Astérisque

PETER BUSER
HERMANN KARCHER

Gromov's almost flat manifolds

Astérisque, tome 81 (1981)
<http://www.numdam.org/item?id=AST_1981__81__1_0>

PREFACE

This exposê gives a detailed proof of M. Gromov's pinching theorem for almost flat manifolds. We have two reasons for spending so much effort to rewrite a proof. One is that Gromov's original publication [1] assumes that the reader is very familiar with several rather different fields and has no difficulties in completing rather unconventional arguments - we hope our presentation requires less background. Secondly, we consider the full proof an ideal introduction to qualitative Riemannian geometry since the characteristic interplay between local curvature controlled analysis and global geometric constructions occurs at several different levels. These considerations persuaded us to write the following chapters in a selfcontained and hopefully accessible way:

§ 6 treats curvature controlled constructions in Riemannian geometry; § 7 develops metric properties of Lie groups; § 8 explains nonlinear averaging methods; § 2 contains commutator estimates in the fundamental group which are the heart of the Gromov-Margulis discrete group technique and 5.1 is a new form of Malcev's treatment of nilpotent groups.

The proof proper is given in § 3 - § 5, while § 1 contains earlier results and examples pertaining to the almost flat manifolds theorem and a guide to its proof. The statement of the theorem is in 1.5.

We are grateful for discussions with M. Gromov at the I.H.E.S. and the Arbeitstagung in 1977 on the present § 3 after which the idea of this manuscript was born, and at the I.H.E.S. in 1980 which helped to get § 5.1 in its final form. After (countably) many discussions between the two of us we hope that our readers profit from the synthesis of two different styles and temperaments.

Finally our thanks go to Mrs. M. Barrôn for carefully typing - and retyping - the manuscript and to Arthur L. Besse who suggested contacting Astêrisque for publication.

Table of contents

1. The theorem, earlier results, examples

1.1 Earlier results which concluded global properties from curvature assumptions.

(i) The Gauß-Bonnet formula for the Euler characteristic, $2\pi \cdot \chi(M) = \int_M K dO$, together with the topological classification of surfaces shows that S^2 and $P^2(\mathbb{R})$ are the only compact surfaces admitting metrics of positive curvature. - Such proofs by integral formulas however play no role in what follows.

(ii) The Hadamard-Cartan theorem states that a complete Riemannian manifold M of nonpositive curvature is covered by $\mathbb{R}^n$. The assumptions imply
a) that the Riemannian exponential map $\exp_p : T_pM \to M$ has maximal rank and
b) that each element of the fundamental group $\pi_1(M,p)$ contains exactly one geodesic loop at p so that, in fact, $\exp_p$ is a covering map. - § 2 starts with an extension of these ideas.

(iii) Gromoll-Meyer proved [9] that a complete noncompact manifold M^n of positive curvature is diffeomorphic to $\mathbb{R}^n$ by exhibiting an exhaustion of M^n with convex balls. Cheeger-Gromoll [6] extended this to nonnegative curvature, in which case M^n is diffeomorphic to the normal bundle of a compact totally geodesic submanifold of M . - The - global - convexity arguments in the proof work because the behaviour of geodesics in this case is under sufficiently precise curvature control.

(iv) The topological sphere theorem [3], [20] states: A simply connected complete Riemannian manifold M^n with sectional curvature bounds $\frac{1}{4} < K \leq 1$ is homeomorphic to S^n ; the result is sharp since $P^n(\mathbb{C})$ carries a metric with $\frac{1}{4} \leq K \leq 1$. One proves first that each $p \in M^n$ has a unique antipode $q \in M^n$ at maximal distance from p ; then the "hemispheres" of these poles (the sets of points closer to one pole than to the other) are diffeomorphic to balls since geodesic segments are at least up to length π distance minimizing. - The details depend, as at several though more complicated instances in the proof of Gromov's theorem on a curvature controlled comparison between the situation described on M^n and the corresponding situation on the model spaces, S^n .

(v) The <u>differentiable sphere pinching theorem</u> [13], [17] states under curvature assumptions $0.7 \leq K \leq 1$ that a complete M^n is diffeomorphic to a space of constant curvature. In addition to (iv) one has to find an isometric action of $\pi_1(M)$ on S^n, construct the (differentiable) map $F : \tilde{M}^n \to S^n$ equivariant for the $\pi_1(M)$ actions on $\tilde{M}$ resp. S^n and prove maximal rank of dF with more refined local curvature control. - Corresponding ideas are behind §§ 5.1, 5.2, 5.4.

(vi) <u>Symmetric spaces</u>. In principle similar results hold if the model space S^n is replaced by other symmetric spaces [23], [24], however they are more complicated to formulate and are proved via partial differential equations - a method which does not occur in the sequel.

1.2 Rather different in spirit but also important for the understanding of Gromov's theorem is the following version of

<u>Bieberbach's theorem</u> [4]: Let M^n be a compact flat Riemannian manifold, $\pi_1(M,p)$ its fundamental group acting on $T_pM = \mathbb{R}^n$ by rigid motions (deck-transformations), and Γ the set of all translations in $\pi_1(M,p)$, then: Γ is a free abelian normal subgroup of rank n; the factor group $G = \Gamma\backslash\pi_1(M)$ has finite order and is obtained as the group of rotational parts of the π_1-action on T_pM; $\Gamma\backslash T_pM$ is a torus which covers M with deckgroup G.

The main step is to show that the rotational part $A \in O(n)$ of each motion $x \to Ax + a$ in the deckgroup $\pi_1(M)$ has all its main rotational angles rational. As a consequence of the theorem A is always either the identity or has a maximal rotational angle $\geq \frac{1}{2}$. It is the discovery of a direct geometric proof of this fact which leads to Gromov's theorem.

1.3 <u>Definition</u>. A compact Riemannian manifold is called <u>ε-flat</u> if the curvature is bounded in terms of the diameter as follows:

$$|K| \leq \varepsilon \cdot d(M)^{-2} .$$

By almost flat we usually mean that the manifold carries ε-flat metrics for arbitrary $\varepsilon > 0$. If one multiplies an ε-flat metric by a constant it remains ε-flat.

1.4 Examples of almost flat manifolds. It is essential to realize that almost flat manifolds which do not carry flat metrics exist and occur rather naturally.

(i) Each nilmanifold (= compact quotient of a nilpotent Lie group) is almost flat (7.7.2).

(ii) An illustrative special case of (i) is obtained if on the nilpotent Lie algebra

$$\mathfrak{g} = \left\{ \begin{pmatrix} 0 & \ddots & a_{ij} \\ & \ddots & \\ 0 & & \ddots\, 0 \end{pmatrix} = A \; ; \; a_{ij} \in \mathbb{R} \; , \; 1 \leq i < j \leq n \right\}$$

the following family of scalar products is introduced

$$\|A\|_q^2 := \sum_{i<j} a_{ij}^2 \, q^{2(j-i)} \; .$$

They give left invariant Riemannian metrics on the corresponding nilpotent Lie group N of upper triangular matrices. From
$\|[A,B]\|_q \leq 2(n-2)\,\|A\|_q \cdot \|B\|_q$ and 7.7.1 one derives the following q-independent (!) bound for the curvature tensor R_q of these metrics:

$$\|R_q(A,B)C\|_q \leq 24(n-2)^2 \, \|A\|_q^2 \cdot \|B\|_q^2 \cdot \|C\|_q^2 \; .$$

Therefore each compact quotient $M^m = \Gamma \backslash N$ is almost flat, since obviously its diameter can be made arbitrarily small by choosing q sufficiently small. On the other hand M cannot carry a flat metric since by Bieberbach's theorem (1.2) the fundamental group $\pi_1(M) = \Gamma$ would then contain an abelian subgroup Z^m of finite index in Γ. Hence we would have Z^m uniform in N, which via the Campbell-Hausdorff-formula implies, that N itself is abelian (5.1.6), a contradiction.

The integer subgroup of N gives an example with the compact fundamental domain $\{A \in N \; ; \; 0 \leq a_{ij} \leq 1\}$, a hypercube. In the 3-dimensional case one can see the deviation from the flat situation in a simple picture:

$$\begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} , \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix} \in \Gamma \text{ act on } \begin{pmatrix} 1 & z & x \\ 0 & 1 & y \\ 0 & 0 & 1 \end{pmatrix} \in N \text{ by}$$

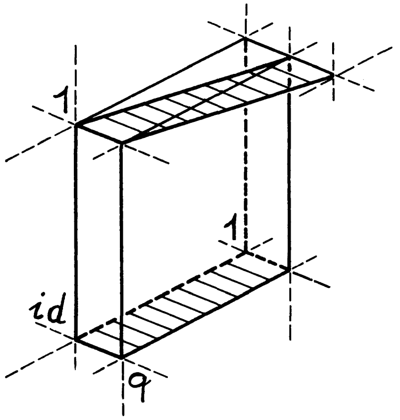
left translations as in the torus case, i.e. by translation in the direction of the x-axis, resp. the y-axis, but

$$\begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad \text{acts as the affine map} \quad \begin{matrix} x \to x + y \\ y \to y \\ z \to z + 1 \end{matrix} \quad .$$

The above left invariant metrics are given by

$$|v|_q^2 = q^2(q^2(\xi - z\eta)^2 + \eta^2 + \zeta^2) \quad \text{if} \quad v = \begin{pmatrix} \xi \\ \eta \\ \zeta \end{pmatrix} \in T_{\begin{pmatrix} x \\ y \\ z \end{pmatrix}} N \ ;$$

$$||ad|| = 1 \ , \text{ independent of } q \ .$$



Only the "vertical" identification differs from the torus case.

Finally, the mulpiplication in N can be called "almost translational" since it deviates from translation only by quadratic errors:

$$||(1 + A) \cdot (1 + B) - (1 + A + B)||_q \leq (n-2) ||A||_q \cdot ||B||_q \ .$$

This notion plays a central role for the fundamental group of almost flat manifolds, see 3.5.

(iii) <u>Parabolic ends</u>. Let H be a simply connected complete Riemannian manifold with curvature bounds $-b^2 \leq K \leq -a^2 < 0$. Each geodesic ray $c : [0,\infty) \to M$ determines its family $\mathcal{H}_t$ of horospheres. If one identifies these horospheres by projecting them onto $\mathcal{H}_0$ with the perpendicular family of geodesics asymptotic to c , then one gets a family of exponentially decreasing metrics on $\mathcal{H}_0$ [15] :

$$g_o \cdot e^{-2bt} \leq g_t \leq g_o \cdot e^{-2at} .$$

On the other hand, the Gauß equations and the nonexistence of focal points for horospheres show that their intrinsic curvatures are bounded independent of t, in fact

$$-(b^2 - a^2) \leq K(\mathcal{H}_t) \leq 2b(b - a) .$$

One therefore gets an almost flat manifold $\Pi\backslash\mathcal{H}_o$ from any parabolic group Π of isometries of H acting on $\mathcal{H}_o$ (fixed point free) with compact quotient - e.g. if Π corresponds to an end of a finite volume quotient of H [8]. A family of ε-flat metrics $(\varepsilon \to 0)$ is naturally given by g_t or equivalently obtained by moving $\mathcal{H}_o$ with the Π-equivariant geodesic flow to $\mathcal{H}_t$ (cf. also 1.5.2).

(iv) The <u>binary icosahedral subgroup</u> of S^3 has as fundamental domain a spherical dodecahedron which is contained in a spherical ball of radius $< \frac{\pi}{8}$. Definition 1.3 works equally well with the diameter replaced by the maximal distance to a distinguished point. The quotient of S^3 by that subgroup therefore is $(\frac{\pi}{8})^2 \approx 0.155$-flat - but still of constant curvature 1 ! An even "flatter" space (in the sense of 1.3) of nonnegative curvature can be obtained by dividing $O(4)$ by the symmetry group of the above dodecahedral tessellation of S^3.

The examples 1.4.(i) are up to finite quotients the only almost flat manifolds:

1.5 <u>Main Theorem</u> (Almost flat manifolds) [1]
Let M^n be a compact Riemannian manifold with sectional curvature bounds

$$|K| \leq \varepsilon_n d^{-2}(M) , \varepsilon_n = \exp(-\exp(\exp n^2)) ,$$

then M is covered by a nilmanifold. More precisely:

(i) The fundamental group $\pi_1(M)$ contains a torsion free nilpotent normal subgroup Γ of rank n (4.6.5).

(ii) The quotient $G = \Gamma\backslash\pi_1(M)$ has order $\leq 2\cdot(6\pi)^{\frac{1}{2}n(n-1)}$ and is isomorphic to a subgroup of $O(n)$ (3.6.4).

(iii) The finite covering of M with fundamental group Γ and deckgroup G is diffeomorphic to a nilmanifold $\Gamma\backslash N$ (5.2).

(iv) The simply connected nilpotent Lie group N is uniquely determined by $\pi_1(M)$, (5.1.7).

Remark 1. It is not known whether M is - as in the flat case - diffeomorphic to the quotient of N by a uniform discrete group of isometries for a suitable left invariant metric on N . Therefore 1.5 does not imply the Bieberbach theorem; but a little more than stated in 1.5 is true in that direction:

In the Bieberbach case $G = \Gamma\backslash\pi_1(M,p)$ is naturally isomorphic to the holonomy group of M at p , and Γ , the set of translations in $\pi_1(M,p)$, can also be described as the set of loops at p with rotational part $\leq \frac{1}{2}$ (1.2). In the almost flat case the group Γ is generated by those "short" loops (i.e. with lengths $\leq 4(6\pi)^{\frac{1}{2}n(n-1)}d(M)$) which have a rotational part (2.3) ≤ 0.48 . Again, these rotational parts are in fact much smaller than 0.48 (3.5); they have an upper bound proportional to the length of the loop so that Γ is almost translational (see example 1.4 (ii)). Moreover, if one chooses shortest loops at p in the equivalence classes of $\pi_1(M,p)$ mod Γ then their holonomy rotations are, after a small correction, a subgroup of $O(n)$ isomorphic to $\Gamma\backslash\pi_1(M,p)$ (3.6.4).

Remark 2. The number ε_n in 1.5 reflects for larger n approximately what the present proof can yield; much better constants can be obtained with the same method for small n (= 3,4).

The twodimensional case of 1.5 follows from the classification of surfaces: Apply 6.4.1 to the Gauß-Bonnet integral to obtain (with curvature normalized to $|K| \leq 1$)

$$|\chi(M)| \leq \frac{1}{2\pi} \int_M |K| dO \leq \int_0^{d(M)} \sinh r\, dr = \cosh d(M) - 1 ;$$

therefore, if $|K|d(M)^2 \leq \operatorname{arccosh}(1+a)$ then $|\chi| \leq a$. In particular

$$|K|d(M)^2 \leq 1.73 \quad \text{implies} \quad \chi = 0 .$$

On the other hand the projective plane is $\pi^2/4 \approx 2.47$ -flat.

The following are corollaries from 1.5 and from the commutator estimate 3.5 for Γ . The proof is given in 5.5.

1.5.1 Corollary (Injectivity radius and commutativity)
Let M^n be ε-flat, $\varepsilon \leq \varepsilon_n$. If the injectivity radius of M is $> 2^{-n^3}(\frac{\varepsilon}{\varepsilon n})^{1/2} d(M)$, then $\Gamma \subset \pi_1(M)$ is abelian and M is covered by a torus.

1.5.2 Corollary (parabolic ends) [11]
Let H with curvature bounds $-b^2 \leq K \leq -a^2 < 0$ and the parabolic group Π acting with compact quotient on the horosphere $\mathcal{H}_o$ be given from 1.4 (iii). Then Π contains a nilpotent normal subgroup Γ of finite index, and the degree of nilpotency of Γ is $\leq \frac{b}{a}$. In particular Γ is abelian if $-1 \leq K \leq -a^2 < \frac{-1}{4}$.

Note that the last statement is sharp: If $H = G/K$ is a symmetric space of rank one, then H has the precise curvature bounds $-1 \leq K \leq -\frac{1}{4}$, and the Iwasawa decomposition $G = KAN$ can be taken such that the nilpotent part N is the restriction of the isometry group to $\mathcal{H}_o$. Now N is 2-nilpotent and so is any uniform discrete subgroup Γ of N (5.1.6).

1.6 Comments about the proof.
In pinching theorems prior to Gromov suitable curvature assumptions allowed to compare the manifold in question to a given model space (1.1). In the almost flat case the nilmanifold $\Gamma\backslash N$ has to be constructed in the proof; moreover curvature inequalities $|K| \leq \varepsilon d(M)^{-2}$ are at first so weak information that with earlier methods nothing could have been said about the universal covering. Gromov achieves his goal with so widely different arguments that it may be helpful to write a guide through the proof; it seems unavoidable that some of the following remarks become clear only after parts of the proof have been read.

§ 2. To each loop at $p \in M$ the affine holonomy associates a rigid motion of T_pM (2.3). Geodesic loops shorter than the maximal rank radius of $\exp_p$ can be multiplied (2.2.3) in a way which is almost compatible with the composition of the corresponding holonomy motions (2.3.1). The error is controlled by curvature and by loop length.

§ 3. This chapter contains what Gromov calls "imitation of the proof of Bieberbach's theorem". Consider the following set of geodesic loops at p, $\Gamma_\rho := \{\alpha; |t(\alpha)| \leq \rho, \|r(\alpha)\| \leq 0.48\}$, where $t(\alpha)$ and $r(\alpha)$ are translational and rotational part of the holonomy motion and $\|r(\alpha)\|$ denotes the distance of $r(\alpha)$ from $id \in O(n)$. The achievement of § 3 is:

Under strong curvature assumptions there exists some $\rho \gg d(M)$ (also a priori bounded above) such that loops $\alpha \in \Gamma_\rho$ satisfy $\|r(\alpha)\| \leq \frac{\Theta}{\rho} |t(\alpha)|$, Θ an adjustable parameter (3.4.2, 3.5).

First, ρ is selected such that the loops $\alpha \in \Gamma_\rho$ with very small rotational part ($\ll \Theta$) have translational parts which are relatively densely distributed in the ball of radius ρ in T_pM (3.2).

This set of "almost translations" is used to show that those loops in Γ_ρ with trivial d-fold iterated commutators in fact have a rotational part $\leq \Theta$ (3.3).

