CUBO A Mathematical Journal
Vol. 7, Nº 2, (111 - 138). August 2005.

Symmetric Spaces of Noncompact type

Martin Moskowitz [1]
Department of Mathematics, CUNY Graduate Center
365 Fifth Ave. New York NY 10016. USA
mmoskowi@mat.uniroma2.it

ABSTRACT

This article gives a detailed introduction to symmetric spaces of non-compact type and their relation to corresponding semisimple lie groups. This is done more or less from scratch and explicitely without the reader having to know large parts of modern differential geometry.

RESUMEN

Este artículo entrega una detallada introducción a espacios simétricos de tipo no compacto y su relación con los correspondientes grupos de Lie semisimples. Esto es hecho mas o menos en términos generales y explícitamente sin que el lector tenga gran conocimiento de la geometría diferencial moderna.

Key words and phrases: *Semisimple Lie group and Lie algebra of non-compact type, Maximal compact subgroup and maximal subalgebra of compact type and their conjugacy. Cartan and polar decomposition, rank, irreducible symmetric space,isometry group, Killing form, exponential map, Cartan involution, law of cosines and two point homogeneous space.*

Math. Subj. Class.: *22E, 57S, 57T*

[1]This article was written while the author was supported by a PSC-CUNY Grant from the Research Foundation of CUNY. He would like to take this opportunity to thank the RF for its generous support.

1 Introduction

In this article we shall give an introduction to symmetric spaces of noncompact type. This subject, largely the creation of Élie Cartan (1869-1951), is of fundamental importance both to geometry and Lie theory. Indeed, one of the great achievements of the mathematics of the first half of the twentieth century was E. Cartan's discovery of the fact that these two categories correspond exactly. Namely, given a connected, centerless, real semisimple Lie group G without compact factors there is associated to it a unique symmetric space of noncompact type. This is G/K, where K is a maximal compact subgroup of G and G/K takes the Riemannian metric induced from the Killing form of G. Conversely, if one starts with an arbitrary symmetric space, X, all of whose irreducible constituents are neither compact nor $\mathbb{R}^n$, then $X = G/K$, where G is the identity component of the isometry group of X. Here G is a centerless, real semisimple Lie group without compact factors. Thus, we have a bijective correspondence between the two categories and this fact underlies an important reason why differential geometry and Lie theory are so closely bound. As one might expect, this close relationship between the two will show up in some of the proofs. For the details of all this, see S. Helgason [4] and G.D. Mostow [9]. Also, [4] has a particularly convenient and useful early chapter on differential geometry. Concerning this correspondence, the same may be said of Euclidean space and its group of isometries, or of compact semisimple groups and symmetric spaces of compact type, which were also studied by E. Cartan. However, we shall not deal with these here. Taken as a whole Cartan's work on symmetric spaces can be considered as the completion of the well known "Erlangen Program" first formulated by F. Klein in 1872. In particular, it ties together Euclidean, elliptic and hyperbolic geometry in any dimension.

Before turning to our subject proper it might be helpful to consider a most important example, namely that of $G = \mathrm{SL}(2,\mathbb{R})$ and X the hyperbolic plane, which we view here as the Poincaré upper half plane, H^+, consisting of all complex numbers $z = x + iy$, where $y > 0$. We let G act on H^+ by fractional linear transformations, $g \cdot z = \frac{az+b}{cz+d}$,

$$g = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

where a, b, c and d are real and $\det g = 1$. Since $\Im(\frac{az+b}{cz+d}) = \frac{\Im(z)}{|cz+d|^2} > 0$, we see that $g \cdot z \in H^+$. That this is an action is easy to verify. Now this action is transitive. Let $c = 0$, then $a \neq 0$ and $d = \frac{1}{a}$. Then $g \cdot i = a^2 i + ab$. Evidently, by varying $a > 0$ and $b \in \mathbb{R}$ this gives all of H^+. A moment's reflection tells us that the isotropy group, $\mathrm{Stab}_G(i)$, is given by $a = d$ and $c = -b$. Since $\det g = a^2 + b^2 = 1$, we see

$$\mathrm{Stab}_G(i) = \{g : g = \begin{pmatrix} \cos t & \sin t \\ -\sin t & \cos t \end{pmatrix} : t \in \mathbb{R}\}.$$

On H^+ we place the Riemannian metric $ds^2 = \frac{dx^2+dy^2}{y^2}$ (meaning the hyperbolic metric $ds = \frac{ds_{Euc}}{\Im z}$) and check that G acts by isometries on H^+ (for this see, for example, p. 118 of [8]). Since G is connected, its image, $\mathrm{PSL}(2,\mathbb{R})$, is contained in $\mathrm{Isom}_0(H^+)$. (Actually it is $\mathrm{Isom}_0(H^+)$, but that will not matter. Also, there are 2 connected components, the other is the anti-holomorphic automorphisms, but that will not matter either. From the point of view of the symmetric space it does not even matter whether we take $\mathrm{SL}(2,\mathbb{R})$ or $\mathrm{PSL}(2,\mathbb{R})$. However, we note that $\mathrm{PSL}(2,\mathbb{R})$, the group that is really acting, is the centerless version.)

Another model for this symmetric space is the unit disk, $D \subseteq \mathbb{C}$, called the disk model. It takes the metric $ds^2 = 4\frac{dx^2+dy^2}{(1-r^2)^2}$ and has the advantage of radial symmetry about the origin, 0. Here r is the usual radial distance from 0. The point of the 4 is, as we shall see, to make D isometric with H^+, or put another way, to normalize the curvature on D to be -1. Now the Cayley transform $c(z) = \frac{z-i}{z+i}$ maps H^+ diffeomorphically onto D. Its derivative $c'(z) = \frac{zi}{(z+i)^2}$. A direct calculation shows that for $z \in H^+$

$$\frac{2|c'(z)|}{1-|c(z)|^2} = \frac{1}{\Im(z)}.$$

Using this we see that if $w = c(z)$, then $|dw| = |c'(z)||dz|$ and so

$$\frac{2|dw|}{1-|w|^2} = \frac{2|c'(z)|}{1-|c(z)|^2}|dz| = \frac{|dz|}{|\Im(z)|}.$$

Thus c is an isometry. Of course in the form of the disk, the group of isometries and its connected component will superficially look different.

2 The Polar Decomposition

Explaining the reasons for the relationship mentioned above will take some time and we shall begin by studying the exponential map on certain specific manifolds.

The $n \times n$ complex matrices will be denoted by $\mathfrak{gl}(n,\mathbb{C})$ and the real ones by $\mathfrak{gl}(n,\mathbb{R})$). Denote by $\mathcal{H}$ the set of all Hermitian matrices in $\mathfrak{gl}(n,\mathbb{C})$ and by H the positive definite ones. It is easy to see that $\mathcal{H}$ is a real (but not a complex!) vector space of $\dim n^2$. Similarly, we denote by $\mathcal{P}$ the symmetric matrices in $\mathfrak{gl}(n,\mathbb{R})$ and by P those that are positive definite. $\mathcal{P}$ is a real vector space of dim $\frac{n(n+1)}{2}$. As we shall see, H and P and certain of their subspaces will actually comprise all symmetric spaces of noncompact type.

Proposition 2.1 *P and H are open in $\mathcal{P}$ and $\mathcal{H}$, respectively. As open sets in a real vector space each is, in a natural way, a real analytic manifold of the appropriate dimension.*

Proof. Let $p(z) = \sum_i p_i z^i$ and $q(z) = \sum_i q_i z^i$ be polynomials of degree n with complex coefficients, let $z_1 \ldots z_n$ and $w_1 \ldots w_n$ denote their respective roots counted according to multiplicity and let $\epsilon > 0$. It follows from Rouché's theorem (see [8]) that there exists a sufficiently small $\delta > 0$ so that if for all $i = 0, \ldots, n$ if $|p_i - q_i| < \delta$, then after a possible reordering of the w_i's, $|z_i - w_i| < \epsilon$ for all i. Suppose H were not open in $\mathcal{H}$. Then there would be an $h \in H$ and a sequence $x_j \in \mathcal{H} - H$ converging to h in $\mathfrak{gl}(n, \mathbb{C})$. Since h is positive definite, all its eigenvalues are positive. Choose ϵ so small that the union of the ϵ balls about the eigenvalues of h lies in the right half plane. Since the coefficients of the characteristic polynomial of an operator are polynomials and therefore continuous functions of the matrix coefficients and x_j converges to h, for j sufficiently large, the coefficients of the characteristic polynomial of x_j are within δ of the corresponding coefficient of the characteristic polynomial of h. Hence all the eigenvalues of such an x_j are positive. This contradicts the fact that none of the x_j are in H, proving H is open in $\mathcal{H}$. Intersecting everything in sight with $\mathfrak{gl}(n, \mathbb{R})$ shows P is also open in $\mathcal{P}$. ∎

Proposition 2.2 *Upon restriction, the exponential map of $\mathfrak{gl}(n, \mathbb{C}) = \mathfrak{gl}(V)$ is a real analytic diffeomorphism between $\mathcal{H}$ and H. Its inverse, is given by*

$$\operatorname{Log} h = \log(\operatorname{tr} h)I - \sum_{i=1}^{\infty}(I - \frac{h}{\operatorname{tr} h})^i/i,$$

which is an analytic function on H.

As a consequence we see that the restriction of exp to any real subspace of $\mathcal{H}$ gives a real analytic diffeomorphism of the subspace with its image. In particular, exp is a real analytic diffeomorphism between between $\mathcal{P}$ and P. In particular, in all these cases exp is a bijection.

Proof. We shall do this for $\mathcal{H}$, the real case being completely analogous. Suppose $h \in H$ is diagonal with eigenvalues $h_i > 0$. Then $\operatorname{tr}(h) > 0$ and $0 < \frac{h_i}{\operatorname{tr}(h)}$ so $\log(\operatorname{tr}(h))$ is well defined and $\log(\frac{h_i}{\operatorname{tr}(h)})$ is defined for all i. But since $0 < \frac{h_i}{\operatorname{tr}(h)} < 1$, we see that $0 < (1 - \frac{h_i}{\operatorname{tr}(h)})^k < 1$ for all positive integers k. Hence $\operatorname{Log}(\frac{I-h}{\operatorname{tr}(h)})$ is given by an absolutely convergent power series $-\sum_{i=1}^{\infty}(I - \frac{h}{\operatorname{tr} h})^i/i$. If u is a unitary operator so that uhu^{-1} is diagonal, then $\operatorname{tr}(uhu^{-1}) = \operatorname{tr}(h)$ and since conjugation by u commutes with any convergent power series, this series actually converges for all $h \in H$ and is a real analytic function Log on H. Because on the diagonal part of H this function inverts Exp, and both Exp and this power series commutes with conjugation, it inverts Exp everywhere on H. Finally, $\log(tr(h))I$ and $\operatorname{Log}(\frac{h}{\operatorname{tr} h})$ commute and Exp of a sum of commuting matrices is the product of the Exp's. Since Log inverts Exp on the diagonal part of H it follows that

$$\operatorname{Log}(h) = \log(\operatorname{tr}(h))I + \operatorname{Log}(\frac{h}{\operatorname{tr} h}) = \log(\operatorname{tr}(h))I - \sum_{i=1}^{\infty}(I - \frac{h}{\operatorname{tr} h})^i/i.$$

∎

We shall need the following elementary fact whose proof is left to the reader.

Lemma 2.3 *For any $g \in \mathrm{GL}(n, \mathbb{C})$, $g^*g \in H$.*

It follows that for all $g \in \mathrm{GL}(n,\mathbb{C})$, $\log(g^*g) \in \mathcal{H}$ and since this is a real linear space also $\frac{1}{2}\log(g^*g) \in \mathcal{H}$. This means we can apply exp and conclude the following:

Corollary 2.4 *$h(g) = \exp(\frac{1}{2}\log(g^*g)) \in H$ is a real analytic function from $\mathrm{GL}(n,\mathbb{C}) \to H$.*

Hence $h(g)^n = \exp(\frac{n}{2}\log(g^*g)) \in H$ for every $n \in \mathbb{Z}$. In particular, $h(g)^{-2} = \exp(\frac{2}{2}\log(g^*g)) = g^*g$. So that

$$gh(g)^{-1}(gh(g)^{-1})^* = gh(g)^{-1}h(g)^{-1*}g^* = gh(g)^{-2}g^*$$

and, since $h(g)^{-1} \in H$, $g(g^*g)^{-1}g^* = I$. Thus, $gh(g)^{-1} = u(g)$ is unitary for each $g \in \mathrm{GL}(n,\mathbb{C})$. Since group multiplication and inversion are analytic, $u(g)$ is also a real analytic function on $\mathrm{GL}(n,\mathbb{C})$ (as is $h(g)$). Now this decomposition $g = uh$, where $u \in \mathrm{U}$ and $h \in H$ is actually unique. To see this, suppose $u_1h_1 = g = u_2h_2$. Then $u_2^{-1}u_1 = h_2h_1^{-1}$ so that $h_2h_1^{-1}$ is unitary. This means $(h_2h_1^{-1})^* = (h_2h_1^{-1})^{-1}$ and hence $h_1^2 = h_2^2$. But since h_1 and $h_2 \in H$, each is an exponential of something in $\mathcal{H}$; $h_i = \exp x_i$. But then $h_i^2 = \exp 2x_i$ and since exp is $1:1$ on $\mathcal{H}$, we get $2x_1 = 2x_2$ so $x_1 = x_2$ and therefore $h_1 = h_2$ and $u_1 = u_2$. The upshot of all this is that we have a real analytic map $\mathrm{GL}(n,\mathbb{C}) \to \mathrm{U}(n,\mathbb{C}) \times H$ given by $g \mapsto u(g)h(g)$. Since $g = u(g)h(g)$ for every g (multiplication in the Lie group $\mathrm{GL}(n,\mathbb{C})$), this map is surjective and has a real analytic inverse. We summarize these facts as the following Polar Decomposition Theorem.

