford product

$$\alpha.\omega = \alpha \wedge \omega + \alpha \lrcorner \omega. \qquad (*)$$

This extends to an associative multiplication on $\Lambda^* T^* V$.

The *Dirac operator* $\mathcal{D}$, a first order operator on differential forms, is the composition of the covariant derivative : $C^\infty(\Lambda^* T^* V) \to C^\infty(T^* V \otimes \Lambda^* T^* V)$ with Clifford multiplication :

$$T^* V \otimes \Lambda^* T^* V \to \Lambda^* T^* V$$

In other words, given an orthonormal basis e_i of TV, with dual basis e_i^* ,

$$\mathcal{D}\beta = \sum_i e_i^* . \nabla_{e_i}\beta$$

Formula (*) shows that

$$\mathcal{D} = d + d^*$$

Let ω be a parallel p-form on V.

Denote by C the operator on differential forms defined by right Clifford multiplication with ω . Clearly

$$\mathcal{D}C\beta = \sum_i e_i^* . \nabla_{e_i}(\beta.\omega) = C\mathcal{D}\beta$$

as follows form the associativity of Clifford multiplication.

Thus the Laplacian $\Delta = (d + d^*)^2 = \mathcal{D}^2$ commutes with C. Splitting

$$C = C^p + C^{p-2} + C^{p-4} + \dots ,$$

according to the degree of forms, we see that C^p, exterior multiplication with ω, commutes with the Laplacian : $[\Delta, C^p] = 0$.

Some of the more basic identities obtained when splitting

$$[\mathcal{D}, C] = 0$$

according to the degree of forms,

$$[d, C^p] = 0 \, ;$$

$$[d, C^{p-2}] + [d^*, C^p] = 0 \, ;$$

and so on, are classical.

The first identity, $d(\alpha \wedge \omega) = (d\alpha) \wedge \omega$, is very general. In the Kähler case, where $C^p = C^2$ is usually denoted by L, the second identity is the familiar

$$[d^*, L] = d^c.$$

All the above formulas extend without change to vector valued forms α (and scalar ω).

4.6.D_4. The vanishing theorem, scalar case :

On a compact manifold V, let ω be a parallel form and α a harmonic 1-form. Using identity

$$[d + d^*, C] = 0$$

we integrate by parts :

$$\int |dC\alpha|^2 = -\int \langle dC\alpha, d^*C\alpha \rangle$$

$$= -\int \langle d^2C\alpha, C\alpha \rangle$$

$$= 0$$

and thus $dC\alpha = d^*C\alpha = 0$. In particular,

$$d^*C^p\alpha = d^*(\alpha \wedge \omega) = 0$$

as announced.

4.6.D_5. <u>The vanishing theorem, general case :</u>

The point is to compute the sign of $<d^2C\alpha, C\alpha>$ when $\alpha = df$, f a map $V \to W$. Let R^W be the curvature tensor of W. It defines a quadratic form on $TW \otimes TW$ (which is non zero only on Λ^2TW), which combined with the metric on TV gives rise to a quadratic form Q on

$$\Lambda^* T^* V \otimes TW \otimes TW$$

<u>Claim :</u>

$$<d^2C\alpha, C\alpha> = 2Q(\alpha.\alpha.\omega)$$

Clearly, this proves the theorem, since Q has the same sign as $\check{R}^W$.

This an algebraic computation. Splitting into degrees of forms, we find

$$<d^2C\alpha, C\alpha> = <d^2C^{p-2}\alpha, C^p\alpha> = <d^2(\alpha \llcorner \omega), \alpha \wedge \omega>$$

Noω d^2 is exterior multiplication with the curvature of the bundle f^*TW, that is $R^W \circ (\alpha \wedge \alpha)$. We get

$$<d^2C\alpha, C\alpha> = <R^W((\alpha \wedge \alpha) \wedge (\alpha \llcorner \omega)), \alpha \wedge \omega> .$$