A certain set of generators for Γ_ρ does have trivial $d(n)$-fold iterated commutators (3.1); $d(n)$ is an a priori bound crucial for the proof.

An induction over wordlength (3.4) and a self-improvement (3.5) complete the argument.

Knowing this much about Γ_ρ is still far from having the finite index torsion free nilpotent subgroup Γ of $\pi_1(M,p)$; but since the loops $\leq \rho$ form a group of equivalence classes mod Γ_ρ (3.6.4) - which is shown in 4.6.5 to be the factor group $G = \Gamma\backslash\pi_1(M,p)$ - one can see the properties of G already.

Note that the enormous curvature assumptions arise since the definitions and estimates of § 2 are repeatedly used for loops of lengths up to ρ.

§ 4. With the results of § 3 one should picture Γ_ρ as consisting of the elements $\leq \rho$ of a slightly deformed lattice in T_pM. After tedious error controls, an induction produces exactly n generators $\gamma_1,\ldots,\gamma_n$ for Γ_ρ with the additional property: All loops in Γ_ρ can uniquely be written as words $\gamma_1^{l_1}\ldots\gamma_n^{l_n}$ $(l_i \in \mathbb{Z})$ and $[\gamma_i,\gamma_j] \in \langle\gamma_1,\ldots,\gamma_{i-1}\rangle$ (4.5). With some suitable R $(d(M) \ll R \ll \rho)$ the set of geodesic loops $\Gamma_R \subset \Gamma_\rho$ generates an (abstract) nilpotent torsionfree group Γ which contains Γ_R as subset (4.6.2). In a final step one shows that the discovered algebraic structure really yields information about the fundamental group: <u>All</u> loops of length $\leq R$ represent different homotopy classes and Γ embeds as a finite index subgroup into $\pi_1(M,p)$; the proof uses that the translational parts of the loops in Γ_R are fairly dense in the ball $B_R \subset T_pM$ (3.6.3) and that all relations in $\pi_1(M,p)$ are generated from "short" relations (i.e. among loops of lengths $\leq 5d(M)$) (2.2.7).

§ 5. The multiplication in Γ is shown to be polynomial in the exponents (of $\gamma_1^{l_1}\ldots\gamma_n^{l_n}$ etc.) (5.1) and by extending this polynomial product from $\mathbb{Z}^n$ to $\mathbb{R}^n$ one embeds Γ as a uniform discrete subgroup in a nilpotent Lie group N (reproving Malcev's result). Only now do we have the model $\Gamma\backslash N$ to which a finite covering of M is to be compared! We find local maps $\exp_\gamma \circ \exp_{\gamma p}^{-1}$ from balls around the "lattice points" $\gamma\cdot p \in \tilde{M}$ $(\gamma \in \Gamma)$ into N which are compatible with the two actions of Γ on $\tilde{M}$ resp. on N (5.2); the local maps are interpolated (with the center of mass averaging of § 8) to a Γ-equivariant differentiable map $F : \tilde{M} \to N$. Finally - as in other pinching theorems - standard curvature control of geodesic constructions applied to $\tilde{M}$ and to a suitable left invariant metric (5.3) on N proves maximal rank for F and completes the proof of 1.5 (5.4).

2. Products of short geodesic loops and their holonomy motions.

In this section the Gromov product for sufficiently short geodesic loops is defined (2.2). It determines the fundamental group already under mild assumptions (2.2.7). The product and the commutator of loops is compared with the easily computable product und commutator of the holonomy motions of the loops, the error is curvature controlled (2.3, 2.4). Immediate applications of this "discrete group technique" are the Margulis lemma (2.5.2) and the corresponding volume bound for compact manifolds of negative curvature (2.5.3). In the almost flat case the commutator estimate (2.4) allows to find nilpotent subgroups of the fundamental group - the starting point of Gromov's proof.

Usually we work with curvature bounds $|K| \leq \Lambda^2$ for a compact Riemannian manifold M ; sometimes it is more convenient to distinguish lower und upper curvature bounds : $\delta \leq K \leq \Delta$. The specialization to the ε-flat case $|K| \leq \varepsilon \cdot d(M)^{-2}$ becomes more important in the next section.

2.1 The many loops at p

2.1.1 The lift of the Riemannian metric to T_pM .
Let M be a Riemannian manifold. Assume curvature bounds $|K| \leq \Lambda^2$ and fix a point $p \in M$. Then we have from 6.4.1 for the maximal rank radius of $\exp_p$

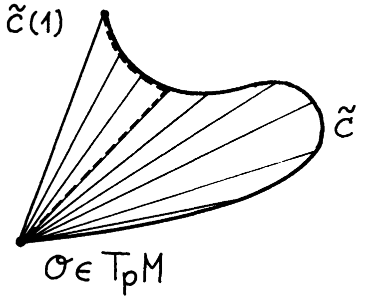
(i) $$r_{max} \geq \pi \cdot \Lambda^{-1} .$$

In a ball B_ρ of radius $\rho < \pi \cdot \Lambda^{-1}$ around $0 \in T_pM$ we have the Riemannian metric lifted from M via $\exp_p^{-1}$. If $v,w \in B_\rho$, $|v|+|w| \leq \rho$ then 6.4.1 gives the following comparison between the lifted Riemannian and the euclidean metric of T_pM .

(ii) $$\frac{\Lambda\rho}{\sinh \Lambda\rho} \cdot d(v,w) \leq |v-w| \leq \frac{\Lambda\rho}{\sin \Lambda\rho} \cdot d(v,w) .$$

2.1.2 Length decreasing homotopies.
Any closed curve c at p of length $\leq \rho$ can be lifted via $\exp_p^{-1}$ to a

curve $\tilde{c}$ in T_pM . By continuously replacing longer and longer arcs of $\tilde{c}$



by geodesic segments (fig.) one has "natural" length decreasing homotopies from $\tilde{c}$ to the geodesic ray $[0,1] \cdot \tilde{c}(1)$. In particular, for any $v \in T_pM$, $|v| \leq \rho - d(M)$, let c be the closed curve which is obtained from the geodesic $t \to \exp tv$ $(0 \leq t \leq 1)$ by joining the endpoint $\exp v$ by a shortest geodesic $(\leq d(M))$ to the initial point p . Then its lift $\tilde{c}$ provides a geodesic triangle $0vw$ with $d(v,w) \leq d(M)$ and $\exp w = p$. This and (2.1.1) proves the

2.1.3 Proposition. (Existence of many loops)
For any $v \in T_pM$, $|v| \leq \rho - d(M)$, there is a geodesic loop $\alpha : [0,1] \to M$ at p with

$$|\dot{\alpha}(0) - v| \leq \frac{\Lambda\rho}{\sinh \Lambda\rho} \cdot d(M) .$$

Or reformulated in terms of 2.1.4 with $\delta := \frac{\Lambda\rho}{\sinh \Lambda\rho} \cdot d(M)$: The initial tangents of geodesic loops at p are δ-dense in the ball $B_{\rho - d(M)}$ around $0 \in T_pM$.

2.1.4 Definition. A discrete subset D of a metric space is called δ-dense in a metric ball B_ρ of radius ρ , if for each $v \in B_\rho$ there is a $w \in D$ such that $d(v,w) \leq \delta$.

Among the geodesic loops of 2.1.3 fairly short ones suffice to generate the fundamental group:

2.1.5 Proposition. (Short generators for $\pi_1(M,p)$)
Let M be a compact Riemannian (or Finsler) manifold. For each $\eta > 0$ the

fundamental group $\pi_1(M,p)$ is generated by geodesic loops of length $\leq 2d(M) + \eta$. If $r_{max} > 2d(M)$ (see 2.1.1) then the loops of length $\leq 2d(M)$ generate $\pi_1(M,p)$. (See also 2.5.6)

Proof. Any closed curve c at p can be subdivided into arcs of lengths $\leq \eta$. The division points are joined to p by geodesic segments of length $\leq d(M)$. Thus c is represented as a product of closed curves of lengths $\leq 2d(M) + \eta$, which then are deformed via length decreasing homotopies to geodesic loops. This is particularly easy if $2d(M) + \eta < r_{max}$ (2.1.2); also, in this case, there are no conjugate points and hence only finitely many loops exist with lengths $l_i \in (2d(M), 2d(M)+\eta]$. These are eliminated by now choosing η sufficiently small.

2.2 Short homotopies and Gromov's product

2.2.1 Definition. A homotopy of loops at p is called short, if each of its curves is shorter than the maximal rank radius r_{max} of $\exp_p$. Equivalence classes under short homotopies will be called short homotopy classes.

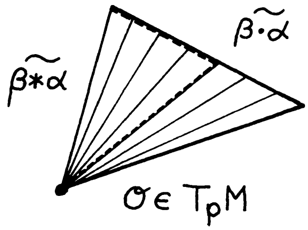
2.2.2 Proposition. There is exactly one geodesic loop in each short homotopy class at p .

Proof. The existence of at least one loop is explained in 2.1.2. If there were two, then there would be a short homotopy between them which by the definition of "short" can be lifted to a homotopy in T_pM with fixed endpoints. On the other hand the lifts of the two geodesic loops are radial segments pointing in different directions, a contradiction.

2.2.3 Definition. (Gromov's product of short geodesic loops)
Let α and β be geodesic loops at p . Denote their lengths by $|\alpha|, |\beta|$ and assume $|\alpha| + |\beta| < r_{max}$. Let $\beta \cdot \alpha$ be the product used in homotopy theory, namely the curve "α followed by β" . Define

> $\beta * \alpha$ is the unique geodesic loop in the short homotopy class of the curve $\beta \cdot \alpha$.

2.2.4 Remarks. (i) If one lifts $\beta * \alpha$ and $\beta \cdot \alpha$ via $\exp_p^{-1}$ to T_pM one obtains a geodesic triangle. The natural length decreasing homotopy by geodesic segments from 2.1.2 is particularly important in this case (see 2.3.1).



(ii) If one has a curvature bound $K \leq 0$, then every homotopy is short in the sense of 2.2.1; every homotopy class is represented by a unique geodesic loop and the Gromov product is the ususal product in the fundamental group $\pi_1(M,p)$. The Riemannian metric lifts to all of T_pM, $\exp_p : T_pM \to M$ is a locally isometric covering and $\pi_1(M,p)$ acts on the universal covering $\tilde{M} = T_pM$ as a group of fixed point free isometries, the so called deck transformations.

2.2.5 Proposition. (Restricted associativity)
If α,β,γ are geodesic loops at p which satisfy $|\alpha| + |\beta| + |\gamma| < r_{max}$ then

$$\alpha * (\beta * \gamma) = (\alpha * \beta) * \gamma .$$

The constant loop $\{p\} =: 1$ is a multiplicative unit. If $2|\alpha| < r_{max}$ and α^{-1} is the loop obtained from α by reversing the parametrization then

$$\alpha^{-1} * \alpha = 1 .$$

Proof. The length assumptions are such that the standard homotopies from homotopy theory are short; for example we have short homotopies from $\alpha * (\beta * \gamma)$ to $\alpha \cdot (\beta \cdot \gamma)$ to $(\alpha \cdot \beta) \cdot \gamma$ to $(\alpha * \beta) * \gamma$. The uniqueness statement 2.2.2 completes the proof.

2.2.6 Notations. 2.1.4, 2.2.4, 2.2.5 and 2.2.7 show that the set of short loops (shorter than r_{max}) together with the Gromov product $*$ are closely related to the fundamental group $\pi_1(M,p)$. We emphasize this by introducing (for $\rho < r_{max}$)

$$\pi_\rho := \{\alpha ; \alpha \text{ a geodesic loop at } p \text{ with } |\alpha| \leq \rho\} .$$

We write $*$-products without brackets: $\alpha_1 * \ldots * \alpha_m$ if $\Sigma |\alpha_i| < r_{max}$ (see 2.2.5). We abbreviate

$$\alpha^k = \alpha * \ldots * \alpha , \; \alpha^{-k} = \alpha^{-1} * \ldots * \alpha^{-1} \quad (\text{if } k \cdot |\alpha| < r_{max}) ,$$

$$[\beta,\alpha] = \beta * \alpha * \beta^{-1} * \alpha^{-1} \qquad (\text{if } 2|\alpha| + 2|\beta| < r_{max}).$$

$[\beta,\alpha]$ is called a 2-fold commutator; m-fold commutators are defined inductively: If γ is a k-fold and δ an l-fold commutator, then $[\gamma,\delta]$ is (k+l)-fold. Note that in this definition it is <u>not</u> assumed that the lengths of the loops involved are so short that the associative law (2.2.5) can unrestrictedly be used; for example, if in $\delta * \gamma * \delta^{-1} * \gamma^{-1}$ we substitute $\gamma = \beta * \alpha * \beta^{-1} * \alpha^{-1}$ we require only that α and β are short enough for γ to be defined and γ and δ are short enough for $[\gamma,\delta]$ to be defined.

2.2.7 <u>Proposition.</u> (π_ρ determines $\pi_1(M,p)$)
Assume $2d(M) < \rho < r_{max} - 2d(M)$.
Let $W(\pi_\rho)$ be the free group of words in the elements of π_ρ ; let $N_o(\pi_\rho)$ be the set of words $\alpha\,\beta\,\gamma^{-1}$ where $\gamma = \alpha * \beta$ (the "short relations", one needs $|\alpha| + |\beta| \leq \rho + 2d(M)$); let $N(\pi_\rho)$ be the smallest normal subgroup in $W(\pi_\rho)$ which contains $N_o(\pi_\rho)$. Then

$$\hat{\pi}_\rho := W(\pi_\rho) \,/\, N(\pi_\rho)$$

is the <u>group presented by</u> π_ρ . There is a natural isomorphism

$$\hat{\Phi} : \hat{\pi}_\rho \longrightarrow \pi_1(M,p) ,$$

which can be defined by mapping the word $w = \alpha_1 \ldots \alpha_m \in W$ to the homotopy class of the closed curve $\Phi(w) := \alpha_1 \cdot \ldots \cdot \alpha_m$.

Remark. The proposition states that $\pi_1(M,p)$ is determined by the short loops and their products if only $4d(M) < r_{max}$. It does not state or imply that the set π_ρ can be identified with a subset of $\pi_1(M,p)$. This so called injectivity of π_ρ is false in general; it is true for sufficiently ε-flat manifolds, but this fact is established only late in the proof of Gromov's theorem (4.6.5).

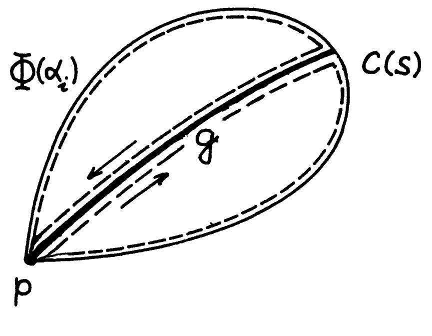
Proof. Since elements of $N(\pi_\rho)$ are products of words of the form $\alpha\,\nu\,\alpha^{-1}$ with $\nu \in N_o$, $\alpha \in W$, it is clear that Φ maps $N(\pi_\rho)$ to nullhomotopic curves in M . Therefore $\hat{\Phi}$ is well defined, clearly a homomorphism and, because of 2.1.5, surjective. To prove injectivity of $\hat{\Phi}$ assume that for a word $w \in W(\pi_\rho)$ the closed curve $c = \Phi(w)$ is nullhomotop in M ; we have to show $w \in N(\pi_\rho)$ by reducing w to the trivial loop with a finite number of substitutions

$$\ldots\alpha\beta\ldots \to \ldots\gamma\ldots \quad \text{(or vice versa), where } \alpha,\beta,\gamma \in \pi_\rho \text{ and } \gamma = \alpha * \beta .$$

The curve $c = \Phi(w) : [a,b] \to M$ comes with a subdivision $a = \sigma_o < \sigma_1 < \ldots < \sigma_n = b$ of $[a,b]$ such that each letter $\alpha_i \in \pi_\rho$ of the word w corresponds to the restriction of c to one of the subintervals:

$$\Phi(\alpha_i) = c|_{[\sigma_{i-1},\sigma_i]} .$$

We first show that words which correspond to further refinements of $[a,b]$ are mod $N(\pi_\rho)$ the same as w . By induction it suffices to include one more subdivision point s in one of the intervals $[\sigma_{i-1},\sigma_i]$. Join $c(s)$ by a



<u>minimizing</u> geodesic g to p. Then each of the closed curves $g \cdot c|[\sigma_{i-1}, s]$ resp. $c|[s,\sigma_i] \cdot g^{-1}$ is at most as long as $\alpha_i \in \pi_\rho$ and therefore short homotop to a geodesic loop α_i' resp. α_i''. We have $\alpha_i'' * \alpha_i' = \alpha_i$ since $|\alpha_i'| + |\alpha_i''| \leq |\alpha_i| + 2d(M) < r_{max}$; this proves $\alpha_i''\alpha_i' = \alpha_i \bmod N(\pi_\rho)$.

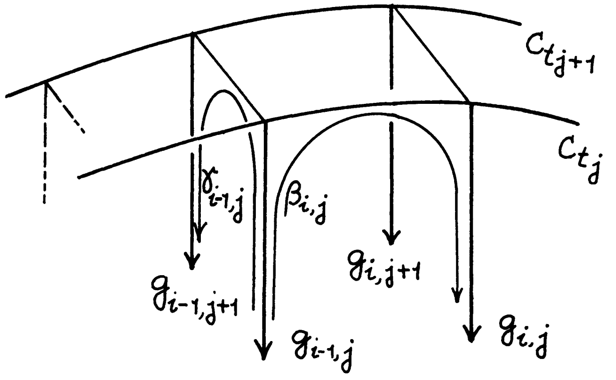
Next let c_t, $0 \leq t \leq 1$, be a piecewise differentiable homotopy from $c = c_o$ to $\{p\} = c_1$. By uniform continuity we choose subdivisions $a = s_o < s_1 < \ldots < s_L = b$ (a refinement of $\{\sigma_i\}$) and $0 = t_o < t_1 < \ldots < t_e = 1$ such that the curves $s \to c_{t_j}(s)$, $s_{i-1} \leq s \leq s_i$, and $t \to c_t(s_i)$, $t_{j-1} \leq t \leq t_j$ have lengths $\leq \frac{1}{3}\eta$, where $\eta = \min(r_{max} - 4d(M)$, injectivity radius of M). Again join the points $c_{t_j}(s_i)$ by minimizing geodesics g_{ij} of length $\leq 2d(M)$ to p. Then each curve c_{t_j} is represented as a product of the short curves

$$g_{ij} \cdot c_{t_j} \Big| [s_{i-1}, s_i] \cdot g_{i-1,j}^{-1}$$

each of which is shorthomotopic to a geodesic loop β_{ij}, $|\beta_{ij}| \leq 2d(M) + \frac{1}{3}\eta$.