Theorem 2.5 *The map $g \mapsto u(g)h(g)$ gives a real analytic diffeomorphism $\mathrm{GL}(n,\mathbb{C}) \to \mathrm{U}(n,\mathbb{C}) \times H$. Identical reasoning also shows that as a real analytic manifold $\mathrm{GL}(n,\mathbb{R})$ is, in the same way, diffeomorphic to $\mathrm{O}(n,\mathbb{R}) \times P$.*

From this it follows that, since H and P are each diffeomorphic with a Euclidean space, and therefore are topologically trivial, in each case the topology of the noncompact group is completely determined by that of the compact one. In this situation, one calls the compact group a *deformation retract* of the noncompact group. Since P and H are diffeomorphic images under exp of some Euclidean space, one calls them *exponential submanifolds*. For example, connectedness, the number of components, simple connectedness and the fundamental group of the noncompact group are each the same as that of the compact one. Thus for all $n \geq 1$, $\mathrm{GL}(n,\mathbb{C})$ is connected and its fundamental group is $\mathbb{Z}$, while for all n, $\mathrm{GL}(n,\mathbb{R})$ has 2 components and the fundamental group of its identity component is $\mathbb{Z}_2$ for $n \geq 3$ and $\mathbb{Z}$ for $n = 2$. These facts follow from the long exact homotopy sequence for a fibration and are explained in C. Chevalley [2].

3 The Cartan Decomposition of a Real Semi-simple Lie Group of Noncompact type

What we have done so far may seem rather special. We now turn to more general groups G and also streamline our notation. Instead of $\mathcal{H}$, we shall consider certain real subspaces of $\mathcal{H}$ denoted by $\mathfrak{p}$ whose exponential image will be P and make the following definition.

Definition 3.1 *Let G be a Lie subgroup of $\mathrm{GL}(n,\mathbb{R})$ with Lie algebra $\mathfrak{g}$. We denote by $K = \mathrm{O}(n,\mathbb{R}) \cap G$, by P the positive definite symmetric matrices in G, by $\mathfrak{p}$ the symmetric matrices of $\mathfrak{g}$, and by $\mathfrak{k}$ the skew symmetric matrices in $\mathfrak{g}$. In the case that G be a Lie subgroup of $\mathrm{GL}(n,\mathbb{C})$ we again denote the Lie algebra by $\mathfrak{g}$, but now $K = \mathrm{U}(n,\mathbb{C}) \cap G$, P is the positive definite Hermitian matrices of G, $\mathfrak{p}$ is Hermitian matrices of $\mathfrak{g}$ and $\mathfrak{k}$ the the skew Hermitian matrices in $\mathfrak{g}$.*

Lemma 3.2 *Let $q(t) = \sum_{j=1}^{n} c_j \exp(b_j t)$ be a trigonometric polynomial, where $c_j \in \mathbb{C}$, and b_j and $t \in \mathbb{R}$. If q vanishes for an unbounded set of real t's, then $q \equiv 0$.*

An immediate consequence is that for a polynomial $p \in \mathbb{C}[z_1, \ldots, z_n]$ in n complex variables with complex coefficients and $(x_1, \ldots, x_n) \in \mathbb{R}^n$, if $p(\exp(tx_1), \ldots, \exp(tx_n))$ vanishes for an unbounded set of real t's, then it vanishes identically in t.

Proof. First we can assume that the t's for which q vanishes tend to $+\infty$. Otherwise, they would have to tend to $-\infty$ and in this case we just let $p(t) = q(-t)$. Then p is also a trigonometric polynomial and if $p = 0$, then so is q. Reorder the b_j's, if necessary, so that they are strictly increasing by combining terms by adding the corresponding c_j's. Of course, we can now assume that all the c_j's are nonzero. Let t_k be a sequence tending to $+\infty$ on which q vanishes. Suppose there are two or more b_j's. Since

$$\frac{q(t)}{c_n \exp(b_n t)} = \sum_{j=1}^{n-1} \frac{c_j}{c_n} \exp((b_j - b_n)t) + 1,$$

it follows that $\frac{q(t)}{c_n \exp(b_n t)} \to 1$ as $k \to \infty$. But since q is identically 0 in k so is this quotient, a contradiction. This means that all the b_j's are equal and so $q(t) = c \exp(bt)$ for some $c \in \mathbb{C}$ and $b \in \mathbb{R}$. This function cannot have an infinite number of zeros unless $c = 0$, that is $q = 0$.

∎

Definition 3.3 *One calls a subgroup $G \subseteq \mathrm{GL}(n,\mathbb{C})$ an* algebraic group *if it is the simultaneous zero set within $\mathfrak{gl}(n,\mathbb{C})$ of a family of polynomials with complex coefficients in the $x_{i,j}$ coordinates of the matrices in $\mathfrak{gl}(n,\mathbb{C})$. Clearly, such a group is a closed subgroup of $\mathrm{GL}(n,\mathbb{C})$ in the usual Euclidean topology and hence by a theorem of E. Cartan is a Lie group. Further we shall call $G_{\mathbb{R}} = G \cap \mathrm{GL}(n,\mathbb{R})$ its $\mathbb{R}$-points. Similarly, the $\mathbb{R}$-points $G_{\mathbb{R}}$ of an algebraic group G is also a Lie group. If the family of polynomials defining G happens to have all its coefficients lying in some subfield F of $\mathbb{C}$, we then say G is defined over F.*

Typical examples of algebraic groups are $\mathrm{GL}(n,\mathbb{C})$ itself (the empty set of polynomials) and $\mathrm{SL}(n,\mathbb{C})$ itself (the single polynomial $\det -1 = 0$). The respective real points are $\mathrm{GL}(n,\mathbb{R})$ and $\mathrm{SL}(n,\mathbb{R})$. The reader should check that each of the complex classical groups (see [4], or [2]) is an algebraic group.

Proposition 3.4 *Suppose M is an algebraic subgroup of* $\mathrm{GL}(n,\mathbb{C})$ *and G be a Lie subgroup of* $\mathrm{GL}(n,\mathbb{R})$ *(alternatively* $\mathrm{GL}(n,\mathbb{C})$*) with Lie algebra $\mathfrak{g}$. Let G have finite index in $M_{\mathbb{R}}$ (alternatively M). If $x \in \mathfrak{p}$ and $\exp x \in G$, then $\exp tx \in P$ for all real t. In particular, $x \in \mathfrak{g}$ and hence $x \in \mathfrak{p}$.*

Proof. To avoid circumlocutions we shall prove the complex case, the real case being completely analogous. Choose $u \in \mathrm{U}(n,\mathbb{C})$ so that uxu^{-1} is diagonal with real eigenvalues λ_j. Replace G by uGu^{-1}, a Lie subgroup of $\mathrm{GL}(n,\mathbb{C})$ which is contained in uMu^{-1} with finite index. Now uMu^{-1} is an algebraic subgroup of $\mathrm{GL}(n,\mathbb{C})$ (and in the real case $uM_{\mathbb{R}}u^{-1} = (uMu^{-1})_{\mathbb{R}}$). Hence we can assume x is diagonal. Let $p(z_{ij})$ be one of the complex polynomials defining M. Since $\exp x \in G$ and G is a group, $\exp kx \in G \subseteq M$ for all $k \in \mathbb{Z}$. But $\exp kx$ is diagonal with diagonal entries $\exp(k\lambda_j)$. Applying p to $\exp kx$, we get $p(\exp kx) = 0$ for all k. By the corollary, $p(\exp tx) = 0$ for all t. Because p was an arbitrary polynomial defining M, it follows that $\exp tx \in M$ for all real t. Since G has finite index in M and the 1-parameter group $\exp tx$ is connected, it must lie entirely in G and therefore in P. Hence $x \in \mathfrak{g}$. ∎

Definition 3.5 *A subgroup G of* $\mathrm{GL}(n,\mathbb{R})$ *(*$\mathrm{GL}(n,\mathbb{C})$*) is called self-adjoint if it is stable under taking transpose ($*$). Here transpose and $*$ refer to any linear involution (conjugate linear involution) on $\mathbb{R}^n$ ($\mathbb{C}^n$).*

For example, $\mathrm{SL}(n,\mathbb{R})$ ($\mathrm{SL}(n,\mathbb{C})$) are self-adjoint since $\det g^t = \det g$ ($\det g^* = \overline{\det(g)}$. The routine calculations showing $\mathrm{O}(n,C)$, $\mathrm{SO}(n,\mathbb{C})$, $\mathrm{O}(p,q)$ and $\mathrm{SO}(p,q)$ are also self-adjoint are left to the reader. In fact, the reader can check that any classical noncompact simple group in E. Cartan's list (see [4]) is self-adjoint. Clearly by their very definition these groups are either algebraic or have finite index in the real points of an algebraic group (essentially algebraic). Now it is an important insight of Mostow [10] that any linear real semisimple Lie group is self-adjoint under an appropriate involution. Moreover, by the root space decomposition the adjoint group of any semisimple group without compact factors is algebraic (actually over $\mathbb{Q}$). Thus here we are really talking about all the semisimple groups without compact factors and, of course, this means our construction actually gives all symmetric spaces of noncompact type. But even if we did not know this, since any classical noncompact simple group is easily seen to be self-adjoint as well as essentially algebraic, we already get a plethora of symmetric spaces from them.

Particular cases of Theorem 3.6 below are the following. We shall leave their routine verification to the reader. $\mathrm{SL}(n,\mathbb{R})$ is real analytically diffeomorphic with $\mathrm{SO}(n) \times P_1$, where the latter is the positive definite symmetric matrices of $\det 1$, which in turn is diffeomorphic under exp with the linear space of real symmetric matrices of trace 0. Similarly, $\mathrm{SL}(n,\mathbb{C})$ is real analytically diffeomorphic with $\mathrm{SU}(n) \times H_1$, where the latter is the positive definite Hermitian matrices of $\det 1$, which in turn is diffeomorphic with the linear space of Hermitian matrices trace 0. As deformation retracts, similar conclusions can be drawn about the topology of these, as well as the other groups mentioned earlier.

The following result is a special case of the Iwasawa decomposition theorem which holds for an arbitrary Lie group with a finite number of components, but with a somewhat more elaborate formulation (see G.P. Hochschild [5]). Here, we content ourselves with the matter at hand. Namely, self-adjoint algebraic groups, or their real points. In this context, it is called the *Cartan decomposition.* By a *maximal compact subgroup* of G we mean one not properly contained in a larger compact subgroup of G. Our next result is the Cartan decomposition.