Using symmetries of R^W , the Bianchi identity $d^2\alpha = 0$, and symmetries of the expression $<\alpha \wedge \alpha \wedge (\alpha \llcorner \omega), \alpha \wedge \omega>$, this becomes

$$2 < R^W(\alpha \wedge \llcorner(\alpha \llcorner \omega), \alpha \llcorner(\alpha \wedge \omega)>$$

$$= 2Q(\alpha.\alpha.\omega) .$$

4.6.E. <u>Negative complex curvature</u>

In the Kähler case, i.e. V is Kähler, ω its Kähler form, the curvature assumption on W in theorem 4.6.D_1 can be weakened. Indeed, $\alpha \llcorner \omega = J\alpha$ and we find that the quadratic form Q is the pull back of the quadratic form $\check{R}^W$ on Λ^2TW by the map

$$\alpha \wedge J\alpha = \Lambda^2TV \to \Lambda^2TW$$

Given a unitary basis e_j of TV, denote by

$$Z_j = \alpha(e_j) + i\alpha(Je_j) \in TW \otimes \mathbb{C}$$

then

$$2Q(\alpha.\alpha.\omega) = 2Q(\alpha \wedge J\alpha)$$

$$= \sum_{i<j} < \check{R}^W(Z_i \wedge Z_j), \bar{Z}_i \wedge \bar{Z}_j>$$

thus for the previous argument to apply, it is sufficient to assume that

$$< \check{R}^W(Z_1 \wedge Z_2, \bar{Z}_1 \wedge \bar{Z}_2> \leq 0$$

for all *complex* tangent vectors $Z_1, Z_2 \in TW \otimes \mathbb{C}$. When this condition is satisfied, one says that W has *nonpositive complex curvature*.

Under this assumptions (V compact Kähler, W has nonpositive complex curvature), theorem $4.6.D_1$ is due to Sampson ([Sam]), elaborating on Siu's work where W was assumed Kähler as well.

4.6.F. Non linear equations satisfied by harmonic maps.

The proof of Theorem $4.6.D_1$ relies on the identity

$$\int_V \left|dC\alpha\right|^2 - 2Q\left(\alpha.\alpha.\omega\right) = 0$$

where Q is non positive. Thus with the linear equations $dC\alpha = 0$ comes the nonlinear condition

$$Q(\alpha.\alpha.\omega) = 0 .$$

For a Kähler domain V, the condition states that for all $Z_1, Z_2 \in df(T^{1,0}V)$, $Z_1 \wedge Z_2 \in$ Ker $\check{R}^W \otimes \mathbb{C}$, and has the following interpretation : the Levi Civita connection endows the bundle f^*TW with the structure of a holomorphic bundle. Furthermore the (1, 0)-component of df is a holomorphic 1-form with values in f^*TW (on other formulation of the fact that f is pluriharmonic). When the range W is a symmetric space G/K, the nonlinear condition says that the subspace $df(T^{1,0}V) \subset \underline{p} \otimes \mathbb{C}$ is an abelian subalgebra (compare 4.6.C). This has been analyzed by Siu, Carlson-Toledo.

4.6.F_2 Proof of Mostow rigidity for hermitian symmetric spaces, following Siu.

Let V, W be compact hermitian locally symmetric spaces. Let $f_0 : V \to W$ be a homotopy equivalence. Deform f_o to a harmonic map f. Since the degree of f is non zero, the differential df is surjective at many points.

Lemma (Carlson-Toledo [Car-Tol]) . *The only abelian subspaces of* $\underline{p}^W \otimes \mathbb{C}$ *of maximal dimension are* $\underline{p}^{1,0}$ *and* $\underline{p}^{0,1}$. *i.e., the subspaces that define the complex structure on* W.

Admitting this fact, we see that f is holomorphic or antiholomorphic on a big set. By elliptic regularity, f is holomorphic (or antihomorphic) everywhere. f is injective since its fibers are complex subvarieties homologous to a point. Lift f to a biholomorphism of symmetric spaces, which are bounded domains in $\mathbb{C}^n$. The symmetric metric coincides with the Bergman metric, which is natural under biholomorphisms. We conclude that f is an isometry.

The assumption that V is locally symmetric was needed to conclude that f_o was homotopic to an isometry only. For a Kähler domain V, one concludes that f_o is homotopic to a holomorphic or antiholomorphic map.

4.6.F_3. Non linear Hodge theory.

Hodge theory is concerned with the interplay between topological objects (de Rham cohomology) and holomorphic objects (cohomology of sheaves of holomorphic forms). Nonlinear Hodge theory extends this correspondance and gives holomorphic counterparts to space of representations of fundamental groups of compact Kähler manifolds.