We also introduce the geodesic loops γ_{ij}, $|\gamma_{ij}| \leq 2d(M) + \frac{1}{3}\eta$ which are short homotopic to $g_{ij+1} \cdot c_t(s_i) \Big| [t_j, t_{j+1}] \cdot g_{ij}^{-1}$. Then: Each of the loops β_{ie}, γ_{oj}, γ_{Lj} is the trivial loop and we have the short relations

$$\gamma_{ij} * \beta_{ij} = \beta_{i,j+1} * \gamma_{i-1,j} \quad \text{or} \quad \gamma_{ij}\,\beta_{ij} = \beta_{i,j+1}\,\gamma_{i-1,j} \bmod N(\pi_\rho),$$



since only curves of length $\leq 4d(M) + \frac{2}{3}\eta$ are involved. With this and associativity in $W(\pi_\rho)$ induction over i gives immediately

$$\beta_{L,j+1} \cdots \beta_{oj+1}\, \gamma_{oj} = \gamma_{Lj}\, \beta_{Lj} \cdots \beta_{oj} \mod N(\pi_\rho) .$$

This is true for $j = 0,\ldots,e$ and proves $w = \beta_{Lo}\ldots\beta_{oo} = 0 \mod N(\pi_\rho)$.

2.3 The map into the affine holonomy

To each geodesic loop α at p we associate its holonomy motion (6.2.2)

$$m(\alpha) : T_pM \to T_pM$$

$$m(\alpha)(x) = r(\alpha) \cdot x + t(\alpha)$$

where $r(\alpha)$ is the Levi-Civita translation around α and is called the rotational part of the motion $m(\alpha)$ or of the loop α ,

and $$t(\alpha) = \dot{\alpha}(1) = r(\alpha) \cdot \dot{\alpha}(o)$$

is the tangent vector at the endpoint of the geodesic loop and is called the translational part of $m(\alpha)$ or of α .

We use the distance on the orthogonal group from (7.3) to compare the holonomy motion of the Gromov product $\beta * \alpha$ to the composition of the holonomy motions of α and β .

2.3.1 Proposition. (The holonomy map is almost homomorphic)
Let M be a complete Riemannian manifold with curvature bounds $|K| \leq \Lambda^2$ (e.g. $\Lambda^2 = \varepsilon \cdot d(M)^{-2}$ in the almost flat case). Consider loops $\alpha,\beta \in \pi_{\Lambda^{-1}}$, then $\beta * \alpha$ is defined and

(i) $$d(r(\beta) \circ r(\alpha), r(\beta * \alpha)) \leq \Lambda^2 \cdot |t(\alpha)| \cdot |t(\beta)|$$

(ii) $$|t(m(\beta) \circ m(\alpha)) - t(\beta * \alpha)| \leq \Lambda^2 \cdot |t(\alpha)| \cdot |t(\beta)| \, (|t(\alpha)| + |t(\beta)|).$$

Proof. By definition $m(\beta) \circ m(\alpha) = m(\beta \cdot \alpha)$. Therefore one has to control the change of m under a good homotopy from $\beta \cdot \alpha$ to $\beta * \alpha$. We choose the one from 2.2.4(i) and apply 6.2 and 6.7. The longest curve in this homotopy is $\beta \cdot \alpha$ with length $|\beta \cdot \alpha| = |t(\alpha)| + |t(\beta)|$. From $|K| \leq \Lambda^2$ we have for the curvature tensor $||R|| \leq \frac{4}{3} \Lambda^2$ (6.1.1), then 6.2.1 gives

$$|\measuredangle(r(\beta) \circ r(\alpha) \cdot x\, ,\, r(\beta * \alpha) \cdot x)| \leq \frac{4}{3} \Lambda^2 \cdot F\,.$$

The area F of the homotopy is estimated with 6.7, 6.7.1 as

$$F \leq 1.25 \frac{|t(\alpha)| \cdot |t(\beta)|}{2}\,.$$

This proves (i), and (ii) follows similarly from 6.2.2.

2.3.2 Remark. The proof shows that 2.3.1 can be regarded as a convenient reformulation of the path-dependence of Levi-Civita and affine parallel translation. It is a typical pinching result, since the composition of motions $m(\beta) \circ m(\alpha)$ is easily obtained and approximates $m(\beta * \alpha)$ up to curvature controlled errors. The estimates 2.3.1 do not allow to take advantage of cancellations usually occuring when commutators are computed; good commutator estimates are derived in section 2.4.

2.4 Commutator estimates

Since the geometric importance of an inequality can only be explained through its applications we refer the reader to 2.5 to appreciate the following theorem. It is also the key result from which Gromov's arguments start. Recall notations and definitions from 2.2.6 and 7.3.

2.4.1 Theorem. (Commutator estimates)
Let M be a Riemannian manifold; assume curvature bounds either

$$\text{(a)} \quad |K| \leq \Lambda^2 \quad \text{or} \quad \text{(b)} \quad -\Lambda^2 \leq K \leq 0\,.$$

Let α, β be loops at p, which in case (a) also satisfy $|\alpha| + |\beta| \leq \pi/3\Lambda$.

Then the Gromov commutator $[\beta,\alpha]$ is defined and satisfies

(i) $\Lambda \cdot |t([\beta,\alpha])| \leq \|r(\alpha)\| \cdot \sinh \Lambda|t(\beta)| + \|r(\beta)\| \cdot \sinh \Lambda|t(\alpha)|$

$$+ \frac{2}{3}\Lambda^2|t(\alpha)| \cdot |t(\beta)| \sinh \Lambda(|t(\alpha)| + |t(\beta)|)$$

and trivially also $\leq 2\Lambda(|t(\alpha)| + |t(\beta)|)$.

(ii) $d(r([\beta,\alpha]), [r(\beta),r(\alpha)]) \leq$

$$\frac{5}{3}\Lambda^2 |t(\alpha)| \cdot |t(\beta)| + \frac{5}{6}\Lambda^2 |t([\beta,\alpha])| \cdot (|t(\alpha)| + |t(\beta)|) .$$

(iii) $|t([\beta,\alpha]) - t([m(\beta),m(\alpha)])| \leq$

$$2(|t(\alpha)| + |t(\beta)|) \cdot (\frac{5}{3}\Lambda^2|t(\alpha)| \cdot |t(\beta)| + \frac{5}{6}\Lambda^2|t([\beta,\alpha])| \cdot (|t(\alpha)| + |t(\beta)|).$$

(If $K \leq 0$ then the constants $\frac{5}{3}$ resp. $\frac{5}{6}$ can be replaced by 1 resp. $\frac{1}{2}$.)

We include two immediate corollaries for quicker reference.

2.4.2 Corollary. (Simplification of 2.4.1)
Choose in definition 7.3.2 the parameter $c \geq 8$. Assume in addition to 2.4.1 $|\alpha|,|\beta| \leq (16\,\Lambda)^{-1}$, then

(i) $|t([\beta,\alpha])| \leq 1.006\,(\|m(\alpha)\| \cdot |t(\beta)| + \|m(\beta)\| \cdot |t(\alpha)|)$.

The error term from 2.4.1 (ii), (iii) simplifies to

(ii) $\frac{5}{3}\Lambda^2|t(\alpha)||t(\beta)| + \frac{5}{6}\Lambda^2|t([\beta,\alpha])| \cdot (|t(\alpha)| + |t(\beta)|) \leq$

$$\leq 0.21\,\Lambda(\|m(\alpha)\| \cdot |t(\beta)| + \|m(\beta)\| \cdot |t(\alpha)|)$$

(iii) $\|m([\beta,\alpha])\| \leq 2.03\,\|m(\alpha)\| \cdot \|m(\beta)\|$.

2.4.3 Remark. In this form the result can easily be compared with commutator estimates in the group of motions:

$$|t([m(\beta),m(\alpha)])| \le ||r(\beta)|| \cdot |t(\alpha)| + ||r(\alpha)|| \cdot |t(\beta)|$$

$$|r([m(\beta),m(\alpha)])| \le ||[r(\beta),r(\alpha)]|| \le 2\,||r(\alpha)|| \cdot ||r(\beta)|| .$$

We see that the homotopy errors are under effective curvature control. It is standard to conclude from these estimates that iterated commutators of a set of motions with rotational parts $< \frac{1}{2}$ and bounded translational parts converge to the identity. Now almost the same follows for Gromov commutators of loops:

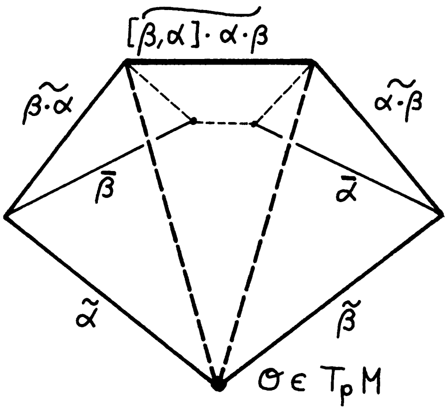
2.4.4 Corollary. (Convergence of commutators)
Assume in addition to 2.4.2 $||m(\alpha)||$, $||m(\beta)|| \le 0.49$, which is essentially an additional assumption on the rotational parts. Then

(i) $$||m([\beta,\alpha])|| \le 0.995 \min(||m(\alpha)|| , ||m(\beta)||)$$

(ii) $$|t([\beta,\alpha])| \le 0.493\,(|t(\alpha)| + |t(\beta)|) .$$

Proof. For the corollaries simplify 2.4.1 (i) with $\frac{2}{3}\sinh\frac{1}{8} \le 0.084$ and $\frac{1}{r}\sinh r \le 1.001$ if $r \le \frac{1}{16}$. This gives 2.4.2 (i) which then implies 2.4.2 (ii). The other inequalities are restatements.

2.4.1 (ii) and (iii) follow with the same arguments as in section 2.3: If we lift the closed null-homotopic curve $\alpha^{-1}\cdot\beta^{-1}\cdot[\beta,\alpha]\cdot\alpha\cdot\beta$ via $\exp_p^{-1}$ to



T_pM we obtain the geodesic pentagon with edgelengths $|\beta|, |\alpha|, |[\beta,\alpha]|$, $|\beta|, |\alpha|$ shown in the figure. We subdivide this pentagon by two (dotted) geodesics of length $\leq |t(\alpha)| + |t(\beta)|$ into three triangles and use the distance decreasing homotopies from 2.1.2 to span "ruled surfaces" into these triangles. The curvature controlled path dependence of Levi Civita translation (6.2.1) along this homotopy is used to estimate the distance between $r([\beta,\alpha])$ and $[r(\beta), r(\alpha)]$; inserting $||R|| \leq \frac{4}{3}\Lambda^2$ (6.1.1) and $F \leq \frac{5}{8}(2|\alpha|\cdot|\beta| + |[\beta,\alpha]|\cdot(|\alpha| + |\beta|))$ (6.7, 6.7.1) gives 2.4.1 (ii) . Similarly the control (6.2.2) of affine translations gives 2.4.1 (iii) ; note that the longest curve in our homotopy from $[\beta,\alpha]$ to $\beta\cdot\alpha\cdot\beta^{-1}\cdot\alpha^{-1}$ has length $2(|\alpha| + |\beta|)$. Finally, to obtain 2.4.1 (i) introduce two more geodesics $\bar{\alpha}$ (resp. $\bar{\beta}$) which start from $\tilde{\beta}(1)$ (resp. $\tilde{\alpha}(1)$) in directions obtained by parallel translation of $\tilde{\alpha}(o)$ along $\tilde{\beta}$ (resp. of $\tilde{\beta}(o)$ along $\tilde{\alpha}$). Rauch's result (6.4.1) gives

$$\Lambda \cdot d(\widetilde{\alpha\cdot\beta}(1), \bar{\alpha}(1)) \leq ||r(\beta)|| \cdot \sinh \Lambda|t(\alpha)| ,$$

$$\Lambda \cdot d(\widetilde{\beta\cdot\alpha}(1), \bar{\beta}(1)) \leq ||r(\alpha)|| \cdot \sinh \Lambda|t(\beta)| ;$$

and 6.6.1 gives

$$\Lambda \cdot d(\bar{\beta}(1), \bar{\alpha}(1)) \leq \frac{2}{3}\Lambda^2 |t(\alpha)| \cdot |t(\beta)| \sinh \Lambda (|t(\alpha)| + |t(\beta)|) .$$

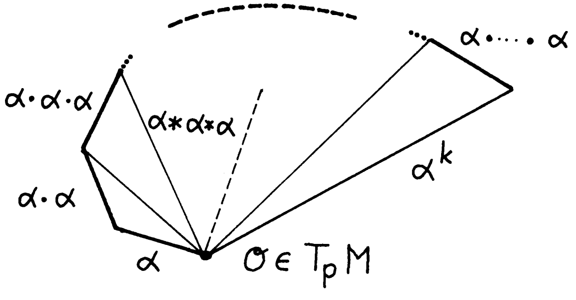
The sum of these distances is the bound 2.4.1 (i) for $|t([\beta,\alpha])|$.

2.5 Applications

We discuss examples of global results, which were used by Gromov for the same purpose, namely to illustrate the power of the discrete group technique and in particular of the commutator estimates. A simple volume comparison proves the existence of loops with small rotational parts; then the Margulis Lemma follows from 2.4.4; it implies a lower volume bound for compact manifolds with curvature bounds $-1 \leq K < 0$. Finally, the volume comparison is used again to obtain curvature controlled information about the fundamental group.

2.5.1 Proposition. (Existence of small rotational parts)
Assume curvature bounds $|K| \leq \Lambda^2$ and choose η, $0 < \eta \leq 0.49$. Put $\tilde{\eta} = 0.99\eta$, $m = 2 \operatorname{int} (2\pi/\tilde{\eta})^{n/2}$. Then: For each geodesic loop α at p of length $|\alpha| \leq \frac{\pi}{8\Lambda} \cdot (\tilde{\eta}/2\pi)^{1+n/2}$ there exists $k \leq m$ such that $||m(\alpha^k)|| \leq \eta$. (Use $c = 8$ in 7.3.2).

Proof. First, with 7.6.1 (ii) choose $k \leq m$ such that $||r(\alpha)^k|| \leq \tilde{\eta}$. The area of the natural homotopy (2.1.2) from $\alpha\cdot\ldots\cdot\alpha$ to $\alpha^k = \alpha * \ldots * \alpha$



is bounded by applying 6.7, 6.7.1 to the $(k-1)$ triangles into which the homotopy can be decomposed:

$$F \leq \frac{5}{8} |\alpha|^2 \cdot (1+2+\ldots+(k-1)) .$$

Now 6.1.1 and 6.2.1 give

$$d(r(\alpha^k), r(\alpha)^k) \leq \Lambda^2 \cdot |t(\alpha)|^2 \cdot \frac{m^2}{2} ,$$

and using the assumption $8 \Lambda m|\alpha| \leq \tilde{\eta} \leq \frac{1}{2}$ we obtain

$$||r(\alpha^k)|| \leq (1 + \frac{1}{256})\tilde{\eta} \leq \eta .$$

2.5.2 Corollary. (Margulis Lemma)
Assume curvature bounds $|K| \leq \Lambda^2$. If $a,b \in \pi_1(M,p)$ are represented by geodesic loops α, β at p which satisfy

$$\Lambda|\alpha| , \Lambda|\beta| \leq 2\cdot 4^{-(n+3)} \quad (n = \dim M) ,$$

then there are $k, l \leq 2(3.6)^n$ such that a^k and b^l generate a nilpotent subgroup of $\pi_1(M,p)$.

Proof. With 2.5.1 choose k and l such that $||m(\alpha^k)||, ||m(\beta^l)|| \leq 0.49$. From the existence of a shortest nontrivial loop at p and the estimate 2.4.4 one finds a number d such that any d-fold commutator (2.2.6) of α^k and β^l is the trivial loop. Now the commutator identity

$$[a,bc] = [a,b] \cdot [b,[a,c]] \cdot [a,c]$$

implies that any d-fold commutator in the group generated by a^k, b^l in $\pi_1(M,p)$ is trivial, since d-fold commutators of products can be rewritten as (much longer) products of at least d-fold commutators of the generators a^k, b^l .

2.5.3 Proposition. (Volume bound from below)
Let M^n be an n-dimensional compact Riemannian manifold with curvature bounds $-1 \leq K < 0$ and diameter $d(M)$.

(i) There exists a point $q \in M$ such that the injectivity radius of $\exp_q$ is

$$r_q \geq \rho := 4^{-(n+3)}$$

(ii) $$\mathrm{vol}(M) \geq \frac{1}{n} \mathrm{vol}(S^{n-1}) \rho^n ,$$

note that this bound depends only on the dimension, in particular not on $\max K$.

(iii) For all $p \in M$ the injectivity radius r_p of $\exp_p$ satisfies

$$r_p \geq \frac{\pi}{n} \frac{\mathrm{vol}(S^{n-1})}{\mathrm{vol}(S^n)} \left(\frac{\rho}{\sinh d(M)} \right)^n \cdot \sinh d(M) .$$

Remark. The last inequality says that in this general situation one has a similar phenomenon as for surfaces of constant curvature -1 : A very small injectivity radius (for surfaces: a very thin handle) can only occur if the diameter is large.