Theorem 3.6 *Let G be a self-adjoint subgroup of* $\mathrm{GL}(n,\mathbb{C})$ *(*$\mathrm{GL}(n,\mathbb{R})$*) with Lie algebra $\mathfrak{g}$. Suppose that G has finite index in an algebraic subgroup M of* $\mathrm{GL}(n,\mathbb{C})$ *(G has finite index in $M_{\mathbb{R}}$, its real points). Then*

1. *$G = K \times P$ as real analytic manifolds.*
2. *$\mathfrak{g} = \mathfrak{k} \oplus \mathcal{P}$ as a direct sum of $\mathbb{R}$-vector spaces.*
3. $\exp : \mathcal{P} \to P$ *is a real analytic manifold diffeomorphism whose inverse is given by the global power series of Proposition 2.2.*
4. *K is a maximal subgroup of G. In particular, P is simply connected and G is a deformation retract of K.*

Proof. Here again we deal with the complex case, the real case being similar. First we show each $g \in G$ can be written uniquely as $g = k \exp x$, where $k \in K$ and $x \in \mathfrak{p}$. By Theorem 2.5, $g = up$, where $u \in \mathrm{U}(n,\mathbb{C})$ and $p \in \mathcal{H}$. Now $g^* = (up)^* = p^*u^* = pu^{-1}$, so $g^*g = pu^{-1}up = p^2$. Since G is self-adjoint, $p^2 \in G$, and therefore so is p^{2k} for every $k \in \mathbb{Z}$. Now $p = \exp x$ for some Hermitian x. Hence $p^{2k} = \exp 2kx = \exp k2x$. Since $2x$ is Hermitian, $\exp 2x \in G$ and $\exp k2X \in P$ for all k. By Proposition 3.4, $\exp t2x \in P$ for all real t. Now, just as above, taking $t = \frac{1}{2}$ we get $\exp x = p \in P \subseteq G$. But then $gp^{-1} = u \in G$. Therefore $u \in K$. Also, since $\exp tX \in P$ for all real t, $x \in \mathfrak{p}$. Thus $g = kp$ where $k \in K$ and $p = \exp x$ for $x \in \mathfrak{p}$. Thus we have a map $g \mapsto (k,p)$ from G to $K \times P$. As above, if we can show uniqueness of the representation $g = kp$, then the map is onto. But since $K \subseteq \mathrm{U}(n,\mathbb{C})$ and $P \subseteq$ the positive definite Hermitian matrices, this follows from the uniqueness result proven earlier. Since multiplication inverts this map it is 1:1 and has a smooth inverse. The formula, $p(g) = \exp(\frac{1}{2}\log(g^*g)) \in P$ derived in the case of $\mathrm{GL}(n,\mathbb{C})$ is still valid, if suitably interpreted, and gives a real analytic map $G \to P$. Arguing exactly as in the case of $\mathrm{GL}(n,\mathbb{C})$ we see that part 1 is true. Part 3 follows immediately from the case of $\mathrm{GL}(n,\mathbb{C})$ treated earlier.

For part 2, write $x = \frac{x-x^*}{2} + \frac{x+x^*}{2}$. Since the first term is skew Hermitian, the second is Hermitian and each is an $\mathbb{R}$-linear function of $x \in \mathfrak{gl}(n,\mathbb{C})$, this proves part 2 for the case $\mathfrak{gl}(n,\mathbb{C})$. To prove it in general we need only show that $\frac{x-x^*}{2} \in \mathfrak{k}$ and $\frac{x+x^*}{2} \in \mathfrak{p}$. Now for x and $y \in \mathfrak{gl}(n,\mathbb{C})$, $[x^*,y^*] = -[x,y]^*$. Hence the map θ sending $x \mapsto -x^*$ is an involutive automorphism of the Lie algebra $\mathfrak{gl}(n,\mathbb{C})$ called a *Cartan involution.* If we show $\mathfrak{g}$ is stable under this map, then $x \mapsto x^*$ also leaves $\mathfrak{g}$ stable since it is an $\mathbb{R}$ subspace of $\mathfrak{gl}(n,\mathbb{C})$. Hence $\frac{x-x^*}{2} \in \mathfrak{k}$ and $\frac{x+x^*}{2} \in \mathfrak{p}$. Now for $x \in \mathfrak{g}$,

$\exp tx \in G$ for all t. Since G is self-adjoint and $(\exp tx)^* = \exp t(x)^*$, it follows that $x^* \in \mathfrak{g}$. ∎

To prove part 4, we first consider the basic cases, $\mathrm{GL}(n, \mathbb{R})$ and $\mathrm{GL}(n, \mathbb{C})$.

Proposition 3.7 *Let L be a compact subgroup of $\mathrm{GL}(n, \mathbb{C})$ ($\mathrm{GL}(n, \mathbb{R})$). Then some conjugate gLg^{-1}, $g \in \mathrm{GL}(n, \mathbb{C})$ ($\mathrm{GL}(n, \mathbb{R})$) is contained in $\mathrm{U}(n, \mathbb{C})$ ($\mathrm{O}(n, \mathbb{R})$). In particular, $\mathrm{U}(n, \mathbb{C})$ is a maximal compact subgroup of $\mathrm{GL}(n, \mathbb{C})$ and $\mathrm{O}(n, \mathbb{R})$ a maximal compact subgroup of $\mathrm{GL}(n, \mathbb{R})$. In $\mathrm{GL}(n, \mathbb{C})$ and $\mathrm{GL}(n, \mathbb{R})$ any two maximal compact subgroups are conjugate.*

Proof. We deal with the complex case, the other being completely analogous. For an exposition of the existence of Haar measure see, for example, [5]. If $(,)$ is a Hermitian inner product on $\mathbb{C}^n$, using (finite) Haar measure dl on L we can form an L-invariant Hermitian inner product on $\mathbb{C}^n$ given by $\langle v, w\rangle = \int_L (lv, lw)dl$. Thus for some $g \in \mathrm{GL}(n, \mathbb{C})$, gLg^{-1} is contained in $\mathrm{U}(n, \mathbb{C})$.

If $L \supset \mathrm{U}(n, \mathbb{C})$, then it would have to have larger dimension, or if not, $\mathrm{U}(n, \mathbb{C})$ would be an open subgroup. As such, it would be the identity component of L since it is connected. Because L is compact it would consist of a finite number of open components of $\mathrm{U}(n, \mathbb{C})$. But some conjugate, gLg^{-1} is contained in $\mathrm{U}(n, \mathbb{C})$ so L can not have larger dimension. Similarly, by continuity, there can only be one component. This is a contradiction, so $L = \mathrm{U}(n, \mathbb{C})$. In the real case we just work with the compact connected group $\mathrm{SO}(n, \mathbb{R})$ instead of $\mathrm{U}(n, \mathbb{C})$. Thus $\mathrm{U}(n, \mathbb{C})$ and $\mathrm{O}(n, \mathbb{R})$ are maximal compact subgroups of $\mathrm{GL}(n, \mathbb{C})$ and $\mathrm{GL}(n, \mathbb{R})$, respectively. That any other maximal compact subgroup is conjugate to one of these now follows from the first statement of the proposition. ∎

In particular, if L is any compact subgroup of $\mathrm{GL}(n, \mathbb{C})$, all its elements have their eigenvalues on the unit circle. From this we see that if an element $l \in L$ has all its eigenvalues equal to 1, then $l = I$. This is because glg^{-1} is unitary. Hence for some u we know $uglg^{-1}u^{-1}$ is diagonal and also has all eigenvalues equal to 1. Thus $uglg^{-1}u^{-1} = I$ and hence l itself equals I.

Finally, we turn to the proof of part 4 itself.

Proof. First suppose L is any compact subgroup of G. Then $L \cap P = (1)$. To see this just observe that, by the theorem below, since L is compact, all its elements have all their eigenvalues on the unit circle. But the eigenvalues of elements of P are all positive. Hence all the elements of L have all their eigenvalues equal to 1 and so, as above, each $l = I$. Now let $L \supseteq K$. Then each $l \in L$ can be written $l = kp$, where $k \in K$ and $p \in P$. But since $k \in L$, so is p. Hence by the above $p = I$ and $l = k$. Hence $L \subseteq K$, so that actually $L = K$. Thus K is a maximal compact subgroup of G. ∎

We have essentially used the conjugacy of maximal compact subgroups in $\mathrm{GL}(n, \mathbb{C})$ and $\mathrm{GL}(n, \mathbb{R})$ to show that K is a maximal compact subgroup of G, in general. However to prove, in general, that any two maximal compact subgroups of G are conjugate will require something more. For this we will rely on the important differential geometric fact, called *Cartan's fixed point theorem*, that a compact group of isometries acting on a complete simply connected Riemannian manifold of nonpositive sectional curvature at every point (Hadamard manifold) always has a unique fixed point and, for the reader's convenience, we will prove Cartan's result as well in the next section. However, we will only prove it for symmetric spaces of noncompact type. This will also establish the fact that for each $p \in P$, $\mathrm{Stab}_G(p)$ is a maximal compact subgroup of G.

We note that the Cartan involution of $\mathfrak{g}$ is given by $k + p \mapsto k - p$. It is an automorphism of $\mathfrak{g}$ whose fixed point set is $\mathfrak{k}$. We also mention the *Cartan relations*, which were also proved earlier. If the Cartan decomposition of $\mathfrak{g}$ is $\mathfrak{g} = \mathfrak{k} \oplus \mathfrak{p}$, since $\mathfrak{k}$ is a subalgebra and $[x^*, y^*] = -[x, y]^*$ and $[x^t, y^t] = -[x, y]^t$ it follows that

1. $[\mathfrak{k}, \mathfrak{k}] \subseteq \mathfrak{k}$,
2. $[\mathfrak{k}, \mathfrak{p}] \subseteq \mathfrak{p}$,
3. $[\mathfrak{p}, \mathfrak{p}] \subseteq \mathfrak{k}$.

We conclude this section by observing that for all the G we are dealing with there is a natural smooth action of G on P given by $(g, p) \mapsto g^t p g$. Now this action is transitive. To see this, consider the G orbit of $I \in P$, $\mathcal{O}_G(I) = \{g^t g : g \in G\}$. As we saw earlier, this is $\{p^2 ; p \in P\}$. But since everything in P is exp of a unique element x of $\mathfrak{p}$, it follows that everything in P has a unique square root in P, namely $\exp \frac{1}{2} x$. This means the action is transitive. What is the isotropy group of $\mathrm{Stab}_G(I)$ of I? This is $\{g \in G : g^t g = I\} = G \cap \mathrm{O}(n, \mathbb{R}) = K$. Hence, by general principles, (G, P) is G-equivariantly diffeomorphic with G acting by left translation on G/K. As we shall see, this transitive action will be of great importance in what follows.

Observe that this action does not have the two-point homogeneity property. That is, given p, q and p', q', all in P, there may not be a $g \in G$ so that $g(p) = p'$ and $g(q) = q'$, even when $\dim P = 1$. Note also that $g^t(\exp x)g$ is not equal to $\exp(g^t x g)$ so this is not equivariant with the $\mathbb{R}$-linear representation of G acting on $\mathfrak{p}$ by $(g, x) \mapsto g^t x g$, $x \in \mathfrak{p}$. Concommitantly, the latter is not a transitive action because it is linear, so 0 is a single orbit. In fact, here the orbit space can be parametrized by the number of positive, negative and zero eigenvalues of a representative.

4 The case of Hyperbolic Space and the Lorentz Group

We now make explicit the Cartan decomposition in an important special case and give the Lorentz model for hyperbolic n space, H^n. We consider $\mathrm{O}(n, 1)$ the subgroup of $\mathrm{GL}(n+1, \mathbb{R})$ leaving invariant the nondegenerate quadratic form $q(v, t) = v_1^2 + \ldots + v_n^2 - t^2$, where $v \in \mathbb{R}^n$ and $t \in \mathbb{R}$. Equivalently, by polarization, this means leaving invariant the nondegenerate symmetric bilinear form $\langle (v, t), (w, s) \rangle = (v, w) - ts$, where (v, w) is the usual (positive definite) inner product in $\mathbb{R}^n$. Thus G is defined by the condition $g^{-1} = g^t$ (with respect to $\langle , \rangle$). It is easy to check that G is the set of $\mathbb{R}$-points of a self-adjoint algebraic group and, in particular, is a Lie group. Now

G is not compact. For example, $\mathrm{SO}(1,1) \subseteq \mathrm{O}(1,1)$, which sits inside $\mathrm{O}(n,1)$, is given as follows.

$$g = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

One checks easily that $g \in \mathrm{SO}(1,1)$ if and only if $a^2 - c^2 = 1$, $ab - cd = 0$ and $b^2 - d^2 = -1$. In particular, taking an arbitrary a and $c = (a^2-1)^{\frac{1}{2}}$, where $a^2 - 1 = c^2 > 0$ and letting b and d be determined by the remaining two equations we see that $b = (a^2-1)^{\frac{1}{2}} = c$ and $d = a$. Now consider the identity component $\mathrm{SO}_0(1,1)$. Since the locus $a^2 - c^2 = 1$ has two connected components, if $g \in \mathrm{SO}_0(1,1)$, then $a > 0$ and so there is a unique $t \in \mathbb{R}$ for which $a = \cosh t$ and $b = \sinh t$. Thus

$$g(t) = \begin{pmatrix} \cosh t & \sinh t \\ \sinh t & \cosh t \end{pmatrix}$$

Because these hyperbolic functions are unbounded, we see even $\mathrm{SO}_0(1,1)$ is not compact. The identities satisfied by the hyperbolic functions show that this is an abelian subgroup. However, we shall see this without these identities; in fact, we will derive the identities. Let

$$X = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

A direct calculation using the fact that $X^2 = I$ shows that $\exp tX = I\cosh t + X\sinh t = g(t)$, from which it follows that $g(s+t) = g(s)g(t)$. This equation gives all the identities satisfied by the hyperbolic functions sinh and cosh and g is a smooth isomorphism of $\mathrm{SO}_0(1,1)$ with $\mathbb{R}$. The geometric importance of such 1-parameter subgroups will be seen in a moment.