Let V be compact Kähler. Start with a finite dimensional representation

$$\rho : \pi_1(V) \to Gl(r, \mathbb{C}).$$

Assume the Zariski closure of $\rho(\pi_1(V))$ in $GL(r, \mathbb{C})$ is reductive. The representation ρ determines a vector bundle

$$F = \tilde{V} \times_\rho \mathbb{C}^r$$

over V. Hermitian metrics on F correspond to twisted maps of V to the space X of hermitian quadratic forms on $\mathbb{C}^r$, i.e., up to an irrelevant scale, the symmetric space $Gl(r, \mathbb{C})/U(r)$.

According to theorem 4.5.D_2, there exists on F a unique "harmonic metric" (i.e., metric corresponding to a harmonic equivariant map f from $\tilde{V}$ to X). The bundle F carries a natural flat connection ∇ (since it lifts to a trivial bundle on $\tilde{V}$). In general, this connection is not unitary with respect to the harmonic metric, thus it splits as

$$\nabla = D + \theta$$

where D is a unitary connection, θ a 1-form with values in (symmetric) endomorphism of F, which is nothing but df.

In this context, paragraph 4.6.F_1 can be improved : D defines a holomorphic structure on F (we recover the general fact that E = End F has a holomorphic structure), and $\theta^{1,0}$ is a End F-valued holomorphic 1-form. Furthermore,

$$(\theta^{1,0})^2 = 0.$$

Such data (E, $\theta^{1,0}$) form a Higgs bundle ([Hit], [Sim]).

When the representation ρ is irreducible, the Higgs bundle is stable.

Conversely, a stable Higgs bundle whose Chern classes vanish arises from a representation of π_1 (V) (ibidem).

4.6.G. Rigid exterior forms and superrigidity

4.6.G_1. Definition : Let ω be an exterior form on the vector space $\mathbb{R}^n$. We say that ω is *rigid* if the subgroup of $Gl(n, \mathbb{R})$ that fixes ω is compact.

Example : The Kähler form on $\mathbb{C}^n$ is not rigid. Indeed, its stabilizer is the non compact symplectic group $Sp(2n, \mathbb{R})$.

Example : On quaternionic vector space $\mathbb{H}^n$, there is a $Sp(n-1)$ $Sp(1)$-invariant 4-form, which is rigid if $n \geq 2$. Let $\omega_i, \omega_j, \omega_k$ denote the three components of the $\mathrm{Im}\mathbb{H}$-valued 2-form

$$\mathrm{Im}\left(\sum_{\ell} \overline{q_\ell} q'_\ell\right)$$

then

$$\omega = \omega_i^2 + \omega_j^2 + \omega_k^2$$

has stabilizer exactly $Sp(n-1)\,Sp(1)$ provided $n \geq 2$.

Similarly, there is on $\mathbb{C}a^2$ a rigid 8-form, whose stabilizer is exactly $Spin(9) \subset SO(16)$.

A Riemannian manifold of dimension $4n$ is called quaternion-Kähler if its holonomy at each point is contained in a copy of $Sp(n-1)\,Sp(1) \subset Gl(4n, \mathbb{R})$. Such a manifold automatically carries a rigid parallel 4-form. Quaternion-Kähler symmetric spaces are listed in ([Bes], chap. 14). These include quaternionic hyperbolic spaces $\mathbb{H}^{4n}$, $n \geq 2$. It is unknown whether there exist non locally symmetric compact quaternionic manifolds with negative Ricci curvature.

4.6.G_2. Theorem : *Let* V *be a compact Riemannian manifold that admits a rigid paralled form. Let* W *have non positive curvature operator. Then every twisted harmonic map of* V *to* W *is totally geodesic.*

Proof : Theorem 4.6.D_1 applies to give $d^*(df \wedge \omega) = 0$. Formula 4.6.$D_2$ implies that the Hessian $\nabla df \in A \otimes f^*TW$ where $A \subset \mathrm{End}\ TV$ is the Lie algebra of the stabilizer of ω in TV. If A integrates to a compact group, then every matrix in A has imaginary eigenvalues. But $ddf = 0$ implies ∇df is symmetric and has real eigenvalues. We conclude that $\nabla df = 0$, i.e. f is totally geodesic.