Proof. (ii) follows from (i) since $\exp_q$ is expanding (6.4.1) and the given bound is the volume of a ball of radius ρ in T_qM.
(iii) follows from (ii) by estimating the normal exponential map of the shortest closed geodesic c on M using 6.3.8 (see [16] for more details):

$$\mathrm{vol}(M^n) \leq \mathrm{length}\,(c) \cdot \mathrm{vol}(S^{n-2}) \cdot \int_0^{d(M)} \sinh^{n-2} r \cdot \cosh r \; dr$$

$$\leq \mathrm{length}\,(c) \cdot \frac{1}{2\pi} \cdot \mathrm{vol}(S^n) \cdot \sinh^{n-1} d(M) \; .$$

To prove (i) recall 2.2.4 (ii) and the following facts from the geometry of compact manifolds M of negative curvature: Each element α of the deck-group Γ has in the universal covering $\tilde{M}$ exactly one invariant geodesic c, the "axis" of α. The set of all $\alpha \in \Gamma$ with a common axis c is a cyclic subgroup Z of Γ since Z is discrete and - if restricted to c - a translational subgroup of $\mathbb{R}$. (The restriction map is an isomorphism since the elements of Z have no fixed points.) Now every nilpotent subgroup N of Γ has an axis and is therefore cyclic (Preismann):

(i) If α and β commute and c is the axis of α then $\alpha = \beta\,\alpha\,\beta^{-1}$ shows that $\beta(c)$ is invariant under α, hence $\beta(c) = c$ so that c is the axis of β.

(ii) For $\alpha,\beta \in \Gamma$ assume that α and $[\beta,\alpha]$ commute, i.e. have a common axis c (by (i)). Then $\alpha(c) = c$ and $\beta\,\alpha\,\beta^{-1}\alpha^{-1}(c) = c$ imply $\alpha(\beta^{-1}(c)) = \beta^{-1}(c)$ so that $\beta^{-1}(c)$ is invariant under α, i.e. $\beta(c) = c$ (uniqueness of the axis). Therefore α and β generate a cyclic subgroup, in particular they commute. By induction over the iterated commutator subgroup of N it follows that the elements of N have a common axis.

These standard results are combined with the commutator estimates to give

2.5.4 Margulis lemma for negative curvature.
Let α,β be loops at $q \in M$ which satisfy $|\alpha|,|\beta| \leq 2\cdot 4^{-(n+3)} = 2\rho$, then α and β have a common axis, hence generate a cyclic subgroup of Γ.

Indeed, by 2.5.2 suitable powers α^k , β^l generate a nilpotent subgroup N of the deckgroup which as we have recalled has a common invariant geodesic c ; clearly c is also the common axis of α and β .

To finish the proof of 2.5.3 (i) let α be a shortest closed geodesic on M . If $|\alpha| \geq 2\rho$ there is nothing to prove. If $|\alpha| < 2\rho$ choose $p = \alpha(o)$ and take $\exp_p : T_pM \to M$ as the locally isometric universal covering (2.1.1, 2.2.4 (ii)). Let $H \subset \Gamma$ be the maximal cyclic subgroup which has the lift $\tilde{\alpha}$ of α as axis and let $\tilde{\gamma}$ be a geodesic which is orthogonal to $\tilde{\alpha}$ at $\tilde{\alpha}(o) = \tilde{\gamma}(o)$. For only finitely many $k \in H$ we can have $\tilde{d}(\tilde{\gamma}(o), k(\tilde{\gamma}(O))) < 2\rho$, therefore 6.5.1 implies:

2.5.5 There is a smallest $\tau > 0$ such that we have for some $k_o \in H$ and for all $k \in H : \tilde{d}(\tilde{\gamma}(\tau), k(\tilde{\gamma}(\tau))) \geq \tilde{d}(\tilde{\gamma}(\tau), k_o(\tilde{\gamma}(\tau))) = 2\rho$.

Now the injectivity radius at $q = \exp_p \tilde{\gamma}(\tau)$ is $\geq \rho$, since otherwise there would be a loop β at q with $|\beta| < 2\rho$. Then on the one hand $\beta \notin H$ because of (2.5.5); on the other hand $|\beta|, |k_o| \leq 2\rho$ and 2.5.4 imply that β and k_o have a common axis (namely $\tilde{\alpha}$), i.e. $\beta \in H$, a contradiction.

The simple volume comparison argument which gave 2.5.1 via 7.6.1 is frequently used in Gromov's paper. In the following it is combined with Toponogow's angle comparison theorem for geodesic triangles (6.4.3) and with the "short basis" trick to produce curvature implied information about the fundamental group.

2.5.6 <u>Proposition</u>. (Number of generators for $\pi_1(M)$)

(i) Let M^n be a complete Riemannian manifold of nonnegative curvature. Then the fundamental group can be generated by $s \leq 2 \cdot 5^{\frac{1}{2}n}$ elements.

(ii) Let M^n be a compact Riemannian manifold with diameter $d(M) < D/2$ and curvature $K \geq -\Lambda^2$, then the fundamental group can be generated by

$$s \leq 2 \cdot (3 + 2 \cosh \Lambda D)^{\frac{1}{2}n}$$

elements.

Proof. Represent each element of $\pi_1(M,p)$ by a shortest geodesic loop α at p and call $|\alpha|$ the length of the homotopy class.

Pick a short basis $\{\alpha_1,\dots,\alpha_s\}$ for $\pi_1(M,p)$ as follows:

(i) α_1 represents a nontrivial homotopy class of minimal length.

(ii) If $\alpha_1,\dots,\alpha_k$ have already been chosen, then α_{k+1} represents a homotopy class of minimal length in the complement of the subgroup generated by $\{\alpha_1,\dots,\alpha_k\}$.

By definition $|\alpha_1| \leq |\alpha_2| \leq \dots$ and

$$(*) \qquad |\alpha_i\alpha_j^{-1}| \geq \max\{|\alpha_i|\,,\,|\alpha_j|\}\,,$$

(otherwise α_i or α_j were not chosen minimally). We lift the loops α_i, α_j to the universal covering $\tilde{M}$ of M and obtain a triangle with edge lengths $|\alpha_i|$, $|\alpha_j|$, $|\alpha_i\alpha_j^{-1}|$. By Toponogow's theorem 6.4.5 the angle ϕ opposite to $|\alpha_i\alpha_j^{-1}|$ is not smaller than the corresponding angle in a triangle with the same edgelengths

a) in the euclidean plane if $K \geq 0$

b) in the hyperbolic plane of curvature $-\Lambda^2$ if $K \geq -\Lambda^2$.

In case a) the inequality $(*)$ implies $\phi \geq 60^\circ$ since

$$|\alpha_i\alpha_j^{-1}|^2 \leq |\alpha_i|^2 + |\alpha_j|^2 - 2|\alpha_i| \cdot |\alpha_j| \cos\phi\,.$$

Now there are at most

$$\left(\frac{1 + \sin\frac{1}{2}\phi}{\sin\frac{1}{2}\phi}\right)^n$$

unit vectors with pairwise angles $\geq \phi$ since the balls of radius $\sin\frac{1}{2}\phi$ around the endpoints of these vectors in R^n are on the one hand disjoint and on the other hand contained in the ball of radius $(1 + \sin\frac{1}{2}\phi)$ around the origin, so that the volume ratio is an obvious bound for the number of vectors. If one takes only the inner half of the small balls, then these are contained in the ball of radius $(1 + \sin^2 \frac{1}{2}\phi)^{\frac{1}{2}}$ which gives the bound $2(1 + \sin^{-2}\frac{1}{2}\phi)^{\frac{1}{2}n}$, an improvement if ϕ is not too small.

To get a lower bound for ϕ in case b) recall from 2.1.5 that it is sufficient to consider generators of length $\leq D = 2d(M) + \eta$, where $\eta > 0$ may be chosen arbitrarily small. Then the cosine formula of hyperbolic geometry gives as in case a) (using (✱) and Toponogow's theorem)

$$\cos\phi \leq \frac{\cosh \Lambda|\alpha_i| \cdot \cosh \Lambda|\alpha_j| - \cosh \Lambda|\alpha_i\alpha_j^{-1}|}{\sinh \Lambda|\alpha_i| \cdot \sinh \Lambda|\alpha_j|}$$

$$\leq \frac{\cosh \Lambda D}{1 + \cosh \Lambda D},$$

hence $1 + \sin^{-2}\frac{1}{2}\phi \leq 3 + 2\cosh \Lambda D$.

3. Loops with small rotational parts

The following fact from the flat case (the Bieberbach theorem 1.2 or $\varepsilon = 0$ in Gromov's theorem 1.5) should be kept in mind throughout this section:

> The translational subgroup $(\approx \mathbb{Z}^n)$ of the deckgroup $\pi_1(M,p)$ of a compact flat manifold M^n can be characterized as the subset of π_1 with not too large rotational parts, more precisely
>
> $$\text{Translations in } \pi_1 = \{\gamma \in \pi_1 ;\ \ ||r(\gamma)|| < \tfrac{1}{2}\} \approx \mathbb{Z}^n .$$

This fact follows from Bieberbach's theorem but it plays no role in the classical proofs. In Gromov's approach however it plays a central role since it has a generalization to the almost flat case: If the holonomy motion of not too long a loop γ has its rotational part $||r(\gamma)|| \leq 0.48$ then it follows $||r(\gamma)|| \leq \Theta$, where Θ can be taken arbitrarily small in the flat case and very small if the manifold is ε-flat (1.3) with ε very small. With this result (3.4.2) one obtains in the flat case the lattice subgroup right away while in the almost flat case one can construct a torsionfree nilpotent subgroup Γ of finite index in $\pi_1(M,p)$ by generating Γ from fairly short loops with small rotational parts.

This section proceeds in the following steps:

3.1 adapts the short basis trick (2.5.6) to Γ_ρ, the set of loops at p of length $\leq \rho$ and rotational parts ≤ 0.48 ; any short basis has at most $d(n) = (3.02)^{\frac{1}{2}n(n+1)}$ elements.

3.2 selects a length ρ such that the translational parts of those loops in π_ρ which have very small rotational part are fairly dense in the ball $B_\rho \subset T_pM$.

3.3 shows that at least those elements $\gamma \in \Gamma_\rho$ which have trivial $d(n)$-fold iterated commutators have their rotational parts much smaller than 0.48 .

3.4 proves by a length controlled induction that all $d(n)$-fold iterated commutators in Γ_ρ are trivial.

3.5 collects the consequences of (3.3) and (3.4) for the multiplication of short loops.

3.6 applies the previous information to obtain for example in the flat case an estimate for the index of the lattice subgroup in the deck group; these arguments are the same in the almost flat case.

3.1 The short basis

3.1.1 Definition. (Length controlled generation)
For any subset $A \subseteq \pi_\rho$ (see 2.2.6) define the set $\langle A \rangle_\rho$ inductively by

(i) $\{1\} \cup A \cup A^{-1} \subseteq A_\rho$

(ii) If $\alpha, \beta \in \langle A \rangle_\rho$ and $|\alpha * \beta| \leq \rho$, then $\alpha * \beta \in \langle A \rangle_\rho$.

Note that one has to be carefull with associativity (2.2.5); for example $\alpha * (\beta * \gamma) \in \langle A \rangle_\rho$ does not imply $(\alpha * \beta) \in \langle A \rangle_\rho$.

3.1.2 Definition. (Loops with small rotational part)

$$\Gamma_\rho = \{\alpha \in \pi_\rho \; ; \; \|r(\alpha)\| \leq 0.48\} \qquad (2.2.6,\ 7.3),$$

$$\tilde{\Gamma}_\rho = \langle \Gamma_\rho \rangle_\rho \qquad (3.1.1).$$

Eventually $\tilde{\Gamma}_\rho = \Gamma_\rho$ will be proved for a suitable ρ (3.4.1).

3.1.3 Definition. (Short basis)
A short basis $\{\alpha_1, \ldots, \alpha_d\}$ for Γ_ρ (3.1.2) will be defined below ((i), (ii)) in such a way that it reflects nilpotent properties of Γ_ρ (3.1.4) and such that the a priori bound 3.1.5 for the number d of its elements can be proved. For the proofs it is necessary that the largest rotational parts allowed in 3.1.2 and the largest translational parts have the same weight; therefore we use here the following distance function on the group of motions

$$\tilde{d}(\tilde{A}, \mathrm{id}) = \max\left(d(A, \mathrm{id}) \; , \; \frac{0.48}{\rho} \cdot |a|\right) =: \; \|\tilde{A}\| \; ; \; \tilde{d}(\tilde{A}, \tilde{B}) = \|\tilde{A}^{-1}\tilde{B}\| \; .$$

This choice determines the parameter c in 7.3, $c = \dfrac{0.48}{\Lambda\rho}$, and since we used

$c \geq 8$ in 2.3 and 2.4 we have to assume the curvature bounds

$$\text{(K)} \qquad \Lambda\rho \leq 0.06$$

for the present arguments.
Now the elements of the short basis are inductively selected:

(i) $\alpha_1 \in \Gamma_\rho$ $(\alpha_1 \neq 1)$ is such that $\|m(\alpha_1)\|$ is minimal in Γ_ρ .

(ii) If $\{\alpha_1 \ldots, \alpha_i\} \subset \Gamma_\rho$ have been selected, then $\alpha_{i+1} \in \Gamma_\rho \setminus \langle\{\alpha_1, \ldots, \alpha_i\}\rangle_\rho$ (3.1.1) is chosen so that $\|m(\alpha_{i+1})\|$ is minimal.

Note that d is finite since $\tilde{\Gamma}_\rho$ is finite. Furthermore $\{\alpha_1, \ldots, \alpha_d\}$ is also a short basis for Γ_ρ since the elements $\beta \in \tilde{\Gamma}_\rho \setminus \Gamma_\rho$ have $\|m(\beta)\| > 0.48$.

3.1.4 <u>Proposition</u>. (Nilpotency of the short basis)
For each α_i and all $\gamma \in \Gamma_\rho$ holds

$$[\alpha_i, \gamma] \in \langle\{\alpha_1, \ldots, \alpha_{i-1}\}\rangle_\rho \cap \Gamma_\rho \ .$$

<u>Proof</u>. From 2.4.1 we have $\|m([\alpha_i, \gamma])\| \leq 2.03 \|m(\gamma)\| \cdot \|m(\alpha_i)\| < \|m(\alpha_i)\|$, hence $[\alpha_i, \gamma] \in \Gamma_\rho$ (3.1.2); moreover, since $\|m(\alpha_i)\|$ is minimal in $\Gamma_\rho \setminus \langle\{\alpha_1, \ldots, \alpha_{i-1}\}\rangle_\rho$ we also have $[\alpha_i, \gamma] \in \langle\{\alpha_1, \ldots, \alpha_{i-1}\}\rangle_\rho$.

3.1.5 <u>Proposition</u>. The number d of elements in a short basis (3.1.3) has the a priori bound

$$d \leq d(n) := \operatorname{int}(3.02)^{\frac{1}{2}n(n+1)} .$$

<u>Proof</u>. By construction $\|m(\alpha_i)\| \leq \|m(\alpha_{i+1})\| \leq 0.48$. Also, if $i < j$, then

$$(*) \qquad \|m(\alpha_j * \alpha_i^{-1})\| \geq \|m(\alpha_j)\| ,$$

since otherwise we get the contradiction $\alpha_j \in \langle\alpha_1, \ldots, \alpha_{j-1}\rangle_\rho$:

$$\|m(\alpha_j * \alpha_i^{-1})\| < \|m(\alpha_j)\| \text{ implies } \alpha_j * \alpha_i^{-1} \in \langle\alpha_1, \ldots, \alpha_{j-1}\rangle_\rho ,$$

hence $$(\alpha_j * \alpha_i^{-1}) * \alpha_i = \alpha_j * (\alpha_i^{-1} * \alpha_i) \qquad (2.2.5)$$

$$= \alpha_j \in \langle\alpha_1,\ldots,\alpha_{j-1}\rangle_\rho \qquad (3.1.1).$$

With $\tilde{d}(m(\alpha_j) \circ m(\alpha_i)^{-1}, m(\alpha_j * \alpha_i^{-1})) \le$

$$\Lambda^2|t(\alpha_j)| \cdot |t(\alpha_i)| \le \frac{1}{64}||m(\alpha_j)|| \cdot ||m(\alpha_i)|| \qquad (2.3.1,\ 7.3)$$

and $||m(\alpha_k)|| < \frac{1}{2}$ we get from $(*)$:

$$||m(\alpha_j) \circ m(\alpha_i)^{-1}|| \ge \max\{ ||m(\alpha_j)|| - \frac{1}{128}||m(\alpha_i)|| , ||m(\alpha_i)|| - \frac{1}{128}||m(\alpha_j)||\} .$$

Finally 7.6.2 shows that there are at most $d(n)$ euclidean motions which pairwise satisty these last inequalities.

3.1.6 Remark. One would like to use 3.1.4 and 3.1.5 to prove along the lines of 2.5.2 that all $d(n)$-fold commutators in Γ_ρ vanish. This is not yet possible since the inductive proof in 2.5.2 requires to use associativity in words the lengths of which are not yet under control. So far there were no strong restrictions on $\Lambda\rho$; the proof will finally succeed only for carefully chosen rather large ρ .

3.2 Proposition. (Relative denseness of loops with small rotational parts) Let $\eta \le 0.48$ and $m \ge 10$ be adjustable parameters. Put

$$L = 3 + 2\left(\frac{7}{\eta}\right)^{\dim SO(n)} , \quad w \ge w(n) := 2\cdot 14^{\dim SO(n)}$$

and assume curvature bounds

$$\text{(K)} \qquad 2wd(M)\cdot\Lambda \le \eta\cdot m^{-L-1} .$$

Then there exists a number $\rho_o = \rho_o(\eta,m,w)$ such that

(i) $2w\,d(M)\cdot m^4 \le \rho_o \le 2wd(M)\cdot m^L$ (hence $\Lambda\rho_o \le 0.06$ for 3.1.3).

(ii) For every $v \in T_pM$ with $|v| \leq \rho_o(1 - \frac{1}{m})$ there exists $\alpha \in \Gamma_{\rho_o}$ (3.1.2) such that

$$\|r(\alpha)\| \leq \eta , \quad |t(\alpha) - v| \leq \frac{1}{m-1}\rho_o .$$

Remark. For the final theorem one fixed choice of all adjustable parameters will be sufficient. However it explains the structure of the proof much better if we show how the parameters enter the arguments and fix them later. The constant $w(n)$ insures that ρ_o will be large enough for the proof of 3.6.2; we did not put that assumption into the relative denseness parameter m, since m does not enter the present arguments as a large parameter. Finally η will be a very small number such that for the purpose of the proof of 3.3.1 the rotational parts of the loops selected in 3.2 are negligible.