By Theorem 3.6 a maximal compact subgroup of G is given by $\mathrm{O}(n+1,\mathbb{R}) \cap \mathrm{O}(n,1)$. Because subgroups of $\mathrm{GL}(n,\mathbb{R})$ can be regarded as subgroups of $\mathrm{GL}(n+1,\mathbb{R})$ via the imbedding $g \mapsto \mathrm{diag}(g,1)$, we may regard $\mathrm{O}(n,\mathbb{R})$ as a subgroup of $\mathrm{GL}(n+1,\mathbb{R})$ and, in fact, of $\mathrm{O}(n,1)$. Thus $\mathrm{O}(n,\mathbb{R}) \subseteq \mathrm{O}(n+1,\mathbb{R}) \cap \mathrm{O}(n,1)$. Clearly these are equal. Since $\mathrm{O}(n,\mathbb{R})$ has two components, so does $\mathrm{O}(n,1)$ which equals $\mathrm{O}(n,\mathbb{R})\times$ an exponential submanifold, P. Therefore, $\mathrm{O}(n,1)_0 = \mathrm{SO}(n,\mathbb{R}) \times P$. To identify this connected group, we note that $g^{-1} = g^t$, $gg^t = I$ and so $(\det g)^2 = 1$. Thus $\det g = \pm 1$, a discrete set. It follows that $\mathrm{SO}(n,1)$ is open in $\mathrm{O}(n,1)$ and hence has the same P. The same is true of $\mathrm{SO}_0(n,1)$ because we are dealing with Lie groups. Thus $\mathrm{SO}_0(n,1) = \mathrm{SO}(n,\mathbb{R}) \times P = G$ and we now work with this connected group. [2]

As with a Lie group defined by any nondegenerate bilinear form, the Lie algebra $\mathfrak{g}$ of $G = \mathrm{SO}_0(n,1)$ is

$$\{X \in \mathfrak{gl}(n+1,\mathbb{R}) : X^t = -X\}.$$

This Lie algebra evidently has dim $= \frac{(n+1)n}{2}$. Now consider the subspace of $\mathfrak{gl}(n+1,\mathbb{R})$ consisting of

[2] Actually, $\mathrm{SO}(n,1)$ is connected if n is even, and has two components if n is odd.

$$\begin{pmatrix} \mathfrak{k} & v \\ v & 0 \end{pmatrix},$$

where $\mathfrak{k}$ is the Lie algebra of $\mathrm{SO}(n, \mathbb{R})$ and $v \in \mathbb{R}^n$. It is clearly a subspace and has dimension $\frac{(n-1)n}{2} + n = \frac{(n+1)n}{2}$. Now a direct calculation, which we leave to the reader, shows that this is a subspace of $\mathfrak{g}$, i.e., it consists of skew symmetric matrices with respect to $\langle , \rangle$. Hence it must coincide with $\mathfrak{g}$. Here the Cartan decomposition is perfectly clear. The $\mathfrak{k}$ part is

$$\begin{pmatrix} \mathfrak{k} & 0 \\ 0 & 0 \end{pmatrix},$$

while the $\mathfrak{p}$ part is

$$\begin{pmatrix} 0 & v \\ v & 0 \end{pmatrix}.$$

Consider the locus of points,

$$H = \{(v, t) \in \mathbb{R}^{n+1} : q(v, t) = -1\}.$$

For $g \in O(n, 1)$, $q(g(v, t)) = q(v, t)$. In particular, if $q(v, t) = -1$, then $q(g(v, t)) = -1$. Thus H is invariant under $\mathrm{O}(n, 1)$. Now H is a hyperboloid of two sheets: $1 + \| v \|^2 = t^2$. So $t = \pm(1 + \| v \|^2)^{\frac{1}{2}}$. Write $H = H^+ \cup H^-$, a disjoint union of the upper and lower sheet. Both sheets are open subsets of H since they are the intersection of the sheet with a half space. Each is diffeomorphic with $\mathbb{R}^n$. In particular, each is connected and simply connected. We show that $G = \mathrm{SO}_0(n, 1)$ leaves both H^+ and H^- invariant.

If $g \in G$, then $g(H^+)$ and $g(H^+) \subseteq H$. So

$$g(H^+) = (g(H^+) \cap H^+) \cup (g(H^+) \cap H^-).$$

But H^+ is connected and g is continuous. Hence $g(H^+)$ is connected. Therefore $g(H^+) \subseteq H^+$ or $g(H^+) \subseteq H^-$. Since g itself is a diffeomorphism, $g(H^+) = H^+$ or $g(H^+) = H^-$. We show the former must hold. Since G is itself connected and therefore arcwise connected, let g_t be a smooth path in G joining $g = g_1$ to $I = g_0$ and let $T^+ = \{t \in [0, 1] : g_t(H^+) = H^+\}$ and $T^- = \{t \in [0, 1] : g_t(H^+) = H^-\}$. Suppose $g(H^+) = H^-$. Since $I(H^+) = H^+$, we have $[0, 1] = T^+ \cup T^-$, a disjoint union of nonempty sets. Each of these is closed. For if $t_k \to t$ and say $g_{t_k}(H^+) = H^+$, for all k, but $g_t(H^+) = H^-$, then for $x \in H^+$, $g_{t_k}(x) \to g_t(x)$. This is impossible as the distance between H^+ and H^- is 2.

We now know G operates on H^+ which we shall call H^n, the Lorentz model of hyperbolic n-space. Consider the lowest point, $p_0 = (0, \ldots, 0, 1) \in H^n$. What is $\mathrm{Stab}_G(p_0)$? This is clearly a subgroup which does not change the t coordinate and is arbitrary in the other coordinates since it is linear and so always fixes 0. Hence, $\mathrm{Stab}_G(p_0) = \mathrm{SO}(n, \mathbb{R})$, a maximal compact subgroup of G. Next we look at G-orbit $\mathcal{O}(p_0)$ and show G acts transitively on H^n. Let p = (v,t), where $t = (1 + \| v \|^2)^{\frac{1}{2}}$, be any point in H^n and apply $\mathrm{SO}(n,\mathbb{R})$ to bring it to $(\| v \|, 0, ...0, t)$. Since we are now essentially in a two-dimensional situation, let us consider (x,y), where $y^2 - x^2 = 1$. We want to transform $(0,1)$ to p by something on the 1-parameter group

$$g(s) = \begin{pmatrix} \cosh s & \sinh s \\ \sinh s & \cosh s \end{pmatrix}.$$

But this is just the fundamental property of the right hand branch of the hyperbola mentioned earlier. Therefore, G acts transitively and H^n is equivariantly equivalent to $\mathrm{SO}_0(n,1)/\mathrm{SO}(n,\mathbb{R})$.

Now consider the hyperplane $t = 1$ in $\mathbb{R}^{n+1}$. This is the tangent space $T(p_0)$ to H^n at p_0. Thus there is a positive definite metric, namely $(,)$ on $T(p_0)$. If p is another point of H^n, choose $g \in G$ such that $g(p) = p_0$. Then $d_g(p)$ maps $T(p)$ to $T(p_0)$. Since g comes from a group, it is invertable. By the chain rule so is its derivative $d_g(p)$, so it maps $T(p)$ to $T(p_0)$ bijectively. Use this to transfer the inner product from $T(p_0)$ to $T(p)$. Now if $h(p)$ also equals p_0, then $gh^{-1} \in \mathrm{Stab}_G(p_0) = \mathrm{SO}(n,\mathbb{R})$. Therefore $d_g(p)d_h(p)^{-1}$ is a linear isometry in $T(p_0)$. This shows the inner product on $T(p)$ is independent of g and is well defined. Hence we get a Riemannian metric on H^n because G is a Lie group acting smoothly on H^n. Evidently, G acts by isometries, the action is transitive and H^n can be identified with $G/\mathrm{Stab}_G(p_0) = \mathrm{SO}_0(n,1)/\mathrm{SO}(n,\mathbb{R})$.

Notice that $\mathrm{SO}(n,\mathbb{R}) = \mathrm{Stab}_G(p_0)$ acts transitively on k-dimensional subspaces for all $1 \leq k \leq n$. In particular, this is so for 2-planes in $\mathbb{R}^n = T(p_0)(H^n)$. Since it acts by isometries, this means the sectional curvature is constant as both the point and the plane section vary.

5 The G-invariant Metric Geometry of P

Here we introduce a Riemannian metric on any P and study its most basic differential geometric properties. From now on we will write exp and log instead of Exp and Log.

Lemma 5.1 *If A and B are $n \times n$ complex matrices, then* $\mathrm{tr}(AB) = \mathrm{tr}(BA)$. *Also* $\mathrm{tr}(B^*B) \geq 0$ *and equals* 0 *if and only if* $B = 0$. *Evidently,* $\mathrm{tr}(B)^- = \mathrm{tr}(B^*)$.

Proof. Suppose $A = (a_{i,j})$ and $B = (b_{k,l})$. Then $(AB)_{i,l} = \sum_j a_{i,j}b_{j,l}$. Therefore $\mathrm{tr}(AB) = \sum_{i,j} a_{i,j}b_{j,i}$. But then $\mathrm{tr}(BA) = \sum_{i,j} b_{i,j}a_{j,i} = \sum_{i,j} a_{j,i}b_{i,j} = \sum_{j,i} a_{i,j}b_{j,i} = \mathrm{tr}(AB)$. Taking B^* for A we get $\mathrm{tr}(B^*B) = \sum_{i,j} b_{j,i}^- b_{j,i} \geq 0$ and equals 0 if and only if $B = 0$. ∎

This enables us to put a Hermitian inner product on $\mathfrak{gl}(n,\mathbb{C})$ called the *Hilbert Schmidt inner product* and a symmetric inner product on $\mathfrak{gl}(n,\mathbb{R})$ by defining

$$< Y, X >= \mathrm{tr}(Y^*X).$$

For X Hermitian (symmetric), we now study the linear operator ad_X on $\mathfrak{gl}(n,\mathbb{C})$ ($\mathfrak{gl}(n,\mathbb{R})$). As we saw from the Cartan relations for $T \in \mathfrak{gl}(n,\mathbb{C})$ and X Hermitian, $[X,T]^* = [T^*, X] = -[X, T^*]$.

Lemma 5.2 *If X is Hermitian, $< \mathrm{ad}_X(T), S >=< T, \mathrm{ad}_X(S) >$ for all S and T; that is, ad_X is self-adjoint.*

In particular, the eigenvalues of such an ad_X are all real. (This gives a direct proof of the fact that $d(\exp)_X$ is invertible for all $X \in \mathfrak{p}$.)
Proof. We calculate $\mathrm{tr}([X,T]^*S) = \mathrm{tr}(-[X,T^*]S) = -\mathrm{tr}((XT^* - T^*X)S) = \mathrm{tr}(T^*XS) - \mathrm{tr}(XT^*S)$. On the other hand, $\mathrm{tr}(T^*[X,S]) = \mathrm{tr}(T^*XS) - tr(T^*SX)$. Thus we must show that $\mathrm{tr}(XT^*S) = \mathrm{tr}(T^*SX)$. But this follows from the lemma above. ∎

A formal calculation, which we leave to the reader, proves the following:

Lemma 5.3 *For each $U \in \mathfrak{gl}(n, \mathbb{C})$, $L_{\exp(U)} = \exp(L_U)$ and $R_{\exp(U)} = \exp(R_U)$.*

Definition 5.4 *For X and $Y \in \mathfrak{gl}(n, \mathbb{C})$ let*

$$d_X(Y) = \frac{d}{dt} \exp(-X/2) \exp(X + tY) \exp(-X/2)_{|t=0}.$$

Proposition 5.5 *For $X \in \mathfrak{p}$, the operator d_X is self-adjoint on $\mathfrak{gl}(n, \mathbb{C})$. Using functional calculus, this operator is given by the formula*

$$d_X = \sinh(\frac{\mathrm{ad}_X}{2})/\frac{\mathrm{ad}_X}{2}.$$

Proof. Let $t \in \mathbb{R}$, X, $Y \in \mathfrak{gl}(n, \mathbb{C})$ and $X(t) = X + tY$. Then

$$d_X(Y) = \exp(-X/2) \frac{d}{dt} \exp(X(t))_{|t=0} \exp(-X/2).$$

Now for all t,

$$X(t) \cdot \exp(X(t)) = \exp(X(t) \cdot X(t).$$

Differentiating we get

$$X'(t) \cdot \exp(X(t)) + X(t) \cdot \frac{d}{dt} \exp(X(t)) = \frac{d}{dt} \exp(X(t)) \cdot X(t) + \exp(X(t) \cdot X'(t).$$