Corollary; (Archimedean superrigidity for quaternionic and Cayley hyperbolic spaces).
Let G *be equal, either to* $Sp(n, 1)$, $n \geq 2$ or F_4^{-20}. *Let* $\Gamma \subset G$ *be a discrete cocompact subgroup. Let* G' *be a semisimple Lie group with maximal compact subgroup* K'.
Let ρ *be a homomorphism of* Γ *into* Γ'. *Assume that the Zariski closure of* $\rho(\Gamma)$ *in* G' *is reductive. Then* ρ *factors through a homomorphism*

$$\Gamma \to G \times K' \to G'$$

Remark : (a) The corollary extends to finite covolume lattice, see ($[\mathrm{Cor}]_1$).

(b) Many symmetric spaces of higher rank admit rigid parallel forms, but not all of them.

References

[Ale] A.D. Aleksandrov, Convex Polytopes, Moscow 1950 (Russian).

[And] E. Andreev, Convex polyhedra of finite volume in Lobachevski space (Russian), Mat. Sb. 83 (125) (1970) p.p. 256-260.

[Bas] H. Bass, The congruence subgroup problem, Proc. Conf. Local Fields, Driebergen 1966, Springer 1967, p.p. 16-22.

[B-G-S] W. Ballmann, M. Gromov, V. Schroeder, Manifolds of non-positive curvature. Progress in Math. Vol. 61, Birkhäuser, Boston-Basel-Stuttgart (1985).

[Bes] A.L. Besse, Einstein manifolds, Ergeb. Bd. 10, Springer Verlag, 1986.

[Bor]$_1$ A. Borel, Introduction aux groupes arithmétiques, Hermann, Paris 1969.

[Bor]$_2$ A. Borel, Compact Clifford-Klein forms of symmetric spaces, Topology 2, (1963), p.p. 111-122.

[Bor-Sha] Z.I. Borevič and I.R. Shafarevič, Number theory, Acad. Press 1966.

[Bus] H. Buseman, Extremals on closed hyperbolic space forms, Tensor, 16 (1965), p.p. 313-318.

[B-Y] S. Bochner and K. Jano, Curvature and Betti numbers, Princeton University Press, Princeton 1953.

[Cal] E. Calabi, On compact Riemannian manifolds with constant curvature I. Proc. Symp. Pure Math. Vol III, p.p. 155-180, A.M.S. Providence, R.I. 1961.

[Cal-Ves] E. Calabi and E. Vesentini, On compact locally symmetric Kähler manifolds, Ann. of Math. (2) 71 (1960), p.p. 472-507.

[Car-Tol] J. Carlson, D. Toledo, Harmonic maps of Kähler manifolds to locally symmetric spaces, Publ. Math. I.H.E.S. 69 (1989), p.p. 173-201.

[Cas] J.W.S. Cassels, An introduction to the geometry of numbers, Springer-Verlag 1959.

[Ch-Eb] F. Cheeger, D. Ebin, Comparison theorems in Riemannian geometry, North-Holland 1975.

[Che] S.S. Chern, On a generalization of Kähler geometry, in "Algebraic Geometry and Topology Symp. in honor of S. Lefschetz", Princeton Univ. Press, (1957), p.p. 103-121.

[Cor]$_1$ K. Corlette, Archimedean superrigidity and hyperbolic geometry, preprint 1990.

[Cor]$_2$ K. Corlette, Flat G-bundles and canonical metrics, J. Diff. Geom. 28 (1988), p.p. 361-382.

[dlH-V] P. de la Harpe et A. Valette. La propriété (T) de Kazhdan pour les groupes localement compacts, Astérisque 175 (1989). Soc. Math. de France.

[Don] S. Donaldson, Twisted harmonic maps and the self duality equations, Proc. London Math. Soc. 55 (1987), p.p. 127-131.

$[E\text{-}L]_1$ J. Eells and L. Lemaire, A report on harmonic maps, Bull. Lond. Math. Soc. 10 (1978), p.p. 1-68.

$[E\text{-}L]_2$ J. Eells and L.Lemaire, Another report on harmonic maps, Bull. Lond. Math. Soc. 20 (1988), p.p. 385-525.

[E-S] J. Eells and J.H. Sampson, Harmonic mappings of Riemannian manifolds, Amer. J. Math. 86 (1964), pp. 109-160.

[Efz] V. Efremovič, The proximity geometry of Riemannian manifolds, Uspehi Mat., Nauk 8 (1953), p. 189 (Russian).

[Ef-Ti] V. Efremovič and E. Tichomirova, Equimorphisms of hyperbolic spaces, Isv. Ac. Nauk. 28 (1964), p.p. 1139-1144.