Proof. With 2.1.3 we find for any ρ satisfying (i) a loop $\alpha' \in \pi_\rho$ such that $|t(\alpha') - v| \leq 2d(M)$; but $\|r(\alpha')\|$ may not be $\leq \eta$. We shall modify α' to $\alpha = \alpha' * (\alpha'')^{-1}$ such that $\|r(\alpha)\| \leq \eta$ and $|t(\alpha) - v| \leq \frac{1}{m-1}\rho$.

Define $\rho_i := 2w\, d(M)m^{i+2}$ $(i=1,\ldots,L-3)$. Then we have

(✱) One of the numbers ρ_i $(i=2,\ldots,L-3)$ has the following property: For each $\alpha' \in \pi_{\rho_i}$ there exists $\alpha'' \in \pi_{\rho_{i-1}}$ such that $d(r(\alpha'),r(\alpha'')) \leq \frac{2\pi}{7} \cdot \eta$.

Now 3.2 follows from (✱) if we choose ρ_o to be that ρ_i which satisfies (✱) and modify the above α' (For which $|t(\alpha')-v| \leq 2d(M)$) to $\alpha = \alpha' * (\alpha'')^{-1}$; since then

$$|t(\alpha)-v| \leq |t(\alpha')-v| + |t(\alpha'')| \leq 2d(M) + \rho_{i-1} \leq \frac{1}{m-1}\rho_i ,$$

$$\|r(\alpha)\| \leq \|r(\alpha') \circ r(\alpha'')^{-1}\| + d(r(\alpha') \circ r(\alpha'')^{-1} , r(\alpha' * (\alpha'')^{-1}))$$

$$\leq \frac{2\pi}{7} \cdot \eta + \Lambda^2\rho_i\rho_{i-1} \leq \eta \qquad \text{(with 2.3 and 3.2 (K)).}$$

Finally, assume that (✱) is false. Then there exist loops $\alpha_i \in \pi_{\rho_i}$

($i=1,\dots,L-3$) with $d(r(\alpha_i),r(\alpha_j)) > \frac{2\pi}{7}\eta$ ($i \neq j$) , since for each i there is at least one loop α_i whose rotational part has a distance $> \frac{2\pi}{7}\eta$ from all rotational parts in $\pi_{\rho_{i-1}}$. Because of 7.6.1 (i) there are less than $2(\frac{7}{\eta})^{\dim SO(n)}$ elements in $O(n)$ with pairwise distance $> \frac{2\pi}{7}\eta$ and L was defined large enough to produce a contradiction, which proves (✱) .

3.3 Small rotational parts and trivial d-fold commutators.

Recall that we are proceeding to prove that the loops in a suitable Γ_ρ have a much smaller rotational part than 0.48 , the bound in definition 3.1.2 . Under an additional assumption (✱) on iterated commutators - which will be removed in 3.4 - we achieve this in the following

3.3.1 Key proposition.

Let $0 < \Theta < \frac{1}{7}$ be an adjustable parameter. Choose the η in 3.2 as $\eta = \frac{1}{3}\Theta\ 2.1^{-d(n)}$ with $d(n) = \text{int } 3.02^{\frac{1}{2}n(n+1)}$ from 3.1.5 . Assume the curvature bounds 3.2 (K) and choose $\rho = \rho_o$ from 3.2 . Then we have:

Each $\gamma \in \Gamma_\rho$ which satisfies

(✱) The d-fold ($d \leq d(n)$) iterated commutator $[\dots[\alpha,\gamma],\dots,\gamma]$ exists and is trivial for all $\alpha \in \Gamma_\rho$

has a much smaller rotational part than 0.48 , namely

$$\|r(\gamma)\| \leq \Theta\ 2.1^{d-d(n)} .$$

Proof. Let $m(\gamma)(x) = Cx + x$ (2.3) be the holonomy motion of γ . Let $\vartheta := \|C\| \leq 0.48$ be the largest rotational angle of C and decompose $T_pM = E \oplus E^\perp$ such that E is a 2-plane and $C|_E$ is a rotation through the angle ϑ . Pick $v \in E$, $|v| = \frac{3}{4}\rho$ and choose with 3.2 a loop α such that $\|r(\alpha)\| \leq \eta$ and $|t(\alpha)-v| \leq \frac{1}{m-1}\rho$. If ϑ is larger than claimed in 3.3.1 we can derive a positive lower bound for the translational part of the d-fold commutator $[\dots[\alpha,\gamma],\gamma\dots,\gamma]$, which contradicts 3.3.1 (✱) .

Assume first that one can even find a loop $\alpha \in \Gamma_\rho$ with $r(\alpha) = \text{id}$ and

$0 \neq t(\alpha) \notin E$ - instead of what can be achieved with 3.2; this oversimplified situation already explains why the proof works. Disregard also the homotopy errors for the moment, i.e. compute the iterated commutator in the group of motions:

$$[\ldots[r(\alpha),r(\gamma)],\ldots,r(\gamma)] = \mathrm{id}$$

$$t([\ldots[m(\alpha),m(\gamma)],\ldots,m(\gamma)]) \in E \ , \ |t(\ldots\text{d-fold}\ldots)| = (2\sin\tfrac{\vartheta}{2})^{d} \cdot |t(\alpha)|.$$

Clearly, iterated commutators do not vanish unless $\vartheta = 0$.

To explain more clearly how the a priori bound 3.1.5 and the control of homotopy errors with 2.4 enter the proof, we once more disregard homotopy errors. The result is then applicable to the Bieberbach case: Because of $\Lambda = 0$ there are no homotopy errors and the curvature assumtion 3.2 (K) is trivially satisfied for any η ; therefore the conclusion of the key proposition holds with arbitrarily small Θ , i.e. $r(\gamma) = \mathrm{id}$ and $m(\gamma)$ is a pure translation. Since in this flat case we multiply loops as in the fundamental group (2.2.4 (ii)) one does not have to wait until 3.4 to remove the extra condition 3.3.1 (✱) but can use induction based on 3.1.4, 3.1.5 immediately to obtain the Bieberbach theorem from 3.2.

The following inductive inequality occurs several times in the present proof:

(U) If $\quad -y\cdot\mu^k \leq x_{k+1} - \lambda x_k \leq y\mu^k \quad (0 \leq 2\lambda \leq \mu \ , \ 0 \leq y)$

then $\quad -2y\mu^k \leq x_{k+1} - \lambda^k x_1 \leq y\mu^k(1 + \frac{\lambda}{\mu} + \ldots + (\frac{\lambda}{\mu})^{k-1}) \leq 2y\mu^k$.

We now estimate the translational parts of the iterated commutators $\alpha_1 = \alpha \ , \ \alpha_k = [\alpha_{k-1} \ , \ \gamma]$ assuming $\Lambda = 0$. Abbreviate

$$m(\alpha_k)(x) = A_k \cdot x + a_k \ , \ m(\gamma)(x) = C \cdot x + c$$

and compute in the group of motions

$$a_{k+1} = t([m(\alpha_k), m(\gamma)])$$

$$= (\mathrm{id} - C)\cdot a_k + (\mathrm{id} - [A_k, C])Ca_k + A_k C(\mathrm{id} - A_k^{-1})C^{-1}c \quad .$$

Since we know the action of C on E we decompose vectors $X \in T_pM$ as $X = X^E + X^{\perp}$ and find

$$|a_{k+1}^E| \geq 2 \sin \frac{\vartheta}{2} |a_k^E| - (\|A_{k+1}\| \cdot |a_k| + \|A_k\| \cdot |c|) \ .$$

We apply 2.4.3 to obtain the bounds

$$\|A_{k+1}\| \leq 2\vartheta \cdot \|A_k\| \leq (2\vartheta)^k \cdot \|A_1\| \quad ,$$

$$|a_{k+1}| \leq \|A_k\| \cdot |c| + \|C\| \cdot |a_k| \underset{(U)}{\leq} \vartheta^k |a_1| + (2\vartheta)^{k-1}\, 2\, \|A_1\| \cdot |c| \leq \rho,$$

which are inserted in the previous estimate:

$$|a_{k+1}^E| \geq 2 \sin \frac{\vartheta}{2} |a_k^E| - 2\rho \|A_1\| \cdot (2\vartheta)^{k-1} \quad .$$

Once again (U) implies (with $\|A_1\| \leq \eta$ and $|a_1^E| \geq \frac{1}{2}\rho$)

$$|a_d^E| \geq (2 \sin \frac{\vartheta}{2})^{d-1} \cdot |a_1^E| - 4\rho \|A_1\| \cdot (2\vartheta)^{d-2}$$

$$\geq \rho \frac{1}{2\vartheta} (2 \sin \frac{\vartheta}{2})^{d-1} \cdot (\vartheta - \Theta\ (2.1)^{d-d(n)}) \ .$$

Unless $\vartheta \leq \Theta\ (2.1)^{d-d(n)}$ this gives the contradiction $|a_d^E| > 0$. (It is clear, that this argument is useless unless one can establish 3.3.1 (*) for $d \leq d(n)$.)

Finally, the inclusion of the homotopy errors changes only some constants in the above computation, provided one works with a suitable distance function (7.3): It will be necessary that translational parts of length $\leq \rho$ are less important than rotational parts of size η , therefore we take

3.3.2 $$\|m(\gamma)\| := \max(\|r(\gamma)\| , \frac{\eta}{\rho} \cdot |t(\gamma)|) \ .$$

To have for the use in 2.4 the parameter $c = \frac{\eta}{\rho\Lambda} \geq 8$ we require the curvature assumption $8\Lambda\rho \leq \eta$, which is implied by 3.2 (K). Note that the use of this distance for the present computation does not interfere with the one used in 3.1.3 for the selection of the short basis.

For the neglected homotopy errors we have from 2.4.1 and 2.4.2 $(a_k = t(\alpha_k))$

$$|t(\alpha_{k+1}) - t([m(\alpha_k),m(\gamma)])| \leq$$

$$2(|t(\alpha_k)| + |t(\gamma)|) \cdot 0.21\Lambda(\|m(\alpha_k)\| \cdot |t(\gamma)| + \|m(\gamma)\| \cdot |t(\alpha_k)|) \; .$$

As in the $\Lambda = 0$ computation we need bounds obtained from 2.4 for $\|m(\alpha_{k+1})\|$:

$$\|r(\alpha_{k+1})\| \leq \|m(\alpha_{k+1})\| \leq 2.03\,\|m(\gamma)\| \cdot \|m(\alpha_k)\|$$

$$\leq (2.03\,\|m(\gamma)\|)^k \cdot \|m(\alpha_1)\| \leq \|m(\alpha_1)\| \; ,$$

$$|t(\alpha_{k+1})| \leq 1.006(\|m(\alpha_k)\| \cdot |t(\gamma)| + \|m(\gamma)\| \cdot |t(\alpha_k)|)$$

$$\leq (2.03\,\|m(\gamma)\|)^{k-1} \cdot \|m(\alpha_1)\| \cdot 2.012|t(\gamma)| + (1.006\,\|m(\gamma)\|)^k \cdot |t(\alpha_1)|$$

$$\overset{(U)}{\leq} (2.03\,\|m(\gamma)\|)^{k-1} \cdot \rho.$$

These bounds simplify the homotopy errors to

$$|t(\alpha_{k+1}) - t([m(\alpha_k),m(\gamma)])| \leq \Lambda\rho^2 \cdot (2.03\,\|m(\gamma)\|)^{k-1} \; .$$

As before we estimate the E-component of the translational parts (with $\|A_{k+1}\| \leq (2\vartheta)^k\,\|A_1\|$ replaced by $\|m(\alpha_{k+1})\| \leq (2.03\,\|m(\gamma)\|)^k \cdot \|m(\alpha_1)\|$) and find

$$|t(\alpha_{k+1})^E| \geq 2\sin\frac{\vartheta}{2}\,|t(\alpha_k)^E| - (2\rho\,\|m(\alpha_1)\| + \Lambda\rho^2) \cdot (2.03\,\|m(\gamma)\|)^{k-1} \; ;$$

once more using (U) we obtain

$$|t(\alpha_d)^E| \geq (2 \sin \frac{\vartheta}{2})^{d-1} \cdot |t(\alpha_1)^E| - 2\rho(2\,||m(\alpha_1)|| + \Lambda\rho) \cdot (2.03\,||m(\gamma)||)^{d-2}.$$

At last , if we assume contrary to 3.3.1 $\vartheta = ||r(\gamma)|| > \Theta\ 2.1^{d-d(n)}$ then the choice of our distance implies $||m(\gamma)|| = \vartheta$ and also $||m(\alpha_1)|| \leq \eta$ hence (with $2.03\vartheta \leq 4.2 \sin \frac{\vartheta}{2}$, $\eta = \frac{\Theta}{3} 2.1^{-d(n)}$)

$$|t(\alpha_d)^E| \geq \frac{\rho}{2\vartheta}(2 \sin \frac{\vartheta}{2})^{d-1}(\vartheta - 3\eta \cdot 2.1^d) > 0 .$$

This contradiction to 3.3.1 (✱) proves $||r(\gamma)|| \leq \Theta\ 2.1^{d-d(n)}$.

3.4 Nilpotency of Γ_ρ

The undesirable extra assumption 3.3.1 (✱) on the vanishing of iterated commutators in the key proposition will now be removed.

3.4.1 Proposition. (Nilpotency of Γ_{ρ_o})
Assume the curvature bounds 3.2 (K) with η from 3.3.1. Choose $\rho = \rho_o$ from 3.2. Let $\{\alpha_1,\ldots,\alpha_d\}$ be the short basis from 3.1.3. Then

(i) $\langle \Gamma_\rho \rangle_\rho = \Gamma_\rho$ (see 3.1.2).

(ii) All $d(n)$-fold commutators in Γ_ρ exist and are trivial; moreover

$$[\langle\{\alpha_1,\ldots,\alpha_{j+1}\}\rangle_\rho, \Gamma_\rho] \subset \langle\{\alpha_1,\ldots,\alpha_j\}\rangle_\rho .$$

The key proposition together with 3.4.1 has the immediate

3.4.2 Corollary. (Small rotational parts are very small)
Assumptions as in 3.4.1. All loops $\alpha \in \pi_\rho$ satisfy

if $||r(\alpha)|| \leq 0.48$ then $||r(\alpha)|| \leq \Theta$.

Proof of 3.4.1 by length controlled induction. We use the distance function 3.1.3 on the group of motions if the size of certain loops is compared to the size of the elements of the short basis; we use 3.3.2 for estimates of translational parts.

(i) Because of 3.1.4 we have for all α_j from the short basis and for all $\beta \in \Gamma_\rho$

$$[\beta,\alpha_j^{\pm 1}],[\alpha_j^{\pm 1},\beta] \in \langle\{\alpha_1,\dots,\alpha_{j-1}\}\rangle_\rho \cap \Gamma_\rho \ =: \ \Gamma^{(j-1)} .$$

(ii) The first element α_1 of the short basis 3.1.3 commutes with all $\gamma \in \Gamma_\rho$ since $\|m([\alpha_1,\gamma])\| < \|m(\alpha_1)\|$ (2.4.2). Therefore all powers $\alpha_1^k \in \langle\alpha_1\rangle_\rho$ commute with all $\gamma \in \Gamma_\rho$, in particular one can apply the key proposition 3.3.1 to all powers $\alpha_1^k \in \langle\alpha_1\rangle_\rho \cap \Gamma_\rho$. We conclude $\|r(\alpha_1^k)\| \le \Theta$ and therefore $\|r(\alpha_1^{k+1})\| \le 3\Theta$ (2.3.1); hence, if $\alpha_1^{k+1} \in \langle\alpha_1\rangle_\rho$ then $\alpha_1^{k+1} \in \langle\alpha_1\rangle_\rho \cap \Gamma_\rho$. This proves

$$\Gamma^{(1)} = \langle\alpha_1\rangle_\rho , \ [\Gamma^{(1)},\Gamma_\rho] = \{1\} , \ \alpha \in \Gamma^{(1)} \Rightarrow \|r(\alpha)\| \le \Theta .$$

(iii) Assume now for all $i \le j$

$$(H_j) \quad \langle\{\alpha_1,\dots,\alpha_i\}\rangle_\rho = \Gamma^{(i)} , \ \|\mathrm{rot}|_{\Gamma^{(i)}}\| \le \Theta , \ [\Gamma_\rho,\Gamma^{(i)}] \subset \Gamma^{(i-1)} .$$

We proved (H_1) in (ii) and we will conclude (H_{j+1}) using the inductive definition 3.1.1. Take $\alpha,\beta \in \langle\{\alpha_1,\dots,\alpha_{j+1}\}\rangle_\rho \cap \Gamma_\rho$ and assume

(a) $\qquad [\gamma,\alpha] , [\gamma,\beta] \in \Gamma^{(j)}$ for all $\gamma \in \Gamma_\rho$,

(b) $\qquad \|r(\alpha)\| , \|r(\beta)\| \le \Theta$.

Note that the generators $\alpha_1,\dots,\alpha_{j+1}$ and their inverses indeed satisfy (a) and (b) : (a) holds because of part (i) above; then (a) and (H_j) imply that assumption (*) of 3.3.1 is satisfied, therefore also (b) follows. To prove (H_{j+1}) inductively following 3.1.1 we assume $|t(\alpha * \beta)| \le \rho$ and have to show (a) and (b) for $\alpha * \beta$, which implies $\alpha * \beta \in \langle\{\alpha_1,\dots,\alpha_{j+1}\}\rangle_\rho \cap \Gamma_\rho = \Gamma^{(j+1)}$. (Inverses cause no problems since $\langle A\rangle_\rho = \langle A\rangle_\rho^{-1}$ and $\Gamma_\rho = \Gamma_\rho^{-1}$.)