Evaluating at $t = 0$ and subtracting gives $X \cdot \frac{d}{dt} \exp(X(t))_{|t=0} - \frac{d}{dt} \exp(X(t))_{|t=0} \cdot X = \exp(X)Y - Y \exp(X)$. Multiplying on both the left and right by $\exp(-X/2)$ and taking into account the fact that $\exp(-X/2)$ and X commute, we get
$X \cdot \exp(-X/2) \frac{d}{dt} \exp(X(t))_{|t=0} \exp(-X/2) - \exp(-X/2) \frac{d}{dt} \exp(X(t))_{|t=0} \exp(-X/2) X = \exp(X/2) Y \exp(-X/2) - \exp(-X/2) Y \exp(X/2)$. Substituting for $d_X(Y)$, the left hand side becomes $X d_X(Y) - d_X(Y) X = \mathrm{ad}_X\, d_X(Y)$, while the right hand side is

$$L_{\exp(X/2)} R_{\exp(-X/2)}(Y) - L_{\exp(-X/2)} R_{\exp(X/2)}(Y).$$

But by the lemma above

$$L_{\exp(U)} = \exp(L_U) \text{ and } R_{\exp(U)} = \exp(R_U).$$

Substituting we get

$$\operatorname{ad}_X d_X(Y) = \exp(L_{X/2})\exp(R_{-X/2})(Y) - \exp(L_{-X/2})\exp(R_{X/2})(Y).$$

Since L_U and $R_{U'}$ commute for all U and U', we see that

$$\exp(L_{X/2})\exp(R_{-X/2}) = \exp(L_{X/2} + R_{-X/2}) = \exp(L_{X/2} - R_{X/2}) = \exp(\operatorname{ad}_X/2).$$

Similarly,

$$\exp(L_{-X/2})\exp(R_{X/2}) = \exp(L_{-X/2} + R_{X/2}) = \exp(-\operatorname{ad}_X/2).$$

So for all Y,

$$\operatorname{ad}_X \cdot d_X(Y) = (\exp(\operatorname{ad}_X/2) - \exp(-\operatorname{ad}_X/2))(Y).$$

Now let

$$f(z) = e^{z/2} - e^{-z/2} = z + 2(z/2)^3/3! + 2(z/2)^5/5! + \dots.$$

Then f is an entire function and 0 is a removable singularity with $f(0) = 0$. In terms of f, the equation above says $\operatorname{ad}_X d_X = f(\operatorname{ad}_X)$. This means if we let

$$g(z) = f(z)/z = 1 + (z/2)^2/3! + (z/2)^4/5! + \dots,$$

with $g(0) = 1$, then g is also entire and $d_X = g(\operatorname{ad}_X)$. Now $\sinh z = z + z^3/3! + z^5/5! + \dots$ so $g(z) = \sinh(z/2)/(z/2)$ and hence the conclusion. Finally, because $d_X = g(\operatorname{ad}_X)$, ad_X is self-adjoint and the Taylor coefficients of g are real, d_X is also self-adjoint. ∎

Corollary 5.6 *For $X \in \mathfrak{p}$,* $\operatorname{Spec}(\frac{\sinh(\operatorname{ad}_X)}{\operatorname{ad}_X})$ *consists of real numbers greater than or equal to 1. The same is so for the operator d_X.*

Proof. Since for $t \in \mathbb{R}$, $\frac{\sinh t}{t} = 1 + t^2/3! + t^4/5! + \dots$, we see that $\frac{\sinh t}{t} > 1$ unless $t = 0$. Now

$$\operatorname{Spec}(\frac{\sinh(\operatorname{ad}_X)}{\operatorname{ad}_X}) = \{\frac{\sinh(\lambda)}{\lambda} : \lambda \in \operatorname{Spec}\operatorname{ad}_X\} \subseteq \{\frac{\sinh(\lambda_i - \lambda_j)}{\lambda_i - \lambda_j}\} : \lambda_i, \lambda_j \in \operatorname{Spec} X\}.$$

If $\lambda = \lambda_i - \lambda_j$ for distinct eigenvalues of X, then $\frac{\sinh(\lambda)}{\lambda} > 1$. If λ_i and λ_j are equal, then $\lambda = 0$ and $\frac{\sinh(\lambda)}{\lambda} = 1$. ∎

We now work exclusively over $\mathbb{R}$. The same type of arguments also work just as well over $\mathbb{C}$.

Corollary 5.7 *For $X \in \mathfrak{p}$ and $Y \in \mathfrak{gl}(n, \mathbb{R})$, $\operatorname{tr}(Y^2) \leq \operatorname{tr}(d_X(Y))^2)$. Equality occurs if and only if $[X, Y] = 0$.*

Proof. Because ad_X is self-adjoint, we can choose an orthonormal basis of real eigenvectors of ad_X, $Y_1, \dots Y_j \in \mathfrak{gl}(n, \mathbb{R})$ which, since $d_X = g(\mathrm{ad}_X)$ are also eigenvectors for d_X with corresponding real eigenvalues $\mu_1, \dots \mu_j$. Then $d_X(Y_k) = \mu_k Y_k$ for all k. If $Y = \sum_k a_k(Y) Y_k$, then $d_X(Y) = \sum_k a_k(Y) d_X(Y_k) = \sum_k a_k(Y)\mu_k Y_k$. Since the Y_k form an orthonormal basis, we see $\mathrm{tr}(d_X(Y)^2) = \sum_k a_k(Y)^2 \mu_k^2$, while $\mathrm{tr}(Y^2) = \sum_k a_k(Y)^2$. Thus we are asking whether $\sum_k a_k(Y)^2 \le \sum_k a_k(Y)^2 \mu_k^2$. Since each $\mu_k \ge 1$, this is clearly so and equality occurs only if $\mu_k = 1$ whenever $a_k(Y) \neq 0$. Rearrange the eigenvectors so that the $\mu_k = 1$ come first and for $k \ge k_0$, $\mu_k > 1$. Hence $\mathfrak{gl}(n, \mathbb{R}) = W_1 \oplus W_\infty$ is the orthogonal direct sum of two ad_X-invariant subspaces. Here W_1 is the 1-eigenspace, and W_∞ the sum of all the others. But since $a_k(Y) = 0$ for $k \ge k_0$, $Y \in W_1$. But on W_1 all eigenvalues of $g(\mathrm{ad}_X) = d_X$ are 1, and the eigenvalues of ad_X are 0 so $\mathrm{ad}_X = 0$ on W_1 and hence $[X, Y] = 0$.

Conversely, if $[X, Y] = 0$, then $\mathrm{ad}_X(Y) = 0$. Therefore $d_X = g(\mathrm{ad}_X) = I$. Hence $W_1 = \mathfrak{gl}(n, \mathbb{R})$ and $W_\infty = (0)$. Therefore all $\mu_k = 1$ and equality holds. ∎

Theorem 5.8 *Along any smooth path $p(t)$ in P we have*

$$\mathrm{tr}((\frac{d}{dt}\log p(t))^2) \le \mathrm{tr}(p^{-1}p'(t))^2).$$

with equality if and only if $p(t)$ and $p'(t)$ commute for that t. In particular, $\mathrm{tr}(p^{-1}p'(t))^2) \ge 0$ since if $X(t) = \log p(t) \in \mathfrak{p}$, then $X'(t) \in \mathfrak{p}^{\cdot} = \mathfrak{p}$. Hence $\mathrm{tr}(X'(t)^t X'(t)) \ge 0$.

Proof. For each t, it is easy to see that

$$p^{\frac{1}{2}} p^{-1} p' p^{-\frac{1}{2}} = (p^{-\frac{1}{2}} p' p^{-1} p^{-\frac{1}{2}})^2.$$

It follows that $\mathrm{tr}(p^{-1}p'(t))^2) = \mathrm{tr}(p^{-\frac{1}{2}} p' p^{-1} p^{-\frac{1}{2}})^2)$. Set $X(t) = \log p(t)$. Then $X(t)$ is a smooth path in $\mathfrak{p}$ and $p^{-\frac{1}{2}}(t) = \exp(-X(t)/2)$. Let t be fixed and $Y = X'(t)$. Since $d_X(Y) = \exp(-X/2)\frac{d}{ds}\exp(X + sY))_{|s=0}\exp(-X/2)$, this is $p^{-\frac{1}{2}} p' p^{-\frac{1}{2}}$, where $p' = \frac{d}{ds}\exp(X + sY))_{|s=0}$ (the tangent vector to curve $p(t)$ at $p = \exp X$). Hence $\mathrm{tr}(p^{-\frac{1}{2}} p' p^{-\frac{1}{2}} 2 = \mathrm{tr}(d_X(X'))^2$. Also $\mathrm{tr}(\frac{d}{dt}\log p(t))^2 = \mathrm{tr}\, X'(t)^2$. Now by the corollary, for each t, $\mathrm{tr}\, X'(t)^2 \le \mathrm{tr}(d_{X(t)}(X'(t))^2$ with equality if and only if $X(t)$ and $X'(t)$ commute for that t. Finally we show that if $\exp X(t) = p(t)$, then for fixed t, $X(t)$ and $X'(t)$ commute if and only if $p(t)$ and $p'(t)$ commute. For by the chain rule and the formula for $d(\exp)$ (see L.S. Varadarajan [11]),

$$p'(t) = d(\exp)_{X(t)} X'(t) = \phi(-\mathrm{ad}\, X(t)) X'(t),$$

where ϕ is the entire function given by $\phi(z) = \sum_{n=0}^{\infty} \frac{z^n}{(n+1)!}$.

If X' commutes with X for fixed t, then since $\phi(0) = 1$, we see that $\phi(-\mathrm{ad}(X))X' = X'$ so that $p' = X'$. In particular, p' commutes with X and therefore with $\exp X = p$. On the other hand, if $\phi(-\mathrm{ad}(X))X'$ commutes with $\exp X = p$, then since $\log : P \to \mathfrak{p}$ is given by a convergent power series in p (see Theorem 3.6, part 3), it must also commute with $\log p = X$. Looking at the specific form of the function ϕ, it follows that

$$[X, X' - \mathrm{ad}(X)(X')/2! + \mathrm{ad}^2(X)(X')/3! \dots] = 0.$$

That is, $\operatorname{ad}(X)(X') - \operatorname{ad}^2(X)(X')/2! + \operatorname{ad}^3(X)(X')/3! \ldots = 0$. Hence $\exp(-\operatorname{ad}(X)(X')) = X'$. Taking $\exp(\operatorname{ad}_X)$ of both sides tells us $\exp(\operatorname{ad}_X)(X') = X'$. Therefore $\operatorname{Ad}_{\exp X}(X') = X'$ so X' commutes with $\exp X$. But then, reasoning as above, X' must commute with $\log(\exp X) = X$. ∎

Since what is inside the square root is real and positive, we make the following definition.

Definition 5.9 *Let $p(t)$ be a smooth path in P, where $a \leq t \leq b$. Then its length $l(p) = \int_a^b \operatorname{tr}(p^{-1}p'(t))^2)^{\frac{1}{2}}dt$. The Riemannian metric is given by $ds^2 = \operatorname{tr}((p^{-1}p')^2)dt^2$. We call this metric d.*

Proposition 5.10 *G acts isometrically on P.*

Proof. We calculate that

$$(g^tpg)^{-1}(g^tpg)' = g^{-1}p^{-1}(g^t)^{-1}g^tp'g = g^{-1}p^{-1}p'g.$$

Hence $((g^tpg)^{-1}(g^tpg)')^2 = g^{-1}(p^{-1}p')^2g$. Taking traces we get

$$\operatorname{tr}((g^tpg)^{-1}(g^tpg)')^2) = \operatorname{tr}(p^{-1}p')^2).$$

∎

On $\mathfrak{p}$ we place the metric given infinitesimally by $ds^2 = \operatorname{tr}((\frac{d}{dt}\log p(t))^2)dt^2$, that is, if $X(t)$ is a smooth path in $\mathfrak{p}$, then $ds^2 = \operatorname{tr}(X'(t)^2)dt^2$. We call this metric $d_{\mathfrak{p}}$. Earlier we defined an inner product on $\mathfrak{gl}(n, \mathbb{R})$ by $< Y, X >= \operatorname{tr}(Y^tX)$. Hence the linear subspace $\mathfrak{p}$ has an inner product on it by restriction, namely $< Y, X >= \operatorname{tr}(YX)$. The associated norm is $\| Y \|^2 = \operatorname{tr}(Y^2)$. This, together with the formula above, shows $d_{\mathfrak{p}}$ is the Euclidean metric. If we transfer $d_{\mathfrak{p}}$ to P, then $d_{\mathfrak{p}}(p, q) = \| \log p - \log q \|$. This will give us the opportunity to compare $d_{\mathfrak{p}}$ and d on P. Since along any smooth path $p(t)$ in P we have $\operatorname{tr}(\frac{d}{dt}\log p(t))^2) \leq \operatorname{tr}(p(t)^{-1}p'(t))^2$, we see that infinitesimally and hence globally $d_{\mathfrak{p}} \leq d$.