[Fl] W. J. Floyd, Group completion and limit sets of Kleinian groups, Inv. Math. 57 (1980), p.p. 205-218.

$[\text{Für}]_1$ H. Fürstenberg, Poisson boundaries and envelopes of discrete groups, Bull. Am. Math. Soc. 73 (1967), p.p. 350-356.

$[\text{Für}]_2$ H. Fürstenberg, Rigidity and cocycles for ergodic actions of semisimple Lie groups. Lect. Notes Math. 842, p.p. 273-292, Springer-Verlag 1981.

$[\text{Für}]_3$ H. Fürstenberg, Boundary theory and stochastic processes on homogeneous spaces. Proc. Symp. Pure Math. XXVI, 1973, p.p. 193-233.

[Geh] F.W. Gehring, Rings and quasiconformal mappings in space, Trans. Amer. Math. Soc. 103 (1962), p.p. 353-393.

$[\text{Gro}]_1$ M. Gromov, Hyperbolic manifolds according to Thurston and Jorgensen, in Springer Lecture Notes, 842 (1981), p.p. 40-53.

$[\text{Gro}]_2$ M. Gromov, Groups of polynomial growth and expanding maps, Publ. Math. IHES, 53 (1981), p.p. 53-78.

$[\text{Gro}]_3$ M. Gromov, Hyperbolic manifolds groups and actions, in Riem. surfaces and related topics, Ann. Math. Studies 97 (1981), p.p. 183-215.

$[\text{Gro}]_4$ M. Gromov, Hyperbolic groups, in Essays in group theory, S.M. Gersten ed., p.p. 75-265, Springer-Verlag 1987.

[G-L-P] M. Gromov, J. Lafontaine, and P. Pansu, Structures métriques pour les variétés riemanniennes, CEDIC- Fernand Nathan, Paris,1981.

[G-P] M. Gromov and I. Piatetski-Shapiro, Non-arithmetic groups in Lobachevsky spaces, Publ. Math. I.H.E.S. 66 (1988), p.p. 93-103.

[G-S] M. Gromov, R. Schoen, in preparation.

[Ha-Mu] U. Haagerup and M. Munkholm, Simplices of maximal volume in hyperbolic n-space. Act. Math. 1941 (1981), p.p. 1-11.

[Hit] N. Hitchin, The selfduality equations on a Riemann surface, Proc. London Math. Soc. 55 (1987), p.p. 59-126.

[Kar] H. Karcher, Riemannian center of mass and mollifier smoothing, Comm. Pure and Appl. Math. 30 (1977), p.p. 509-541.

[Klin] W. Klingenberg, Geodätischer Fluss auf Mannigfaltigkeiten vom hyperbolischen Typ, Inv. Math. 14 (1971), p.p. 63-82.

[Law-Mic] B. Lawson and M.-L. Michelsohn, Spin geometry, Princeton University Press, Princeton 1989.

[Kaz] D. Kazhdan, Connection of the dual space of a group with the structure of its closed subgroups, Func. Anal. and Appl. 1 (1967), p.p. 63-65.

[Mal] A.I. Malcev, On a class of homogeneous spaces, Amer. Math. Soc. Transl. 39, (1951), p.p. 276-307.

$[Mar]_1$ G. Margulis, Discrete groups of motions of manifolds of non-positive curvature (Russian), ICM 1974. Translated in A.M.S. Transl. (2) 109 (1977), p.p. 33-45.

$[Mar]_2$ G. Margulis, Discrete subgroups of semisimple Lie groups, Springer-Verlag 1990.

$[Mar]_3$ G. Margulis, The isometry of closed manifolds of constant negative curvature with the same fundamental group, Dokl. Akad. Nauk. SSSR 192 (1970), p.p. 736-737.

[Mau] F.I. Mautner, Geodesic flows on symmetric Riemannian spaces, Ann. of Math. 65 (1957), p.p. 416-431.

$[Mil]_1$ J. Milnor, Whitehead torsion, Bull. Amer. Math. Soc. 72 (1966), p.p. 358-426.

$[Mil]_2$ J. Milnor, A note on curvature and fundamental group, J. Diff. Geom. 2 (1968), p.p. 1-7.

[Mok] N. Mok, Metric rigidity theorems on Hermitian locally symmetric manifolds, World Sci. Singapore-New Jersey-London-Hong Kong, 1989.