Our curvature assumptions are such that we have associativity (2.2.5) certainly for products of 18 factors of length $\le \rho$, therefore

$$[\gamma,\alpha * \beta] = [\gamma,\alpha] * ([\alpha,[\gamma,\beta]] * [\gamma,\beta]) .$$

Each of the three factors on the right is already in $\Gamma^{(j)}$ because of (a). To see that the product is in $\Gamma^{(j)}$ we first estimate the lengths of the translational parts with 2.4.2 and 3.3.2, assuming (H_j) .

$$|t([\gamma,\beta])| \le 1.006 \, (\|m(\beta)\| \cdot |t(\gamma)| + \|m(\gamma)\| \cdot |t(\beta)|)$$

$$\le 1.006 \, (\Theta \rho + 0.48 \rho) \le 0.63 \rho ,$$

$$|t([\alpha,[\gamma,\beta]])| \le 1.006 \, (\Theta \rho + \Theta \rho) \le 0.29 \rho ,$$

therefore (2.3.1, 3.1.3 (K))

$$|t([\alpha,[\gamma,\beta]] * [\gamma,\beta])| \le \rho ,$$

$$\|r([\alpha,[\gamma,\beta]] * [\gamma,\beta])\| \le 3 \Theta .$$

This gives first $[\alpha,[\gamma,\beta]] * [\gamma,\beta] \in \Gamma^{(j)}$ and then $[\gamma,\alpha * \beta] \in \Gamma^{(j)}$, which is (a). As in (ii) this implies (*) in 3.3.1 and therefore $\|r(\alpha * \beta)\| \le \Theta$. This proves (b) and completes the inductive proof of (H_{j+1}) , but (H_d) is 3.4.1.

Note that in this proof no bound on the algebraic word length of elements of Γ_ρ (as products of the α_i) is needed or obtained.

3.5 Proposition. (Almost translational behaviour of Γ_ρ)
Assume the curvature bounds 3.2 (K), 3.3.1; choose $\rho = \rho_0$ from 3.2. For loops $\alpha,\beta \in \Gamma_\rho$ with $|t(\alpha)|,|t(\beta)| \le \frac{1}{3}\rho$ we have

(i) $$\|r(\alpha)\| \le 1.6 \, \frac{\Theta}{\rho} \, |t(\alpha)| ,$$

(ii) $$|t(\alpha * \beta) - t(\alpha) - t(\beta)| \le 2 \cdot \frac{\Theta}{\rho} \, |t(\alpha)| \cdot |t(\beta)| ,$$

(iii) $$|t([\beta,\alpha])| \le 4 \cdot \frac{\Theta}{\rho} \, |t(\alpha)| \cdot |t(\beta)| .$$

Proof. For each $k \leq \rho \cdot |t(\alpha)|^{-1}$ we have $|t(\alpha^k)| \leq \rho$ hence $\|r(\alpha^k)\| \leq \Theta$ (3.4). Then 2.3.3 implies $\Theta \geq \|r(\alpha^k)\| \geq k \cdot \|r(\alpha)\| - \frac{1}{2} k^2 \Lambda^2 |t(\alpha)|^2$, or $\|r(\alpha)\| \leq \frac{1}{k} \cdot \Theta + \frac{1}{2} k\Lambda\ |t(\alpha)|^2$.

Use this for $k_{max} \geq \rho \cdot |t(\alpha)|^{-1} - 1$ and note $\Lambda^2 \rho^2 \leq (\eta/m)^2 << \Theta$ (3.3.1 and 3.2 (K)), then (i) follows:

$$\|r(\alpha)\| \leq \left(\frac{\Theta}{\rho - t(\alpha)} + \frac{1}{2} \Lambda^2 \rho \right) \cdot |t(\alpha)| \leq 1.6 \frac{\Theta}{\rho} \cdot |t(\alpha)| .$$

Now (ii) and (iii) follow with (i) and $\Lambda^2 \rho^2 << \Theta$ from 2.3.1 and 2.4.1, namely

$$|t(\alpha * \beta) - t(\alpha) - t(\beta)| \leq \Lambda^2 |t(\alpha)| \cdot |t(\beta)| \cdot (|t(\alpha)| + |t(\beta)|) + \|r(\alpha)\| \cdot |t(\beta)|$$

$$\leq (1.6\ \Theta + \Lambda^2 \rho^2) \cdot \frac{1}{\rho} |t(\alpha)| \cdot |t(\beta)| ,$$

$$|t([\beta,\alpha])| \leq \frac{\sinh 2\Lambda\rho}{2\Lambda\rho} (3.2\ \Theta + \frac{4}{3} \Lambda^2 \rho^2) \cdot \frac{1}{\rho} \cdot |t(\alpha)| \cdot |t(\beta)| .$$

3.5.1 Remark. Note that in proposition 3.4 and 3.5 the statements about rotational parts and translational parts are once again separated. The distances 3.1.3 and 3.3.2 on the group of motions and the short basis 3.1.3 have served their technical purpose and will not appear again.

We conclude this chapter with a rather immediate application of 3.4 and 3.5 which shows already the far reaching consequences of these technical propositions. The main application is the development of still more precise information about the multiplication of loops in chapter 4.

3.6 The short loops modulo the almost translational ones.

3.6.1 Definition. (Equivalent loops)
Assumptions as in 3.4, 3.5. Call loops $\alpha, \beta \in \pi_{\rho/2}$ equivalent mod Γ_ρ ($\alpha \sim \beta$ mod Γ_ρ) if $\alpha * \beta^{-1} \in \Gamma_\rho$.

Justification of the definition: Take $\alpha, \beta, \gamma \in \pi_{\rho/2}$ such that $\alpha \sim \beta$, $\beta \sim \gamma$ mod Γ_ρ. First, $\Gamma_\rho = \Gamma_\rho^{-1}$ implies $\beta \sim \alpha$. Next, $\Lambda\rho$ is small enough to have $(\alpha * \beta^{-1}) * (\beta * \gamma^{-1}) = \alpha * \gamma^{-1}$ (associativity 2.2.5); therefore we obtain from 2.3 and 3.4.2 $\|r(\alpha * \gamma^{-1})\| \leq 2\,\Theta + \Lambda^2\rho^2 < 0.48$, hence $\alpha * \gamma^{-1} \in \Gamma_\rho$. This proves that we have indeed defined an equivalence relation.- Note how essential 3.4.2 enters: If rotational parts ≤ 0.48 were not automatically $\leq \Theta$, there would be no equivalence relation.

3.6.2 Proposition. (Description of equivalence classes)
Assumptions as in 3.4, 3.5, 3.6.1.

(i) If $\alpha,\beta \in \pi_{\rho/2}$ satisfy $d(r(\alpha),r(\beta)) \leq 0.47$ $(\leq 0.48 - \Lambda^2\rho^2)$

then $\alpha \sim \beta \bmod \Gamma_\rho$.

(ii) If $\alpha \sim \beta \bmod \Gamma_\rho$ then $d(r(\alpha),r(\beta)) \leq \Theta + \Lambda^2\rho^2$.

(iii) There are at most $w(n) = 2\cdot 14^{\dim SO(n)}$ equivalence classes mod Γ_ρ .

(iv) In each equivalence class is a loop of length $\leq 2w(n)d(M)$; this is a short representative since $2w(n)d(M) \leq 2m^{-4}\cdot\frac{\rho}{2}$.

Proof. If $d(r(\alpha),r(\beta)) \leq 0.47$ then 2.3 implies $\|r(\alpha * \beta^{-1})\| \leq 0.48$, hence $\alpha * \beta^{-1} \in \Gamma_\rho$ which is (i) . If $\alpha = \gamma * \beta$, $\gamma \in \Gamma_\rho$ then 2.3 implies $d(r(\alpha),r(\beta)) \leq \|r(\gamma)\| + \Lambda^2\rho^2$, therefore 3.4.2 gives (ii) . (iii) follows from (i) and 7.6.1 since there are at most $w(n)$ elements in $O(n)$ with pairwise distance $\geq \frac{\pi}{7} \approx 0.45$.

To see (iv) note first that 2.1.5 applied to any loop $\alpha \in \pi_\rho$ gives a decomposition $\alpha = \alpha_1 \ldots \alpha_k$ into a product of loops $\alpha_i \in \pi_{2d(M)}$ such that any partial product of consecutive factors $\alpha_j * \alpha_{j+1} \cdot\cdot * \alpha_{j+\ell}$ has length $\leq |t(\alpha)|$. Therefore we take $\pi_{2d(M)}$ as a set of generators for $\pi_{\rho/2}$ and choose in each of the $(\leq w(n))$ equivalence classes a representative which as a word in the $\alpha_i \in \pi_{2d(M)}$ has minimal word length. Since all words of word length $\leq w(n)$ have a loop length $\leq 2w(n)d(M) \ll \rho/2$ we can use (iii) and the pigeon hole argument to find a number $w \leq w(n)$ such that all words of word length $= w+1$ are equivalent to words of word length $\leq w$.

Assume by induction that each loop $\alpha \in \pi_{\rho/2}$ which can be written as a word of word length $= \ell \geq w+1$ has a representative $\omega \sim \alpha$ of word length $\leq w$; then each $\alpha * \alpha_i \in \pi_{\rho/2}$ $(\alpha_i \in \pi_{2d(M)})$ of word length $\ell+1$ is equivalent to

$\omega * \alpha_i$, which is a word of length $\leq w+1$ and therefore again equivalent to a word of length $\leq w$.

3.6.3 Corollary. (Relative denseness of Γ_ρ)
Assumptions as in 3.4, 3.5, 3.6.1.
With $\sigma = 3w(n)d(M)$ $(<< \frac{1}{m-1}\rho$, compare 3.2) we have: The translational parts of loops $\gamma \in \Gamma_\rho$ are σ-dense (2.1.4) in the ball of radius $\rho/2 - \sigma$ in T_pM.
Recall $\|r(\gamma)\| \leq \Theta$ (3.4.2).

Proof. To each $v \in T_pM$ with $|v| \leq \rho/2 - \sigma$ we find with 2.1.3 a loop $\alpha \in \pi_{\rho/2}$ such that $|v - t(\alpha)| \leq 2d(M)$. Then take a shortest representative $\omega \sim \alpha$ (3.6.2 (iv)), i.e. $|t(\omega)| \leq 2w(n)d(M)$. Now $\alpha * \omega^{-1} \in \Gamma_\rho$ and (with 2.3.1) $|t(\alpha * \omega^{-1}) - t(\alpha)| \leq |t(\omega)| + \Lambda^2\rho^2|t(\omega)|$, hence $|v - t(\alpha * \omega^{-1})| \leq \sigma$.

3.6.4 Corollary. (Multiplication of equivalence classes)

(i) Γ_ρ has the following property of a normal subgroup:
If $\alpha \in \pi_{\rho/3}$, $\gamma \in \Gamma_{\rho/3}$ then $\alpha * \gamma * \alpha^{-1} \in \Gamma_\rho$.

(ii) If $\omega_i \in [\alpha_i]$ (i=1,2) are short representatives of equivalence classes mod Γ_ρ (3.6.2 (iv)) then

$$[\alpha_1] * [\alpha_2] \quad := \quad [\omega_1 * \omega_2]$$

is a well defined product.

(iii) The equivalence classes mod Γ_ρ with the product (ii) form a group G of order $|G| \leq w(n)$ which is isomorphic to a subgroup $\bar{G}$ of $O(n)$. If $A \in \bar{G}$ and $\|A\| \leq 0.47 - 3\Theta$ then $A = id$.

Proof. From 2.3.1 (i) we have $d\big(r(\alpha * \gamma * \alpha^{-1}),\ r(\alpha) \circ r(\gamma) \circ r(\alpha^{-1})\big) \leq \Lambda^2\rho^2$, hence $\|r(\alpha * \gamma * \alpha^{-1})\| \leq \|r(\gamma)\| + \Lambda^2\rho^2 < 0.48$ and $|t(\alpha * \gamma * \alpha^{-1})| \leq 2|t(\alpha)| + |t(\gamma)| \leq \rho$, which proves (i) .

To see (ii) note first that $|t(\omega_1)| + |t(\omega_2)| < \rho/2$ so that $[\omega_1 * \omega_2]$ is defined; if $\omega_i' \sim \omega_i$ mod Γ_ρ are other short representatives, $\omega_i' = \gamma_i * \omega_i$, then 2.2.5 gives $(\omega_1' * \omega_2') * (\omega_1 * \omega_2)^{-1} = \gamma_1 * (\omega_1 * \gamma_2 * \omega_1^{-1})$, hence

$$\|r((\omega_1' * \omega_2') * (\omega_1 * \omega_2)^{-1})\| \leq \|r(\gamma_1)\| + \|r(\omega_1 * \gamma_2 * \omega_1^{-1})\| + \Lambda^2\rho^2 \quad (2.3)$$

$$\leq 2\,\Theta + \Lambda^2\rho^2 \leq 0.48 \quad (3.4.2,\ 3.6.4\ (i)).$$

To prove (iii) note first that $\omega * \omega^{-1} = 1$ (2.2.5) implies $[\omega] * [\omega^{-1}] = |1|$. Associativity of the product follows from 3.6.2 (iv) and 2.2.5. Therefore we have a group G and 3.6.2 (iii) implies $|G| \leq w(n)$. To map G into $O(n)$ pick in each equivalence class one short representative ω and map $[\omega]$ to $h([\omega]) := r(\omega) \in O(n)$. This map $h : G \to O(n)$ is almost a homomorphism since 3.6.2 (ii) implies

$$d(h([\omega_1] * [\omega_2]),\ h([\omega_1]) \circ h([\omega_2])) \leq \Theta + \Lambda^2\rho^2 .$$

With 8.2 we change h slightly to a homeomorphism $\bar{h} : G \to \bar{G} \subset O(n)$; since $[\omega_1] \neq [\omega_2]$ implies $d(h([\omega_1]), h([\omega_2])) \geq 0.47$ (3.6.2 (i)) we also have from 8.2 $d(\bar{h}([\omega_1]), \bar{h}([\omega_2])) \geq 0.47 - 3\Theta$, which shows injectivity of $\bar{h}$ and proves the last statement of (iii) .

3.6.5 Outlook. We have now derived enough properties of the multiplication of short loops to describe the end of the proof of Gromov's theorem. From 3.5 we prove that a set of exactly n generators $\gamma_1,\dots,\gamma_n$ for Γ_ρ can be chosen such that the not too long loops of Γ_ρ have a unique representation $\gamma = \gamma_1^{l_1} * \dots * \gamma_n^{l_n}$ and such that $[\gamma,\gamma_{i+1}] \in \langle\{\gamma_1,\dots,\gamma_i\}\rangle$. The multiplication in Γ_ρ can be expressed by polynomials in the exponents $(l_1,\dots,l_n) \in \mathbb{Z}^n$; therefore Γ_ρ generates a uniform discrete subgroup Γ of an n-dimensional nilpotent Lie group N . Moreover Γ is isomorphic to a normal subgroup of the fundamental group with factor group G (3.6.4). In the universal covering $\tilde{M}$ we have the subset $\Gamma \cdot p$ which we map onto the subset $\Gamma \subset N$. With the aid of a suitable left invariant metric on N we interpolate this map to a Γ-equivariant diffeomorphism from $\tilde{M}$ to N , thus proving that ${}_{\Gamma}\backslash M$ is diffeomorphic to the nilmanifold ${}_{\Gamma}\backslash N$.

4. The embedding of Γ_ρ into a nilpotent torsionfree subgroup of finite index in $\pi_1(M)$.

This chapter achieves the first part of 3.6.5 in the following steps:

4.1 defines λ-normal bases in $\mathbb{R}^n$, shows that they exist for lattice subgroups of $\mathbb{R}^n$ and proves estimates which are familiar for orthonormal bases.

4.2 introduces almost translational subsets of $\mathbb{R}^n$ and gives bounds for the deviation of products from the purely translational behaviour.

4.3 studies in the almost translational case (but guided by 4.1) orbits, suitable representatives and their projections into a hyperplane.

4.4 proves that the projections (4.3) of almost translational sets are again almost translational.

4.5 uses the above to find a λ-normal basis $c_1,\ldots,c_n$ for an almost translational set T and to represent short elements $x \in T$ uniquely as $x = c_1^{l_1} \ldots c_n^{l_1}$ $(l_i \in \mathbb{Z})$.

4.6 embeds some ball in Γ_ρ (3.5) using the representation 4.5 into a torsionfree nilpotent group Γ generated by that ball; finally Γ is embedded as a subgroup of finite index into the fundamental group $\pi_1(M)$ with the help of 2.2.7 and 3.6.2.

Remark. Most of the more unpleasant portions of this chapter, notably (4.2.3), are devoted to length estimates in almost translational sets; they are crucial in 4.5 and 4.6, but they do not contribute much to the geometric idea of the proof.

4.1 Normal bases for lattices in $\mathbb{R}^n$

4.1.1 Definition. By induction over n we introduce λ-normal bases for $\mathbb{R}^n$ as follows:

(i) Any basis for $\mathbb{R}^1$ is λ-normal for each $\lambda \geq 1$.
(ii) A basis $\{\gamma_1,\ldots,\gamma_n\}$ for $\mathbb{R}^n$ is λ-normal if it satisfies:

$$|\gamma_i'| \le |\gamma_i| \le \lambda\cdot|\gamma_i'| \ , \text{ where } \ \gamma_i' := \gamma_i - \frac{\langle\gamma_i,\gamma_1\rangle}{\langle\gamma_1,\gamma_1\rangle}\cdot\gamma_1 \quad (i=2,\dots,n)$$

are the projections of γ_i into $\{\gamma_1\}^{\perp}$;

$\{\gamma_2',\dots,\gamma_n'\}$ is a λ-normal basis for $\{\gamma_1\}^{\perp} \cong \mathbb{R}^{n-1}$.

4.1.2 <u>Remarks.</u> The γ_i are pairwise orthogonal if and only if $\lambda = 1$. If $\{\gamma_1,\dots,\gamma_n\}$ is λ-normal for $\mathbb{R}^n$, then $\{\gamma_1,\dots,\gamma_k\}$ is λ-normal for $\mathbb{R}^k = \text{span}\{\gamma_1,\dots,\gamma_k\}$.

Again, we study this notion first in the purely translational case ($\Theta = 0$ in 3.5) and later modify the procedure to handle errors caused by $\Theta \neq 0$.

4.1.3 <u>Proposition.</u> (Normal basis of a lattice)
Any uniform discrete subgroup Γ of $\mathbb{R}^n$ has a $\sqrt{2}$-normal basis.