Now for $X \in \mathfrak{p}$, $d_X = \frac{\sinh(\operatorname{ad}_X/2)}{\operatorname{ad}_X/2}$. It follows that

$$\operatorname{Spec} d_X = \{\frac{\sinh(\lambda/2)}{\lambda/2} : \lambda \in \operatorname{Spec} \operatorname{ad}_X\}.$$

As $\frac{\sinh t}{t}$ is analytic, by continuity $\frac{\sinh t}{t} \to 1$ as $t \to 0$. This tells us that from the formulas for $\operatorname{tr}(d_X(Y))^2$ and $\operatorname{tr}(Y^2)$, if $X \to 0$, then independently of Y, $\operatorname{tr}(d_X(Y))^2$ can be made as near as we want to $\operatorname{tr}(Y^2)$. This last statement implies that for p and q in a sufficiently small neighborhood of a point p_0, which by transitivity of G we may assume to be I, the nonpositively curved symmetric space and Euclidean distances approach one another.

$$\lim_{p,q \to p_0} \frac{d(p,q)}{d_{\mathfrak{p}}(p,q)} = 1.$$

This has the interesting philosophical consequence that in the nearby part of the universe that man inhabits, because of experimental error in making measurements, nonpositively curved symmetric space distances and Euclidean ones are indistinguishable. As we shall show below, angles at I are in any case identical. This means no experiment can tell us if we "really" live in a hyperbolic or Euclidean world.

Corollary 5.11 *If $p = \log X \in P$, the 1-parameter subgroup $\exp tX$ is the unique geodesic in (P, d) joining I with p. Moreover, any two points of P can be joined by a unique geodesic. (We shall see rather explicitly and directly how this geodesic is determined by its initial and terminal points.)*

This Corollary also follows from more general facts in differential geometry. This is because as a 1-parameter subgroup every geodesic emanating from I has infinite length. Since G acts transitively by isometries, this is true at every point. Hence by the Hopf-Rinow theorem (see J. Milnor [7]) P is complete. In particular, any two points can be joined by a shortest geodesic (also Hopf-Rinow). Being diffeomorphic to Euclidean space, P is simply connected. If P had nonpositive sectional curvature in every section and at every point, then this geodesic would be unique. This last fact is actually valid for any Hadamard manifold and is called the Cartan-Hadamard theorem. We will give a direct proof of completeness of P shortly.
Proof. Consider a path $p(t)$ in P which happens to be a 1-parameter subgroup. Since $p(t) = \exp tX$, $\log p(t) = tX$ and its derivative is X. Thus for each t, $\log p(t)$ and its derivative commute. Hence, as we showed, $p(t)$ and $p'(t)$ also commute. This tells us that all along $p(t)$, $d_{\mathfrak{p}}$ and d coincide. But the 1-dimensional subspaces of $\mathfrak{p}$ are geodesics for $d_{\mathfrak{p}}$. Hence if $p = \log X \in P$, the 1-parameter subgroup $\exp tX$ is the unique geodesic in (P, d) joining I with p. Let p and q be distinct points of P. Since G acts transitively on P, we can choose g so that $g(q) = I$. Connect I with $g(p)$ by its unique geodesic γ. Since G acts isometrically, $g^{-1}(I) = q$, $g^{-1}(g(p)) = p$ and $g^{-1}(\gamma)$ is the unique geodesic joining them. ∎

Corollary 5.12 *A curve $p(t)$ in P is a geodesic through $p_0 \in P$ if and only if $p(t) = g(\exp tX)g^t$, where $X \in \mathfrak{p}$ and $g \in G$.*

Proof. Since G acts transitively by isometries on P, choose $g \in G$ so that $gIg^t = p_0$. The result follows from the above since the 1-parameter subgroup $\exp tX$ is the unique geodesic in (P, d) beginning at I in the direction X. ∎

Corollary 5.13 *At I the angles in the two metrics coincide.*

Proof. Let X and Y be two vectors in $\mathfrak{p}$ and $p(t)$ and $q(t)$ be curves in P passing through I with tangent vectors X and Y, respectively, and let $p_0(t) = \exp tX$ and $q_0(t) = \exp tY$ be two 1-parameter groups in P. Then since X and Y are also the tangent vectors of p_0 and q_0, respectively, the angle between p and q equals that between p_0 and q_0. We may therefore replace p and q by p_0 and q_0. Now $p^{-1}p'q^{-1}q'(0)$ is just XY so that $\operatorname{tr}(p^{-1}p'q^{-1}q'(0)) = \operatorname{tr}(XY)$. ∎

Corollary 5.14 *For $X \in \mathfrak{p}$, $d(I, \exp X) = \operatorname{tr}^{\frac{1}{2}}(X^2)$.*

Proof. The 1-parameter group $\exp tX$ is a geodesic in P passing through I at $t = 0$. Hence, infinitesimally along this curve, $d = d_{\mathfrak{p}}$. This implies the same is true globally along it. Put another way, at each point of $\exp tX$, for $0 \leq t \leq 1$, the theorem tells us the metric is $\operatorname{tr}(\frac{d}{dt}(tX)^2) = \operatorname{tr}(X^2)$. Since this is independent of t, integrating from 0 to 1 gives $\operatorname{tr}(X^2)$. ∎

Corollary 5.15 *For X and $Y \in \mathfrak{p}$, $d(\exp X, \exp Y) \geq \operatorname{tr}^{\frac{1}{2}}((X-Y)^2)$.*

Corollary 5.16 *P is complete.*

Proof. Let p_k be a Cauchy sequence in (P, d). By the inequality above, $X_k = \log p_k$ is a Cauchy sequence in $(\mathfrak{p}, d_{\mathfrak{p}})$ which must converge to X since Euclidean space is complete. By continuity, p_k converges to $\exp X = p$. ∎

Corollary 5.17 *(Law of Cosines). Let a, b and c be the lengths of the sides of a geodesic triangle in P and A, B and C be the corresponding vertices. Then*

$$c^2 \geq a^2 + b^2 - 2ab\cos C$$

and the sum of the angles $A + B + C \leq \pi$. Moreover, if the vertex C is at I then equality holds if and only if

1. *The triangle lies in a connected abelian subgroup of P, or equivalently,*
2. *$A + B + C = \pi$.*

Proof. Put C at the identity via an isometry from G. Then the Euclidean angle at C equals the angle in the metric d. Also, $l_{\mathfrak{p}}(c) \leq c$ and $l_{\mathfrak{p}}(a) = a$ and $l_{\mathfrak{p}}(b) = b$. The inequality now follows from the Euclidean Law of Cosines. Equality holds if and only if $l_{\mathfrak{p}}(c) = c$. This occurs if and only if log takes side c to a geodesic in $\mathfrak{p}$ (i.e. a straight line) of the same length. This is also equivalent to $\operatorname{tr}((\frac{d}{dt}\log p(t))^2) = \operatorname{tr}(p^{-1}p'(t)^2)$, for all t, where $p(t)$ denotes the geodesic side of length c. This occurs if and only if $p(t)$ satisfies the condition that $p(t)$ and $p'(t)$ commute for all t which, as we showed, is equivalent [X,Y] $= 0$, where X and Y are the infinitesimal generators of the sides a and b. Thus equality in the Law of Cosines holds if and only if the Euclidean triangle lies in a two-dimensional abelian subalgebra of $\mathfrak{g}$ contained in $\mathfrak{p}$. Equivalently, the geodesic triangle lies in a two-dimensional abelian subgroup of G contained in P. Such a subgroup is called a *flat* of P.

Next we show that in general the sum of the angles $\leq \pi$. Since d is a metric and $c = d(A, B)$, etc., it follows that each length a, b, or c is less than the sum of the other two. Therefore there is an ordinary plane triangle with sides a, b and c. Denote its angles by A', B' and C'. Then $A \leq A'$, $B \leq B'$ and $C \leq C'$. For by the Law of Cosines $c^2 \geq a^2 + b^2 - 2ab\cos C$ and $c^2 = a^2 + b^2 - 2ab\cos C$. This means $\cos C' \leq \cos C$. But then because C and C' are between 0 and π and cos is monotone decreasing there, we see $C \leq C'$. Similarly, this holds for the others. Since $A' + B' + C' = \pi$, it follows that $A + B + C \leq \pi$.

If $c^2 > a^2 + b^2 - 2ab\cos C$, then, as above, construct an ordinary plane triangle with sides a, b and c and angles A', B' and C'. Then since here we have strict inequality, it follows as above that $C < C'$. But it is always the case that $A \leq A'$ and $B \leq B'$. Hence $A + B + C < A' + B' + C' = \pi$. Conversely, if $A + B + C = \pi$, then $c^2 = a^2 + b^2 - 2ab\cos C$ and $[X, Y] = 0$. Therefore X and Y generate an abelian subalgebra, and the triangle lies in a flat. ∎

Our next result is of fundamental importance. Nonpositive and positive sectional curvature distinguish the symmetric spaces of noncompact type from those of compact type.

Corollary 5.18 *The sectional curvature of P is ≤ 0 and strictly < 0 off flats. In particular, P is a Hadamard manifold.*

Before turning to the proof we remark that when $X, Y \in \mathfrak{p}$ and are orthonormal with respect to the Killing form, one actually has $K(X, Y) = - \parallel [X, Y] \parallel^2$ (see J. Cheeger and D. Ebin [1]). However, we shall not need this formula.

Proof. Each geodesic triangle lies in a plane section. We have just shown that each geodesic triangle in each such section has the sum of the angles $\leq \pi$ and the sum of the angles $< \pi$ if we are off a flat. It is a standard result of two-dimensional Riemannian geometry (Gauss-Bonnet theorem) that these conditions are equivalent to $K \leq 0$ and $K < 0$, respectively, where K denotes the Gaussian curvature of the section, that is, the sectional curvature. ∎

Definition 5.19 *A submanifold N of a Riemannian manifold M is called totally geodesic if given any two points of N and a geodesic γ in M joining them, γ lies entirely in N.*

Corollary 5.20 *P is a totally geodesic submanifold in the set of all positive definite symmetric matrices.*

Proof. Let p and $q \in P$. Since $p^{\frac{1}{2}}$ and $p^{-\frac{1}{2}}$ are self-adjoint, $p^{-\frac{1}{2}}qp^{-\frac{1}{2}}$ is positive definite and symmetric. But as we showed earlier, $p^{-\frac{1}{2}} \in G$. Hence $p^{-\frac{1}{2}}qp^{-\frac{1}{2}} \in G$. Because $p^{-\frac{1}{2}}$ is self-adjoint we see that $p^{-\frac{1}{2}}qp^{-\frac{1}{2}} \in P$. Let $X \in \mathfrak{p}$ be its log. Then $\exp tX$ lies in P, for all real t. Therefore $\gamma(t) = p^{\frac{1}{2}}(\exp tX)p^{\frac{1}{2}}$ is a geodesic in P. Clearly, $\gamma(0) = p$ and $\gamma(1) = q$. Since there is a unique geodesic joining these points in both P and in the positive definite matrices, this completes the proof. ∎

We conclude this section with the standard definition of a *Symmetric Space*.

Definition 5.21 *A Riemmanian manifold M is called a Symmetric Space if for each point $p \in M$ there is an isometry σ_p of M satisfying the following conditions.*

1. $\sigma_p^2 = I$, *but* $\sigma_p \neq I$

2. *σ_p has only isolated fixed points among which is p. (In our case, p is actually the only fixed point.)*

3. *$d(\sigma_p)$ on $T_p(M) = -I$.*

Thus the main feature of the definition is that for each point p there is an isometry which leaves p fixed and reverses geodesics through p.

Corollary 5.22 *P is a symmetric space.*

Proof. Since G acts transitively and by isometries, we may restrict ourselves to the case $p = I$. Take $\sigma_I = \sigma(p) = p^{-1}$, for each $p \in P$. This map is clearly of order 2. If p is σ fixed, then $p^2 = I$. Hence every conjugate of p also has order 2. This means $p \in K \cap P$, which as we saw earlier $= (I)$, Thus I is the only fixed point. Let $p = \exp X$, then $\sigma(p) = \exp(-X)$ so that $d(\sigma_p)_{\mathfrak{p}} = -I$, where here we identify $T_I(P)$ with $\mathfrak{p}$.

It remains to see that σ is an isometry. For a curve $p(t)$ in P, since $p(t)p(t)^{-1} = I$, differentiating tells us

$$\frac{d}{dt}(p(t)^{-1}) = -p(t)^{-1}\frac{dp}{dt}p(t)^{-1}.$$

Substituting we get

$$\operatorname{tr}(p\frac{dp^{-1}}{dt})^2 = \operatorname{tr}(p - p(t)^{-1}\frac{dp}{dt}p(t)^{-1}p - p(t)^{-1}\frac{dp}{dt}p(t)^{-1}p).$$

Cancelling the minus signs and pp^{-1}, we have $\operatorname{tr}(\frac{dp}{dt}p(t)^{-1}\frac{dp}{dt}p(t)^{-1})$. Since $\operatorname{tr}(AB) = \operatorname{tr}(BA)$, this is

$$\operatorname{tr}(p(t)^{-1}\frac{dp}{dt}p(t)^{-1}\frac{dp}{dt}) = \operatorname{tr}((p(t)^{-1}\frac{dp}{dt})^2.$$

Thus for every t,

$$\operatorname{tr}(p(t)\frac{dp^{-1}}{dt})^2 = \operatorname{tr}((p(t)^{-1}\frac{dp}{dt})^2.$$

Hence σ is an isometry of P and the latter is a symmetric space. ∎

6 The Conjugacy of Maximal Compact Subgroups

The theorem on the conjugacy of maximal compact subgroups of G in the present context is due to E. Cartan. Actually, the result is true for an arbitrary connected Lie group and is due to K. Iwasawa and the case of a finite number of components to G.D. Mostow. In this more general context see [5]. We shall deal with this problem in the present context by means of Cartan's fixed point theorem which states that a compact group of isometries acting on a complete, simply connected Riemannian manifold of nonpositive sectional curvature (Hadamard manifold) has a unique fixed point. However, here we will prove the fixed point theorem where we need it, namely, in the special case when the manifold is a symmetric space of noncompact type.