[Mors] M. Morse, Geodesics on negatively curved surfaces, Trans. Amer. Math. Soc 22 (1921) p.p. 84-100.

$[Mos]_1$ G. Mostow, Quasi-conformal mappings in n-space and the rigidity of hyperbolic space form, Publ. Math. I.H.E.S., vol. 34, 1968, p.p. 53-104.

$[Mos]_2$ G. Mostow, Equivariant embeddings in Euclidean space, Ann. of Math. 65 (1957), 432-446.

[Mos]$_3$ G. Mostow, Factor spaces of solvable groups, Ann. of Math. 60 (1954), p.p. 1-27.

[Mos]$_4$ G.D. Mostow, Strong rigidity of symmetric spaces, Ann. Math. Studies 78, Princeton 1973.

[Mos]$_5$ G.D. Mostow, Some new decomposition theorems for semisimple groups, Memoirs Amer. Math. Soc. 14 (1955), p.p. 31-54.

[Pal] R. Palais, Equivalence of nearby differentiable actions of a compact group, Bull. Am. Math. Soc. 61(1961), p.p. 362-364.

[Pan]$_1$ P. Pansu, Croissance des boules et des géodésiques fermées dans les nilvariétés, Ergod. Th. Dynam. Syst. 3 (1983), p.p. 415-445.

[Pan]$_2$ P. Pansu, Métriques de Carnot-Caratheodory et quasi-isométries des espaces symétriques de rang un, Ann. of Math. 129:1 (1989), p.p. 1-61.

[Pont] L.S. Pontryagin, Topological groups, Gordon and Breach 1966.

[Pra] G. Prasad, Strong rigidity of Q-rank 1 lattices, Inv. Math. 21 (1973), p.p. 255-286.

[Rag] M.S. Raghunathan, Discrete subgroups of Lie groups, Springer-Verlag, 1972.

[Resh] Yu. G. Reshetniak, On conformal mappings of a space, Dokl. Acad. Nauk. SSSR 130 (1960) p.p. 981-983 = Soviet. Math. Dokl. 1 (1960), p.p. 122-124.

[Rei] H.M. Reimann, Capacities in the Heisenberg group, Preprint University of Bern (1987).

[Sel] A. Selberg, On discontinuous groups in higher dimensional symmetric spaces, Contribution to Function Theory, Tata Inst., Bombay 1960, p.p. 147-164.

[Sha] I. R. Shafarevič, Basic Algebraic geometry, Springer-Verlag 1974.

[Sam] J.H.Sampson, Applications of harmonic maps to Kähler geometry, Contemp. Math. 49 (1986), p.p. 125-133.

[Sim] C. Simpson, Constructing variations of Hodge structure using Yang-Mills theory and applications to uniformization, J. of the Amer. Math. Soc. 1 (1988), p.p. 867-918.

[Siu] Y.-T. Siu, The complex analyticity of harmonic maps and the strong rigidity of compact Kähler manifolds, Ann. of Math. 112 (1980), p.p. 73-111.

[Str] R. Strichartz, Sub-Riemannian geometry, Journ. of Diff. Geom. 24, (1986), p.p. 221-263.

[Šv] A. Švarc, A volume invariant of coverings, Dokl. Akad. Nauk SSSR 105 (1955), p.p. 32-34.

[Th] W. Thurston, Geometry and topology of 3-manifolds, Lecture notes Princeton, 1978.

[Tuk] P. Tukia, Quasiconformal extension of quasisymmetric maps compatible with a Möbius group, Acta Math. 154 (1985), p.p. 153-193.

[Tuk-Väi] P. Tukia, J. Väisälä, A remark on 1-quasiconformal maps, Ann. Acad. Sci. Fenn. 10 (1985), p.p. 561-562.

[Ul] S.M. Ulam, A collection of mathematical problems. Interscience Publ. N.Y. 1960.

[Väi] J. Väisälä, Lectures on n-dimensional quasiconformal mappings, Lecture Notes in Math. Vol 129, Springer, Berlin (1971).

[Wang] H.-C. Wang, Topics on totally discontinuous groups, in "Symmetric spaces" edited by W.M. Boothby and G.L. Weiss, p.p. 459-487, Marcel Dekker, Inc. N.Y. 1972.

[Wei] H. Weil, On discrete subgroups of Lie groups II, Ann. of Math. vol. 75 (1962), p.p. 578-602.