<u>Proof.</u> For $n = 1$ the statement is trivial; assume 4.1.3 by induction for $n-1$. Then choose $\gamma_1 \in \Gamma$ as a shortest element and define a uniform discrete subgroup Γ' of $\mathbb{R}^{n-1}$ by projecting the orbits $\{\gamma_1^i\cdot\gamma\}_{i\in\mathbb{Z}}$ (which are sets of points at distance $|\gamma_1|$ apart on lines parallel to $\mathbb{R}\cdot\gamma_1$) orthogonally onto $\{\gamma_1\}^{\perp} \cong \mathbb{R}^{n-1}$. Γ' is uniform, since for each $x \in \mathbb{R}^{n-1} \subset \mathbb{R}^n$ holds $d(x,\Gamma') \le d(x,\Gamma)$.

Now define for each orbit $\{\gamma_1^i\cdot\gamma\}$ a representative $\tilde\gamma$ by the inequalities

$$(*) \qquad \langle\gamma_1,\tilde\gamma\rangle > 0 \ , \ \langle\gamma_1,\tilde\gamma-\gamma_1\rangle \le 0 \ .$$

This definition of $\tilde\gamma$ and the minimality of γ_1 (here $|\gamma_1| \le |\tilde\gamma-\gamma_1|$) give $0 < \langle\tilde\gamma,\gamma_1\rangle \le |\gamma_1|^2$, $2\langle\tilde\gamma,\gamma_1\rangle \le |\tilde\gamma|^2$ hence $\cos^2 \sphericalangle(\gamma_1,\tilde\gamma) \le \frac{1}{2}$. This implies for the projection γ' of $\tilde\gamma$

$$(**) \qquad |\gamma'| \le |\tilde\gamma| \le \sqrt{2}\,|\gamma'| \ .$$

Now $(**)$ shows discreteness of Γ' ; then, by induction hypothesis, Γ' has a $\sqrt{2}$-normal basis $\{\gamma_2',\dots,\gamma_n'\}$ and, again with $(**)$, $\{\gamma_1,\tilde\gamma_2,\dots,\tilde\gamma_n\}$ is $\sqrt{2}$-normal for Γ .

For handling the error terms caused by $\Theta \neq 0$ the following is usefull:

4.1.4 Proposition. (Uniformity properties of λ-normal bases)

(i) $\det(\gamma_1,\ldots,\gamma_n) \geq \lambda^{-\frac{1}{2}n(n-1)} \cdot |\gamma_1| \cdot \ldots \cdot |\gamma_n|$

(ii) $|\sum_{i=1}^{j} x_i\gamma_i| \geq 2^{-\frac{1}{2}(j-1)} \cdot \lambda^{-\frac{1}{2}j(j-1)} \cdot \sum_{i=1}^{j} |x_i\gamma_i|$ $(j=1,\ldots,n)$

(iii) If $x' = \sum_{i=2}^{n} x_i\gamma_i'$ then $x = x_1\gamma_1 + \sum_{i=2}^{n} x_i\gamma_i$

with the same $x_i\,(i=2,\ldots,n)$ in both decompositions.

Note that (ii) allows the components of a vector to be much longer than the vector itself; we cannot conclude more than

$$|x_i\gamma_i| \leq 2^{+\frac{1}{2}(n-1)} \cdot \lambda^{\frac{1}{2}n(n-1)} \cdot |\sum_{i=1}^{n} x_i\gamma_i| \,.$$

Proof. (i) and (ii) are trivial for $n = 1$ and are assumed by induction to hold for $(n-1)$. Then

$$\begin{aligned} |\det{}_n(\gamma_1,\ldots,\gamma_n)| &= |\gamma_1| \cdot \det{}_{n-1}(\gamma_2',\ldots,\gamma_n') \\ &\geq |\gamma_1| \cdot \lambda^{-\frac{1}{2}(n-1)(n-2)} |\gamma_2'| \cdot \ldots \cdot |\gamma_n'| \qquad \text{(induction)} \\ &\geq \lambda^{-\frac{1}{2}n(n-1)} |\gamma_1| \cdot |\gamma_2| \cdot \ldots \cdot |\gamma_n| \qquad (\lambda\text{-normal}) \end{aligned}$$

gives (i) . To prove (ii) first observe: If $\phi = \sphericalangle(x,y) \in (0,\pi)$ then

$$\begin{aligned} |x+y|^2 &= x^2 - 2|x||y| \left(\cos^2\tfrac{\phi}{2} - \sin^2\tfrac{\phi}{2}\right) + |y|^2 \\ &\geq \sin^2\tfrac{\phi}{2}\,(|x| + |y|)^2 \,, \\ |x \pm y| &\geq \min\left(\sin\tfrac{\phi}{2}, \cos\tfrac{\phi}{2}\right)(|x| + |y|) \geq 2^{-\frac{1}{2}} \sin\phi \cdot (|x| + |y|). \end{aligned}$$

Next define $\phi_j := \sphericalangle(\gamma_j\,,\ \mathrm{span}(\gamma_1,\ldots,\gamma_{j-1})) \in [0,\frac{\pi}{2}]$ and obtain from the previous inequality inductively

$$|x_j\gamma_j + \sum_{i=1}^{j-1} x_i\gamma_i| \geq 2^{-\frac{1}{2}} \sin\phi_j(|x_j\gamma_j| + |\sum_{i=1}^{j-1} x_i\gamma_i|)$$

$$\geq 2^{-\frac{1}{2}(j-1)} \cdot \prod_{i=2}^{j} \sin\phi_i \cdot \sum_{i=1}^{j} |x_i\gamma_i| .$$

Finally, since $|\det(\gamma_1,\ldots,\gamma_j)| = \prod_{i=2}^{j} \sin\phi_i \cdot |\gamma_1|\cdot\ldots\cdot|\gamma_j|$ and since $\{\gamma_1,\ldots,\gamma_j\}$ is λ-normal in $\mathrm{span}(\gamma_1,\ldots,\gamma_j)$ (4.1.2) we can use 4.1.4 (i) to eliminate $\prod_{i=2}^{j} \sin\phi_i$ from the last inequality, proving (ii).

(iii) is clear since $\gamma_i - \gamma_i' \in \mathbb{R}\cdot\gamma_1$ $(i=2,\ldots,n)$.

4.2 Almost translational subsets of $\mathbb{R}^n$

Additional problems arise, if one tries to generalize 4.1 to the case $\Theta \neq 0$:

Orbits are not straight and therefore more difficult to project. Products are defined for sufficiently short elements only and the admissable length decreases in the inductive steps; also the almost translational behaviour deteriorates. After the first inductive step the elements to work with can no longer be considered as loops. Even after one has found a normal basis, the representation corresponding to 4.1.4 (iii) needs further arguments. The following notion helps to generalize the arguments from 4.1:

4.2.1 Definition. We call a finite set $T_\rho \subset \mathbb{R}^n$ a Θ-translational, σ-dense subset of radius ρ if it satisfies (with the parameters restricted to $\Theta \leq \frac{1}{50}$, $\sigma \leq \frac{1}{20}\rho$ for convenience):

(i) $0 \in T_\rho$; if $c \in T_\rho$ then $|c| \leq \rho$.

(ii) For all $x \in \mathbb{R}^n$, $|x| \leq \frac{1}{2}\rho$, there is some $c \in T_\rho$ such that $|x - c| \leq \sigma$.

(iii) For all $a,b \in T_\rho$, $|a + b| \leq \rho\cdot(1-\Theta)$ a product $a * b \in T_\rho$ is defined; for each $a \in T_\rho$, $|a| \leq \rho\cdot(1-\Theta)$ there exists a unique $a^{-1} \in T_\rho$ such that $a * a^{-1} = a^{-1} * a = 0$.

(iv) Associativity $(a * b) * c = a * (b * c)$ holds, if the existence of all products involved follows from (iii) .

(v) The product satisfies

$$|a * b - a - b| \le \frac{\Theta}{\rho} |a| \cdot |b| ,$$
$$|[a,b]| \le \frac{2\Theta}{\rho} |a| \cdot |b| .$$

4.2.2 Remark. Because of 3.5, 2.2.5, 3.6.3 we have with $\rho = \frac{1}{3}\rho_o$ that the set

$$T_\rho = \{a = t(\alpha); \alpha \in \Gamma_{\rho_o} (3.5), |t(\alpha)| \le \rho\}$$

together with the product $t(\alpha) * t(\beta) := t(\alpha * \beta)$ satisfies 4.2.1 (i) - (v). The induction starts from this example.

In order to control the deviation of products from vector sums we collect some immediate consequences of definition 4.2.1 in the following.

4.2.3 Proposition. Notations and assumptions as in 4.2.1. Then

(i) Shortest elements $0 \ne c_1 \in T_\rho$ satisfy $|c_1| \le 2\sigma$, $[c, c_1] = 0$, if $[c, c_1]$ is defined.

(ii) Let $a \in T_\rho$, $|a| \le (1-\Theta) \cdot \rho$, then a^{-1} satisfies
$|a^{-1} + a| \le \frac{\Theta}{\rho} |a^{-1}| \cdot |a| \le \frac{\Theta(1+\Theta)}{\rho} |a|^2$
$|a^{-1}| \le |a| \cdot (1 - \frac{\Theta}{\rho} |a|)^{-1} \le (1+\Theta)|a|$.

(iii) Let $a, b \in T_\rho$ with $|a|, |b|, |a-b| \le (1-3\Theta)(1-\Theta^2)^{-1} \cdot \rho$, then
$|a * b^{-1} - (a-b)| \le \frac{\Theta}{\rho} |b| \cdot |a * b^{-1}| \le \frac{\Theta(1+\Theta)}{\rho} |b| \cdot |a-b|$,
$|a - b| \le (1 + \frac{\Theta}{\rho} |b|) \cdot |a * b^{-1}|$,
$|a * b^{-1}| \le (1 - \frac{\Theta}{\rho} |b|)^{-1} \cdot |a-b| \le (1+\Theta)|a-b|$.

(iv) Let $a_1, \ldots, a_k \in T_\rho$, $\sum_{i=1}^k |a_i| \le (1-\Theta)\rho$. Then $a_1 * \ldots * a_k$ is defined, $|a_1 * \ldots * a_k| \le \sum_{i=1}^k |a_i| \cdot (1+\Theta)$ and

$$|a_1 * \ldots * a_k - (a_1 + \ldots + a_k)| \le \frac{\Theta(1+\Theta)}{\rho} \sum_{i<j}^{k} |a_i| \cdot |a_j| .$$

In particular, if $|ka| \le (1-3\Theta)\rho$, then $(a^{\pm 1})^j$ is defined for

$1 \leq j \leq k$ and

$$|a^j - j\cdot a| \leq \frac{\Theta(1+\Theta)}{\rho} \cdot \frac{j(j-1)}{2} |a|^2 , \quad (1-\Theta)|ja| \leq |a^j| \leq (1+\Theta)\cdot|ja| ,$$

$$|a^{-j} + ja| \leq \frac{\Theta(1+\Theta)^3}{\rho} \cdot \frac{j(j+1)}{2} |a|^2 , \quad (1-\Theta)|ja| \leq |a^{-j}| \leq (1+\Theta)\cdot|ja| .$$

(v) Let $c_1,\dots,c_n \in T_\rho$ be a λ-normal basis with $\lambda \leq 2$.
If $\sum_{i=1}^n |l_i c_i| \leq (1-2\Theta)^2\rho$ then $c_i^{l_i}$ and $c_1^{l_1} * \dots * c_n^{l_n}$ are defined and

$$|c_1^{l_1} * \dots * c_n^{l_n} - \sum_{i=1}^n l_i c_i| \leq \frac{\Theta(1+\Theta)^3}{\rho} \cdot (\sum_{i=1}^n |l_i c_i|)^2$$

$$\leq \frac{\Theta}{\rho} 2^{\frac{1}{2}n^2} \sum |l_i c_i| \cdot |\sum l_i c_i| \leq \frac{\Theta}{\rho} 2^{n^2} \cdot |\sum l_i c_i|^2 .$$

Therefore λ-normal bases are useful if $\sum_{i=1}^n |l_i c_i| \cdot 2^{\frac{1}{2}n^2} \leq \rho$:

$$|c_1^{l_1} * \dots * c_n^{l_n} - \sum_{i=1}^n l_i c_i| \leq \Theta|\sum l_i c_i| \leq \frac{\Theta}{1-\Theta} |c_1^{l_1} * \dots * c_n^{l_n}| .$$

<u>Proof</u>. (i) is clear from 4.2.1 (ii) and (v) . For (ii) use $a^{-1} * a = 0$ in 4.2.1 (v) . To see (iii) check first that the products $a * b^{-1}$ and $(a * b^{-1}) * b$ are defined:

$$|a + b^{-1}| \leq |a - b| + |b + b^{-1}| \leq (1-\Theta)\rho \quad \text{and}$$

$$|a * b^{-1} + b| \leq |a + b^{-1} + b| + \frac{\Theta}{\rho} |a|\cdot|b^{-1}| \leq |a| + \frac{\Theta}{\rho} (|a|+|b|)\cdot|b^{-1}| \leq (1-\Theta)\rho.$$

Therefore we have

$$|a * b^{-1} - (a-b)| = |a * b^{-1} * b - a * b^{-1} - b| \leq \frac{\Theta}{\rho} |a * b^{-1}|\cdot|b| ;$$

this inequality implies

$$(1 - \frac{\Theta}{\rho}|b|)\cdot|a * b^{-1}| \leq |a - b| \leq (1 + \frac{\Theta}{\rho}|b|)\cdot|a * b^{-1}|$$

and both estimates give the remaining inequalities of (iii) . (iv) is proved by induction, $k = 1$ being trivial: For $1 \leq m \leq k$ the product $(a_1 * \dots * a_m) * (a_{m+1} * \dots * a_{k+1})$ exists since

$$|a_1 * \ldots * a_m| + |a_{m+1} * \ldots * a_{k+1}| \underset{(iv)}{\leq} (1+\Theta)\cdot\left(\sum_{i=1}^{m} |a_i| + \sum_{i=m+1}^{k+1} |a_i|\right) \leq (1+\Theta)(1-2\Theta)\rho \leq (1-\Theta)\rho \; ;$$

moreover the product is independent of the brackets, since we have associativity from 4.2.1 (iv) . The estimate follows from

$$|(a_1 * \ldots * a_k) * a_{k+1} - \sum_{i=1}^{k+1} a_i| \leq$$

$$|a_1 * \ldots * a_k - \sum_{i=1}^{k} a_i| + \frac{\Theta}{\rho}|a_1 * \ldots * a_k|\cdot|a_{k+1}| \leq$$

$$\frac{\Theta}{\rho}(1+\Theta)\sum_{i\,j}^{k} |a_i||a_j| + \frac{\Theta}{\rho}(1+\Theta)\sum_{i=1}^{k} |a_i|\cdot|a_{k+1}| \, .$$

The application to a^j is direct; for a^{-1} observe $|j\cdot a^{-1}| \leq (1-2\Theta)\rho$ and

$$|a^{-j} + ja| \leq j\cdot|a + a^{-1}| + \frac{\Theta}{2\rho}(1+\Theta)\, j(j-1)\, |a^{-1}|^2$$

$$\underset{(ii)}{\leq} \frac{\Theta}{\rho}(1+\Theta)^3\cdot|ja|^2\cdot\frac{j+1}{2j}$$

$$\leq \Theta\cdot|ja| \qquad \text{(since } (1+\Theta)^3|ja| \leq \rho) \, .$$

(v) The powers and their product are defined by (iv) , from which also the first estimate follows:

$$|c_1^{l_1} * \ldots * c_n^{l_n} - \sum_{i=1}^{n} c_i^{l_i}| \leq \frac{\Theta(1+\Theta)}{\rho}\sum_{i\,j}^{n} |l_i c_i|\cdot|l_j c_j|\cdot(1+\Theta)^2 \, ,$$

$$\sum_{i=1}^{n} |c_i^{l_i} - l_i c_i| \leq \frac{\Theta(1+\Theta)^3}{\rho}\cdot\sum_{i=1}^{n} |l_i|\cdot\frac{|l_i|+1}{2}\cdot|c_i|^2$$

Finally use 4.1.4 (ii) with $\lambda \leq 2$.

4.3 Orbits, representatives, projection

4.3.1 Almost straight orbits.

Let T_ρ be as in 4.2.1 and let $0 \neq c_1 \in T_\rho$ be a shortest element. For each $c \in T_\rho$, $|c| \leq (1-3\Theta)\rho - 2\sigma$, the products $c_1^{\pm 1} * c$ are defined and $[c, c_1] = 0$ (4.2.3). To follow the orbit through c towards the hyperplane $\{c_1\}^{\perp}$ consider the difference vector between consecutive orbit points:

$$w^- = c_1^{-1} * c - c , \quad \text{if} \quad \langle c,c_1\rangle > 0 ,$$

$$w^+ = c_1 * c - c , \quad \text{if} \quad \langle c,c_1\rangle \leq 0 ,$$

and compare them with $\mp c_1$ using 4.2.3 (iii) :

(i) $$|w^+ - c_1| \leq \frac{\Theta}{\rho}\cdot |c|\cdot|c_1| , \quad |w^- + c_1| \leq \frac{\Theta(1+\Theta)}{\rho} |c-c_1|\cdot|c_1| .$$

Therefore

(ii) $$|\sin \measuredangle (w^+,c_1)| \leq \frac{\Theta}{\rho}\cdot|c| \leq \Theta := \sin\vartheta ,$$

$$|\sin \measuredangle (w^-,-c_1)| \leq \frac{\Theta}{\rho}(1+\Theta)|c-c_1| \leq \Theta = \sin\vartheta .$$

The second angle estimate is independent of c so that the angle between c_1 and the difference vector of <u>any two orbit points</u> is $\leq \vartheta$. In particular we obtain that the lengths of the vectors of the orbit through c towards the hyperplane $\{c_1\}^{\perp}$ do not exceed $|c|\cdot\frac{1}{\cos\vartheta} \leq ((1-3\Theta)\rho - 2\sigma)(1-\Theta^2)^{-\frac{1}{2}}$, which is enough for the above application of 4.2.3 (iii) . Therefore 4.3.1 (i) , (ii) hold along the orbit through c towards the hyperplane. We conclude now from (i) that the scalar product with c_1 changes along the orbit almost as in the $\Theta = 0$ case, namely

(iii) $$(1-\Theta)\cdot|c_1|^2 \leq \pm\langle c_1,w^{\pm}\rangle \leq (1+\Theta)\cdot|c_1|^2 .$$

4.3.2 <u>Representatives of orbits.</u>
Consider $c \in T_\rho$ such that $|c| \leq (1-4\Theta)\rho - 2\sigma$; then 4.3.1 (i) - (iii) show that we find on the c_1-orbit through c after $|k|$ steps, where $\max(-k,k+1) \leq 1 + |\langle c_1,c\rangle| \cdot (1-\Theta)^{-1}|c_1|^{-2}$, a <u>unique representative</u> $\tilde{c}$ which is characterized by the same inequalitties as in 4.1:

(*) $$\langle c_1,\tilde{c}\rangle > 0 , \quad \langle c_1,c_1^{-1} * \tilde{c}\rangle \leq 0 , \quad c = c_1^k * \tilde{c} .$$

Moreover, with 4.2.3 and 4.3.1 we have the estimates

(**) $$(1-\Theta)|kc_1| \leq |c_1^k| \leq (1+\Theta)|kc_1| ,$$

$$|kc_1| \leq |c|\cdot(1-\Theta)^{-1} + 2\sigma \leq (1-3\Theta)\rho ,$$

$$0 < \langle c_1, \tilde{c} \rangle \leq (1+\Theta)|c_1|^2 \quad , \quad |\tilde{c}| \cdot \cos \vartheta \leq |c| \ ,$$

$$|\sin \measuredangle (c-\tilde{c}, c_1)| \leq \frac{\Theta}{\cos \vartheta} \cdot \frac{|c|}{\rho} \leq \sin \vartheta \ .$$

4.3.3 <u>Projection into</u> $\{c_1\}^{\perp}$.