Theorem 6.1 *Let $f : C \to (P, d)$ be a continuous map where d denotes the distance on a symmetric space P of noncompact type and C is a compact space with a positive finite regular measure, μ, on it. Then the functional*

$$J(p) = \int_C d^2(p, f(c)) d\mu(c), p \in P$$

attains its minimum value at a unique point of P called the center of gravity of $f(C)$ with respect to μ.

Proof. Without loss of generality we may normalize the measure so that $\mu(C) = 1$. We first prove the result for Euclidean space $\mathbb{R}^n$. Then $f(c) = (f_1(c), \ldots, f_n(c))$, $p = (p_1, \ldots, p_n)$ and

$$J(p) = \int_C \sum_i (p_i - f_i(c))^2 d\mu(c) = \int_C (\sum_i p_i^2 - 2 \sum_i p_i f_i(c) + \sum_i f_i(c)^2) d\mu(c).$$

By our normalization, this is $|p|^2 - 2(a, p) + \sum_i b_i$, where for $i = 1, \ldots, n$, $a_i = \int_C f_i(c) d\mu(c)$ and $b_i = \int_C f_i(c)^2 d\mu(c)$. Completing the square we get

$$J(p) = |p|^2 - 2(a, p) + |a|^2 - |a|^2 + \sum_i b_i = |p - a|^2 + \sum_i b_i - |a|^2.$$

This is clearly minimized exactly at the point $p = a$.

Now let P be a symmetric space as above. We first show J is continuous on P. If $p_n \to p$ in P, then since d is continuous, $d^2(p_n, f(c)) \to d^2(p, f(c))$ for each $c \in C$. Hence for fixed $c \in C$, $d^2(p_n, f(c)) \leq$ a constant independent of n. Since the measure on C is finite this constant function is integrable and so by the dominated convergence theorem $J(p_n) \to J(p)$.

We will now find a compact set K in M and an $r > 0$ so that $J(p) > r^2$ on $M - K$ and $J(p_0) \leq r^2$ at some point $p_0 \in K$. Then the minimum value of J, if any, would have to be on K and there would be one since K is compact and J continuous. This would prove existence of a fixed point. To do this, choose p_0 arbitrarily in $P - f(C)$ (since $f(C)$ is compact and P is not, the complement of $f(C)$ is nonempty) and let $\inf_{c \in C} d(p_0, f(c)) = r$. Then r is finite and positive. Let $K = \{p \in P : d(p, f(C)) \leq r\}$. If B is a compact set in P, then since exp is a global diffeomorphism, $\log B$ is compact in $\mathfrak{p}$. The formula $d(\exp x, \exp y) \geq d_{\mathfrak{p}}(x, y)$ tells us that

$$\log\{p \in P : d(p, B) \leq r\} \subseteq \{x \in \mathfrak{p} : d_{\mathfrak{p}}(x, \log B) \leq r\}.$$

The right hand side is closed and bounded so since $\mathfrak{p}$ is a Euclidean space, this set is compact. Similarly, by elementary properties of the metric d, $\{p \in P : d(p, B) \leq r\}$ is closed in P and, since we have a diffeomorphism, the left side is closed and therefore is compact in $\mathfrak{p}$. But then exp of this set, i.e., K is compact in P. Evidently $p_0 \in K$ and integrating and making use of the normalization of μ tells us $J(p_0) \leq r^2$. If $p \in M - K$, then by compactness of $f(C)$ there is a $\delta > 0$ such that $d(p, f(c)) > r + \delta$ for all $c \in C$. Upon integration we get $J(p) > r^2$.

Turning to the uniqueness, let $\log \cdot f = f_{\mathfrak{p}}$. This is a continuous map $C \to \mathfrak{p}$ and we can construct the functional $J_{\mathfrak{p}}$ on $\mathfrak{p}$ defined by

$$J_{\mathfrak{p}}(x) = \int_C d_{\mathfrak{p}}^2(x, f_{\mathfrak{p}}(c))d\mu(c), x \in \mathfrak{p}.$$

For a point $p \in P$ and $x = \log p$, since $d_{\mathfrak{p}}(x, y) \leq d(p, q)$ for all $q = \exp y \in P$, we see that $d_{\mathfrak{p}}(x, f_{\mathfrak{p}}(c)) \leq d(p, f(c))$ for all c. Squaring and integrating tells us that

$$J_{\mathfrak{p}}(x) \leq J(p), p \in P.$$

Let p_0 be a point of P where a minimum value of J is attained and let $p = \log x \in P$ approach p_0. Then $J_{\mathfrak{p}}(\log p_0) \leq J_{\mathfrak{p}}(x)$. For if for some $x \to \log p_0$, $J_{\mathfrak{p}}(x) < J_{\mathfrak{p}}(\log p_0)$, which is, in turn, $\leq J(p_0)$. Then since J and $J_{\mathfrak{p}}$ involve integrating d and $d_{\mathfrak{p}}$, respectively, over a compacta and

$$\lim_{p,q \to p_0} d(p, q)/d_{\mathfrak{p}}(\log p, \log q) = 1,$$

there is some sufficiently nearby point y to x for which $J(\log y)$ is near enough to $J_{\mathfrak{p}}(y)$ so that $J_{\mathfrak{p}}(y) \leq J(\log y)$ and so is $< J(p_0)$, contradicting the minimality of p_0. Thus for all $p \to p_0$, $J_{\mathfrak{p}}(\log p_0) \leq J_{\mathfrak{p}}(\log p)$. This means $\log p_0$ is also a minimum value for $J_{\mathfrak{p}}$. Therefore if $\log p_1$ is another point of P where the minimum value of J is attained, then $\log p_1$ is also a minimum value for $J_{\mathfrak{p}}$. By uniqueness in the Euclidean case $\log p_0 = \log p_1$ and so $p_0 = p_1$. ∎

As usual, G is a self-adjoint essentially algebraic subgroup of $\mathrm{GL}(n, \mathbb{R})$, or $\mathrm{GL}(n, \mathbb{C})$ acting on P by $(g, p) \mapsto g^t pg$.

Corollary 6.2 *If C is a compact subgroup of G, then C has a simultaneous fixed point acting on P.*

Proof. Let $\mu = dc$ be normalized Haar measure on C, p_0 a point of P and $f : C \to P$ be the continuous function given by $f(c) = c \cdot p_0$. Then $J(p) = \int_C d^2(p, c \cdot p_0)dc$. Now for $c' \in C$, $J(c'p) = \int_C d^2(c'p, c \cdot p_0)dc$. Since C acts by isometries this is $\int_C d^2(p, (c')^{-1}c \cdot p_0)dc$. By left invariance of dc we get $\int_C d^2(p, c \cdot p_0)dc$. Thus $J(p_0) = J(c \cdot p_0)$ for all $c \in C$ and $p_0 \in P$. But by the fixed point theorem, J has a unique minimum value at some $p_0 \in P$. This means $c(p_0) = p_0$ for all $c \in C$. ∎

We now prove the conjugacy theorem for maximal compact subgroups of G. The proof in [5] is similar to the one given here, but rather than involve differential geometry itself, it uses a convexity argument and a function which mimics the metric.

Theorem 6.3 *Let G be a self-adjoint essentially algebraic subgroup of* $\mathrm{GL}(n, \mathbb{R})$, *or* $\mathrm{GL}(n, \mathbb{C})$. *Then all maximal compact subgroups of G are conjugate. Any compact subgroup of G is contained in a maximal one.*

Proof. Let C be a compact subgroup of G. By Corollary 6.2 there is a point $p_0 \in P$ fixed under the action. Thus $C \subseteq \mathrm{Stab}_G(p_0)$. But this action is transitive so $\mathrm{Stab}_G(p_0) = gKg^{-1}$ for some $g \in G$. Since K is a maximal compact subgroup by Theorem 3.6, so is the conjugate gKg^{-1}. This proves the second statement. If C were itself maximal then $C = gKg^{-1}$. ∎

7 The Rank and Two-Point Homogeneous Spaces

Let $\mathfrak{g}$ be the real (or complex) linear Lie algebra of G, as above, and $\mathfrak{g} = \mathfrak{k} \oplus \mathfrak{p}$ be a Cartan decomposition. By abuse of notation we shall call a subalgebra of $\mathfrak{g}$ contained in $\mathfrak{p}$ a subalgebra of $\mathfrak{p}$. When abelian, such subalgebras will play an important role in what follows. By finite dimensionality, maximal abelian subalgebras of $\mathfrak{p}$ clearly exist. In fact, any abelian subset of $\mathfrak{p}$ is contained in a maximal abelian subalgebra of $\mathfrak{p}$.

Consider the adjoint representation of K on $\mathfrak{g}$. Then the subspace $\mathfrak{p}$ is invariant under this action. Since $\mathrm{Ad}_k(\mathfrak{p}) \subseteq \mathfrak{g}$, to see this we need only check that $\mathrm{Ad}_k(\mathfrak{p})$ is symmetric (Hermitian). We shall always deal with the symmetric case, unless the Hermitian one is harder. So for $p \in \mathfrak{p}$ and $k \in K$ we have $\mathrm{Ad}_k(p) = kpk^{-1} = kpk^t$. Hence the transpose is $(kpk^t)^t = kpk^t = \mathrm{Ad}_k(p)$.

Theorem 7.1 *In $\mathfrak{g}$ any two maximal abelian subalgebras $\mathfrak{a}$ and $\mathfrak{a}'$ of $\mathfrak{p}$ are conjugate by some element of K. In particular, their common dimension is an invariant of $\mathfrak{g}$ called $r = \mathrm{rank}(\mathfrak{g})$.*

This theorem was origianally proved by E. Cartan; however, here we will use the following argument which is essentially due to G.A. Hunt [6].

Proof. Let $(,)$ be the Killing form on $\mathfrak{g}$. This is positive definite on $\mathfrak{p}$ and negative definite on $\mathfrak{k}$. Since K is compact and acts on $\mathfrak{g}$, by averaging with respect to Haar measure on K we can, in addition, assume this form to be K-invariant. That is, each Ad_k preserves the form. Let $a \in \mathfrak{a}$ and $a' \in \mathfrak{a}'$ and consider the smooth numerical function on K given by $f(k) = (\mathrm{Ad}_k\, a, a')$. By compactness of K, this continuous function has a minimum value at k_0 and by calculus, at this point the derivative is zero. Thus for each $x \in \mathfrak{k}$, $\frac{d}{dt}(\mathrm{Ad}_{\exp tx\cdot k_0}\, a, a')_{|t=0} = 0$. But $(\mathrm{Ad}_{\exp tx\cdot k_0}\, a, a') = (\mathrm{Ad}_{\exp tx} \mathrm{Ad}_{k_0}\, a, a') = (\mathrm{Exp}\, t\, \mathrm{ad}(x) \mathrm{Ad}_{k_0}\, a, a')$. Hence differentiating with respect to t at $t = 0$ gives $(\mathrm{ad}(x) \mathrm{Ad}_{k_0}\, a, a') = 0$ for all $x \in \mathfrak{k}$. A calculation similar to the one just given shows that the K-invariance of the form on $\mathfrak{k}$ has an infinitesimal version, $([x, y], z) + (y, [x, z]) = 0$, valid for all $x \in \mathfrak{k}$ and $y, z \in \mathfrak{p}$. Hence, also for all $x \in \mathfrak{k}$, we get $(x, [\mathrm{Ad}_{k_0}\, a, a']) = 0$. Now $\mathrm{Ad}_{k_0}\, a$ and $a' \in \mathfrak{p}$ and $[\mathfrak{p}, \mathfrak{p}] \subseteq \mathfrak{k}$. Hence $[\mathrm{Ad}_{k_0}\, a, a'] \in \mathfrak{k}$ and because $(x, [\mathrm{Ad}_{k_0}\, a, a']) = 0$ for all $x \in \mathfrak{k}$ and $(.)$ is nondegenerate on $\mathfrak{k}$, it follows that $[\mathrm{Ad}_{k_0}\, a, a'] = 0$. Finally, since a and a' are arbitrary, $[\mathrm{Ad}_{k_0}(\mathfrak{a}), \mathfrak{a}'] = 0$. Now hold $a \in \mathfrak{a}$ fixed. Because $[\mathrm{Ad}_{k_0}\, a, \mathfrak{a}'] = 0$ we see by maximality of $\mathfrak{a}'$ that $\mathrm{Ad}_{k_0}\, a \in \mathfrak{a}'$ and since a is arbitrary $\mathrm{Ad}_{k_0}\, \mathfrak{a} \subseteq \mathfrak{a}'$. Thus $\mathfrak{a} \subseteq \mathrm{Ad}_{k_0^{-1}}(\mathfrak{a}')$. The latter is an abelian subalgebra of $\mathfrak{p}$ and by maximality of $\mathfrak{a}$ they coincide. Thus $\mathrm{Ad}_{k_0}(\mathfrak{a}) = \mathfrak{a}'$. ∎

It might be helpful to mention the significance of this theorem in the most elementary situation, namely, when $G = \mathrm{GL}(n, \mathbb{R})$, or $\mathrm{GL}(n, \mathbb{C})$. As usual, we restrict our remarks to the real case. Here $\mathfrak{p}$ is the set of all symmetric matrices of order n. Let $\mathfrak{d}$ denote the diagonal matrices. These evidently form an abelian subalgebra of $\mathfrak{p}$. Now $\mathfrak{d}$ is actually maximal abelian. To see this, suppose there were a possibly larger abelian subalgebra $\mathfrak{a}$. Each element of $\mathfrak{a}$ is diagonalizable being symmetric. Since all these elements commute they are simultaneously diagonalizable. This means, in effect, that $\mathfrak{a} = \mathfrak{d}$. Thus $\mathfrak{d}$ is a maximal abelian subalgebra of $\mathfrak{p}$. Now since any commuting family of symmetric matrices can be imbedded in a maximal abelian subalgebra of $\mathfrak{p}$, Theorem 7.1 tells us that this commuting family can be simultaneously diagonalized by an orthonormal change of coordianates. Similarly, over $\mathbb{C}$ it says any commuting family of Hermitian matrices is simultaneously conjugate by a unitary matrix to the diagonal matrices. This is exactly the content of the theorem in these two cases. Thus Theorem 7.1 is a generalization of the classic result on simultaneous diagonalization of commuting families of quadratic or Hermitian forms.

We also note that the statement of Theorem 7.1 without the stipulation that the subalgebras are in $\mathfrak{p}$ is false. That is, in general, maximal abelian subalgebras of $\mathfrak{g}$ are not conjugate. For example, in $\mathfrak{g} = \mathfrak{sl}(2, \mathbb{R})$, the diagonal elements, the skew symmetric elements and the unitriangular elements are each maximal abelian subalgebras of $\mathfrak{g}$, but no two of them are conjugate (by an element of K or anything else). We leave the verification of these facts to the reader.

Corollary 7.2 *In $\mathfrak{g}$ let $\mathfrak{a}$ be a maximal abelian subalgebra of $\mathfrak{p}$. Then the conjugates of $\mathfrak{a}$ by K fill out $\mathfrak{p}$, that is, $\cup_{k \in K} \operatorname{Ad}_k(\mathfrak{a}) = \mathfrak{p}$. Of course, exponentiating and taking into account that* exp *commutes with conjugation, this translates on the group level to $P = \cup_{k \in K} kAk^{-1}$, where A is the abelian analytic subgroup of G with Lie algebra $\mathfrak{a}$.*

Proof. Let $p \in \mathfrak{p}$ and choose a maximal abelian subalgebra $\mathfrak{a}'$ containing it. By our theorem there is some $k \in K$ conjugating $\mathfrak{a}'$ to $\mathfrak{a}$. In particular, $\operatorname{Ad}_k(p) \in \mathfrak{a}$ for some $k \in K$ and so $p \in \operatorname{Ad}_{k^{-1}}(\mathfrak{a})$. ∎

Our next corollary, also called the Cartan decomposition, follows from this last fact together with the usual Cartan decomposition, Theorem 3.6.

Corollary 7.3 *Under the same hypothesis $G = KAK$.*

An important use of this form of the Cartan decomposition is that it reduces the study of the asymptotics at ∞ on G to A. That is, suppose g_i is a sequence in G tending to ∞. Now $g_i = k_i a_i l_i$, where k_i and $l_i \in K$ and $a_i \in A$. Since both k_i and l_i have convergent subsequences, again denoted by k_i and l_i, which converge to k and l, respectively, the sequence a_i must also tend to ∞. Thus in certain situations we can assume the original sequence started out in A.

Proof. $G = KP \subseteq KKAK = KAK \subseteq G$ ∎

We now make explicit the notions of a *homogeneous space* and *two-fold transitivity* from differential geometry mentioned earlier. If X is a connected Riemannian manifold, we shall say X is a homogeneous space if the isometry group $\operatorname{Isom}(X)$ acts transitively on X. Now even when the action may not be transitive it is a theorem of Myers and Steenrod (see [4]) that $\operatorname{Isom}(X)$ is a Lie group and the stabilizer K_p of any point p is a compact subgroup. In the case of a transitive action it follows from general facts about actions that X is equivariantly equivalent as a Riemannian manifold to $\operatorname{Isom}(X)/K_p$ with the quotient structure. Of course, if some subgroup of the isometry group acted transitively then these same conclusions could be drawn replacing the isometry group by the subgroup. Clearly, by its very construction, every symmetric space of noncompact type is a homogeneous space.

Now suppose in our symmetric space P we are given points p and q and p' and q' of P with $d(p,q) = d(p',q')$. We shall say a subgroup of the isometry group acts two-fold transitively if there is always an isometry g in the subgroup taking p to p' and q to q' for any choices of such points. When this occurs we shall say P is a two-point homogeneous space. Clearly, every two-point homogeneous space is a homogeneous space. As we shall see the converse is not true and learn which of our symmetric spaces is actually a two-point homogeneous space. Before doing so, we make a simple observation which follows immediately from transitivity.

Proposition 7.4 *Let G be as above and K be a maximal compact subgroup. Then $G/K = P$ is a two-point homogeneous space if and only if K acts transitively on the unit geodesic sphere U of P.*

For example, when $G = \mathrm{SO}_0(n,1)$ and $K = \mathrm{SO}(n)$, then $G/K = H^n$, hyperbolic n-space. Here K acts transitively on U. Hence $\mathrm{SO}_0(n,1)$ acts two fold-transitively on H^n. As we shall see in Theorem 7.5, this fact is a special case of a more general result. We also remark that this definition can be given for any connected Riemannian manifold and indeed such a manifold is of necessity a symmetric space (see [4]).

Our last result tells us the significance of the rank in this connection. Before proving it we observe that for all semisimple or reductive groups under consideration $\dim \mathfrak{p} \geq 2$. The lowest dimension arising is the case of the upper half plane introduced at the very beginning of this article. Indeed, suppose $\dim \mathfrak{p} = 1$. Then since $\mathfrak{p}$ is abelian and exp is a global diffeomorphism $\mathfrak{p} \to P$, it follows easily from $\exp x + y = \exp x \exp y$, where $x, y \in \mathfrak{p}$, that P is a connected one-dimensional abelian Lie group. Now since K acts on P by conjugation and in this case these form a connected group of automorphisms of P we see that this action is trivial because $\mathrm{Aut}(P)_0 = (1)$. Thus K centralizes P and we have a direct product of groups. Such a group is not semisimple. It is clearly also not $\mathrm{GL}(n, \mathbb{R})$ or $\mathrm{GL}(n, \mathbb{C})$ for $n \geq 2$.

We now characterize two-point homogeneous symmetric spaces.

Theorem 7.5 *Let G be as above, $\mathfrak{g}$ be its Lie algebra and K be a maximal compact subgroup. Then G/K is a two-point homogeneous space if and only if* $\mathrm{rank}(\mathfrak{g}) = 1$.

Proof. We first assume $\mathrm{rank}(\mathfrak{g}) = 1$. By Proposition 7.4, to see that G/K is a two-point homogeneous space, it is sufficient to show $Int(K)$ acts transitively on geodesic spheres of P. Of course, we know $\mathrm{Ad}(K)$ acts linearly and isometrically on $\mathfrak{p}$. Now by Corollary 7.2 $\cup_{k \in K} \mathrm{Ad}_k(\mathfrak{a}) = \mathfrak{p}$. Hence each point $p \in U$ is a conjugate by something in K of a point on the unit sphere of $\mathfrak{a}$. Since the dimension of this sphere is zero, it consists of two points, $\pm a_0$. Hence $U = \mathrm{Ad}(K)(a_0) \cup \mathrm{Ad}(K)(-a_0)$. In any case, U is a union of a finite number of orbits all of which are compact and therefore closed since K itself is compact. Since these are closed, so is the union of all but one of them. Hence U is the disjoint union of two nonempty closed sets. This is impossible since U is connected because $\dim \mathfrak{p} \geq 2$. Thus there is only one orbit and therefore K acts transitively on U. ∎

Before proving the converse, the following generic example will be instructive. Let $G = \mathrm{SL}(n, \mathbb{R})$, $n \geq 2$. We shall see $\mathrm{SL}(n, \mathbb{R}) / \mathrm{SO}(n)$ is a two-point homogeneous space if and only if $n = 2$. This suggests that unless the rank $= 1$, one can never have a two-point homogeneous symmetric space.

To see this, observe that since $G/K = P$ is the set of positive definite $n \times n$ symmetric matrices of det 1, it follows that $\dim P = \frac{n(n+1)}{2} - 1$. Also $\dim K = \frac{n(n-1)}{2}$. Hence if U denotes the geodesic unit sphere in P, its dimension is $\frac{n(n+1)}{2} - 2$. Let K act on P and U by $(k,p) \mapsto kpk^{-1} = kpk^t$. For $p \in U$ the dimension of $\mathcal{O}_K(p)$, the K-orbit of p, is

$$\dim \mathcal{O}_K(p) = \frac{n(n-1)}{2} - \frac{(n-1)(n-2)}{2} = n - 1.$$

Now if K were to act transitively on U, then $\dim \mathcal{O}_K(p) = \dim U$. That is, $n - 1 = \frac{n(n+1)}{2} - 2$. Alternatively, $(n-2)(n+1) = 0$. Since $n \geq 2$, this holds if and only if $n = 2$. We conclude by proving the converse.

Proof. Suppose (P, G) is a two-point homogeneous space and hence K acts transitively (by conjugation) on the unit geodesic sphere U in $\mathfrak{p}$. Then $U = \mathcal{O}_K(a_0)$, where $a_0 \in \mathfrak{p}$ and $\| a_0 \| = 1$. Since a_0 is conjugate to something in $\mathfrak{a}$, we may assume $a_0 \in \mathfrak{a}$. In particular, everything in $U \cap \mathfrak{a}$ is K-conjugate to everything else. Because these matrices commute, they can be simultaneously diagonalized by some u_0 (which may not be in K). By replacing these a's by their u_0 conjugates we may assume they are all diagonal. Being conjugate under K these matrices have the same spectrum S. Since S is finite and K is connected, K can not permute this finite set. Thus the action of K leaves each of these matrices fixed. But K acts transitively on $U \cap \mathfrak{a}$ so $U \cap \mathfrak{a}$ must be a point. Hence it has dim 0 and $\dim \mathfrak{a} = 1$. ∎

Received: June 2003. Revised: August 2003.

References

[1] J. CHEEGER AND D. EBIN, *Comparison Theorems in Riemannian Geometry*, North Holland, Amsterdam, 1975.

[2] C. CHEVALLEY, *The Theory of Lie Groups*, Princeton University Press, Princeton, 1946.

[3] E. CARTAN, *Groupes simples clos et ouverts et geometrie Riemannienne*, Journal de Math. Pures et Appliqués **8**, (1929) 1-33.

[4] S. HELGASON, *Differential Geometry, Lie Groups and Symmetric Spaces*, Academic Press, New York, 1978.

[5] G. P. HOCHSCHILD, *The Structure of Lie Groups*, Holden Day, San Francisco, 1965.

[6] G. A. HUNT, *A Theorem of E. Cartan*, Proc. Amer. Math. Soc., **7** (1956) 307-308.

[7] J. MILNOR, *Morse Theory*, Annals of Math. Studies **51**, Princeton University Press, Princeton, 1963.

[8] M. MOSKOWITZ, *A Course in Complex Analysis in One Variable*, World Scientific Publishing Co., Singapore, 2002.

[9] G. D. MOSTOW, *Strong Rigidity of Locally Symmetric Spaces*, Annals of Math. Studies **78**, Princeton University Press, Princeton, 1973.

[10] G. D. MOSTOW, *Self Adjoint Groups*, Annals of Math., **62** (1955), 44-55.

[11] V. S. VARADARAJAN, *Lie Groups, Lie Algebras and their Representations*, Prentice-Hall Series in Modern Analysis, Englewood Cliffs New Jersey, 1974.