Assumptions as in 4.3.1, 4.3.2. We map the set $\tilde{T}_\rho$ of representatives by orthogonal projection onto

$$T' \subset \{c_1\}^{\perp} \ , \ \tilde{c} \to c' = \tilde{c} - \frac{\langle \tilde{c}, c_1 \rangle}{\langle c_1, c_1 \rangle} \cdot c_1 \ .$$

We claim (compare 4.1)

$$|c'| \leq |\tilde{c}| \leq \lambda \cdot |c'| \ , \quad \text{where} \quad \lambda \leq \left(\frac{1}{2} - 2\Theta\right)^{-\frac{1}{2}} ,$$

and for different representatives $\tilde{c} \neq \tilde{d} \in \tilde{T}_\rho$

$$\cos \measuredangle (\tilde{c}-\tilde{d}, c_1) \leq \left(\frac{1}{2} + 2\Theta\right)^{\frac{1}{2}} ,$$

which shows that the projection $\tilde{T}_\rho \to T'$ is <u>injective.</u>

<u>Proof</u>. We estimate $\langle \tilde{c}-\tilde{d}, c_1 \rangle$ by modifying 4.1.3 to get the second inequality; the same computation works for 0 instead of $\tilde{d}$ if $\tilde{c} \neq c_1$, then $|c'| = |\tilde{c}| \cdot \sin \measuredangle (\tilde{c}, c_1)$ proves the first inequality. We can, after renaming, assume that

$$0 < \langle c_1, \tilde{d} \rangle \leq \langle c_1, \tilde{c} \rangle \leq (1+\Theta)|c_1|^2 \quad \text{or}$$

$$0 \leq \langle c_1, \tilde{c}-\tilde{d} \rangle \leq (1+\Theta)|c_1|^2 \ .$$

Of course only the case where $|\tilde{c}-\tilde{d}|$ is not too large might cause problems. By choice of c_1 and $\tilde{c} \neq \tilde{d}$ we have using 4.2.3

$$|c_1| \leq |\tilde{c} * \tilde{d}^{-1}| \leq (1+\Theta)|\tilde{c}-\tilde{d}|$$

and

$$|c_1| \leq |c_1^{-1} * (\tilde{c} * \tilde{d}^{-1})| \leq |(\tilde{c} - \tilde{d}) - c_1| + \Theta|c_1| + \Theta|\tilde{c} - \tilde{d}|$$

hence

$$2\langle \tilde{c}-\tilde{d},c_1\rangle = |\tilde{c}-\tilde{d}|^2 + |c_1|^2 - |\tilde{c}-\tilde{d}-c_1|^2 \leq$$

$$\leq (1-\Theta)|\tilde{c}-\tilde{d}|^2 + 3\Theta|c_1|^2 \leq (1+2\Theta+7\Theta^2)|\tilde{c}-\tilde{d}|^2 .$$

Multiplying these two bounds for $\langle \tilde{c}-\tilde{d},c_1\rangle$ we obtain $\cos^2 \sphericalangle (\tilde{c}-\tilde{d},c_1) \leq \frac{1}{2} + 2\Theta$.

4.3.4 Denseness of T'.

By assumption 4.2.1 there is for every $x \in \{c_1\}^{\perp} = \mathbb{R}^{n-1}$ with $|x| \leq \frac{1}{2}\rho$ some $c \in T_\rho$ such that $|x-c| \leq \sigma$. The representative $\tilde{c}$ and its projection $c' \in T'$ exist, we claim

$$|x-c'| \leq (1+3\Theta)\sigma =: \sigma' .$$

Proof. The worst case happens, if c and $\tilde{c}$ are on different sides of $\{c_1\}^{\perp}$. But the angle estimate in 4.3.1 still gives with 4.2.3 (i)

$$|x-c'| \leq \frac{\sigma}{\cos\vartheta} + |c_1|\cdot(1+\Theta)\cdot\sin\vartheta \leq \sigma\left(\frac{1}{1-\Theta^2} + 2\Theta(1+\Theta)\right) .$$

4.4 The product in T'.

4.4.1 Definition. Let $a',b' \in T'$ be such that $|a'|,|b'| \leq \rho' := \frac{1}{12}\rho$ and $|a'+b'| \leq (1-\Theta)\rho'$. From 4.3.3 we have for the unique (!) preimages $\tilde{a},\tilde{b} \in \tilde{T}_\rho$ the bounds $|\tilde{a}| \leq \lambda\cdot|a'|$, $|\tilde{b}| \leq \lambda|b'|$, so that the product $\tilde{a} * \tilde{b}$ is defined and the representative $(\tilde{a} * \tilde{b})^{\sim}$ can be obtained with 4.3.2. Therefore the following definition is justified:

$$a' * b' := ((\tilde{a} * \tilde{b})^{\sim})' .$$

4.4.2 Proposition. (Bounds for the product)

(i) $|a' * b' - (a'+b')| \leq \frac{\Theta}{\rho'}|a'||b'|$; $|a' * b'| \leq \rho'$.

(ii) $((\tilde{a}^{-1})^{\sim})' = (a')^{-1} \in T'$, if $|a'| \leq (1-\Theta)\rho'$.

(iii) $(a' * b') * c' = a' * (b' * c')$, if the existence of all products follows from 4.4.1.

(iv) $|[a',b']| \le \frac{2\Theta}{\rho'} |a'|\cdot|b'|$.
Together with 4.3.4 this shows that T'_ρ is Θ-translational, σ'-dense and of radius ρ' in $\mathbb{R}^{n-1} = \{c_1\}^\perp$. 4.2.3 is therefore applicable.

(v) $a'_1 * \ldots * a'_k = ((\tilde a_1 * \ldots * \tilde a_k)^\sim)'$ if $\sum_{i=1}^{k} |a'_i| \le (1-\Theta)\cdot\rho'$.

<u>Proof.</u> (i) By definition

$$|((\tilde a * \tilde b)^\sim)' - (\tilde a + \tilde b)'| \le |((\tilde a * \tilde b)^\sim - (\tilde a * \tilde b))'| + |(\tilde a * \tilde b - (\tilde a + \tilde b))'|.$$

First we bound the second term with 4.2.1 and 4.3.3

$$|\tilde a * \tilde b - \tilde a - \tilde b| \le \frac{\Theta}{\rho} |\tilde a|\cdot|\tilde b| \le \frac{\Theta}{\rho} \lambda^2 |a'|\cdot|b'| \quad .$$

Secondly, from 4.3.2 (ii) we have the angle estimate

$$|\sin \measuredangle ((\tilde a * \tilde b)^\sim - \tilde a * \tilde b \ , \ c_1)| \le \frac{\Theta}{\cos\vartheta}\cdot\frac{|\tilde a * \tilde b|}{\rho} \ ,$$

and the estimate for the scalar product

$$|\langle(\tilde a * \tilde b)^\sim - \tilde a * \tilde b \ , \ c_1\rangle| \le$$

$$|\langle(\tilde a * \tilde b)^\sim - \tilde a - \tilde b \ , \ c_1\rangle| + |\langle\tilde a * \tilde b - \tilde a - \tilde b \ , \ c_1\rangle| \le$$

$$\le 2(1+\Theta)|c_1|^2 + |c_1|\cdot\frac{\Theta}{\rho}|\tilde a|\cdot|\tilde b|$$

Now, if $\sin \measuredangle (x,y) \le \varepsilon$ and $\langle x,y\rangle \le \alpha\cdot|y|$, then
$|x|\cdot\sin \measuredangle (x,y) \le \frac{\alpha\cdot\varepsilon}{\cos \measuredangle (x,y)}$, hence

$$\begin{aligned} |((\tilde a * \tilde b)^\sim - (\tilde a * \tilde b))'| &\le \frac{\Theta}{\cos^2\vartheta}\,\frac{|\tilde a * \tilde b|}{\rho}\,\left(2(1+\Theta)|c_1| + \frac{\Theta}{\rho}|\tilde a|\cdot|\tilde b|\right) \\ &\le 2\,\frac{\Theta^2}{\rho}|\tilde a|\cdot|\tilde b| + 2\,\frac{\Theta}{\rho}(1-\Theta)^{-1}(|\tilde a|\cdot|c_1| + |\tilde b|\cdot|c_1|) \\ &\le 4.2\,\frac{\Theta}{\rho}\cdot|\tilde a|\cdot|\tilde b| \le 4.2\,\frac{\Theta}{\rho}\lambda^2\cdot|a'|\cdot|b'| \end{aligned}$$

Since $5.2\cdot\lambda^2 < 12$, the increase in the estimate is compensated by the decrease in the radius $\rho \to \rho' = \frac{1}{12}\rho$; this proves (i) .

(ii): $((\tilde{a})^{-1})^{\sim}$ exists and by definition

$$(((\tilde{a})^{-1})^{\sim})' * a' = (((\tilde{a})^{-1})^{\sim} * \tilde{a})' = ((\tilde{a}^{-1} * \tilde{a})^{\sim})' = 0 \;;$$

for the second equality we used 4.2.3 (i), namely that c_1 commutes with every element. For the same reason (iii) is true:

$$((\tilde{a} * \tilde{b})^{\sim} * \tilde{c})^{\sim} = (\tilde{a} * (\tilde{b} * \tilde{c})^{\sim})^{\sim} \;.$$

(iv) can be derived from 4.4.1 (i) or quicker from 4.2.1 (v) :

$$[a',b'] = ([\tilde{a},\tilde{b}]^{\sim})'$$

$$|[\tilde{a},\tilde{b}]^{\sim}| \le \frac{2\Theta}{\rho} |\tilde{a}||\tilde{b}| \cdot (1+\Theta) \le \frac{2\Theta}{\rho'} |a'||b'| \;.$$

(v) follows inductively from the definition 4.4.1 since associativity in T' and T is available and, again, since c_1 commutes with other elements.

4.5 The λ-normal basis for T_ρ

Rather large constants will appear in the assumptions of the following theorem; we summarize from where they come:

Shortest elements satisfy $|c_1| \le 2\sigma$ (4.2.3 (i)); for the denseness parameter we know $\sigma \le 3w(n) \cdot d(M)$ (3.6.3), where $w(n) = 2 \cdot 14^{\dim so(n)}$; the (n-1)-fold application of 4.4 looses a factor 12^{-n+1} in the radius; to use 4.2.3 (v) one needs products up to 2^{n^2} times as long as the elements x under consideration - which of course should at least include the generators c_i which we select.

<u>Theorem</u>. Let T_ρ satisfy 4.2.1 with the additional assumptions $\sigma \cdot 4^{n+6} \cdot 2^{\frac{1}{2}n^2} \le R \le (1+\Theta)^{-n-1} \cdot 12^{-n} \cdot 2^{-\frac{1}{2}n^2} \cdot \rho$; the parameter R is used in (iii) below. (When applied to 4.2.2 this can be satisfied by choosing $m = 10$, $w \ge 2^{(n+3)^2} w(n)$ in (3.2).) Then:

(i) T_ρ has a λ-normal basis $\{c_1,\dots,c_n\}$ with $\lambda \le (\frac{1}{2} - 2\Theta)^{-\frac{1}{2}}$, $|c_k| \le 2\sigma((1+3\Theta)\cdot\lambda)^{k-1} \le 2^k\sigma$.

(ii) If $\sum|l_i c_i| \le (1-2\Theta)^2\rho$ - for example if $\sum|l_i| \le 12^n 2^{n^2}$ - then $c_1^{l_1} * \dots * c_n^{l_n}$ is defined and associativity holds.

(iii) Each $x \in T_\rho$, $|x| \le R$, has a unique representation $x = c_1^{l_1} * \ldots * c_n^{l_n}$ in terms of the basis; for the images $x^{(k)}$ of x under the iterated projections $T_\rho \to T'_{\rho'} = T^{(2)} \to T^{(2)'} = T^{(3)} \to \ldots$ (4.3.3, 4.4) we have $x^{(k)} = (c_k^{(k)})^{l_k} * \ldots * (c_n^{(k)})^{l_n}$ with the same exponents.

(iv) We have $\sum_{i=k}^n |l_i c_i^{(k)}| \cdot 2^{\frac{1}{2}n^2} \le \rho_k$ hence by 4.2.3 (v)

$$|x^{(k)} - \sum_{i=k}^{n} l_i c_i^{(k)}| \le \Theta \cdot |\sum_{i=k}^{n} l_i c_i^{(k)}| \le \frac{\Theta}{1-\Theta} |x^{(k)}| .$$

(v) $|[c_i^{(k)}, c_j^{(k)}]| \le \frac{2\Theta}{\rho_k} \cdot |c_i^{(k)}| \cdot |c_j^{(k)}|$ for $k = 1,\ldots,i < j$, hence

$$[c_i, c_j] = c_1^{k_1} * \ldots * c_{i-1}^{k_{i-1}} \quad (i < j) .$$

(vi) For $x, y \in T_\rho$ with $|x| \le (1-2\Theta)R$, $|y| \le 2^{\frac{1}{2}n^2} R$ we have:
If $x = c_1^{l_1} * \ldots * c_j^{l_j}$ then $y * x * y^{-1} = c_1^{m_1} * \ldots * c_j^{m_j}$.

Proof. (i) The assumption on $\frac{\sigma}{\rho}$ allows to go through $(n-1)$ induction steps of 4.1.3 - of course using 4.3 and 4.4 - to find $\{c_1, \ldots, c_n\}$. 4.3.3 shows $\lambda \le (\frac{1}{2} - 2\Theta)^{-\frac{1}{2}}$ and with 4.2.3 (i) gives the bounds for the $|c_k|$.
(ii) restates 4.2.3 (v).

To prove (iii) we have to go through the induction once more and carry the length estimates (iv) along. Note first $|x^{(k+1)}| \le |x^{(k)\sim}| \le (1+\Theta)|x^{(k)}| \le \ldots \le (1+\Theta)^k \cdot |x|$.

Next, in $T^{(n)}$ all elements are (in a unique way) powers of the selected shortest element: $x^{(n)} = (c_n^{(n)})^{l_n}$; namely: let $b \in T^{(n)}$ and a^k the power of $a := c_n^{(n)}$ closest to it, then 4.2.3 (iii) gives
$|a^k * b^{-1}| \le (1+\Theta)|a^k - b| \le \frac{(1+\Theta)^2}{2} |a^{k+1} - a^k| \le \frac{(1+\Theta)^3}{2} |a|$, hence
$a^k * b^{-1} = 0$. Finally $|l_n c_n^{(n)}| \le (1-\Theta)^{-1} |(c_n^{(n)})^{l_n}| \le \frac{(1+\Theta)^{n-1}}{(1-\Theta)} |x| \le 2^{-\frac{1}{2}n^2} \cdot \rho_n$ by 4.3.2 (**).

Now assume by induction that we have in $T^{(k)}$ a unique representation $x^{(k)} = (c_k^{(k)})^{l_k} * \ldots * (c_n^{(k)})^{l_n}$ together with the estimate

$$\sum_{i=k}^{n} |l_i c_i^{(k)}| \cdot 2^{\frac{1}{2}n^2} \le \rho_k .$$

This implies (4.3.3)

$$\sum_{i=k}^{n} |l_i c_i^{(k-1)}| \cdot 2^{\frac{1}{2}n^2} \leq \lambda \cdot \sum_{i=k}^{n} |l_i c_i^{(k)}| \cdot 2^{\frac{1}{2}n^2} \leq \frac{\lambda}{12} \rho_{k-1} .$$

Consequently we have 4.2.3 (v) available for the product

$$y^{(k-1)} := (c_k^{(1-1)} * \dots * (c_n^{(k-1)})^{l_n} \text{ in } T^{(k-1)} ,$$

hence

$$|y^{(k-1)}| \leq (1+\Theta) \Big| \sum_{i=k}^{n} l_i c_i^{(k-1)} \Big| \leq (1+\Theta) \sum_{i=k}^{n} |l_i c_i^{(k-1)}| \leq \frac{1}{6} 2^{-\frac{1}{2}n^2} \rho_{k-1} .$$

Therefore we can use 4.3.2 to find the representative $y^{(k-1)^\sim}$ on the $c_{k-1}^{(k-1)}$ orbit towards $\{c_{k-1}^{(k-1)}\}^\perp$; moreover we can compute that representative from the product $(c_k^{(k-1)})^{l_k} * \dots * (c_n^{(k-1)})^{l_n}$: since the multiplication in $T^{(k)}$ is defined with representatives in $T^{(k-1)}$ (4.4.2 (v)) we see that $y^{(k-1)^\sim}$ must be the unique representative (4.3.3) above $x^{(k)}$, i.e. $y^{(k-1)^\sim} = x^{(k-1)^\sim}$. Observe $y^{(k-1)} = (c_{k-1}^{(k-1)})^{l} * y^{(k-1)^\sim}$ and with 4.3.2 (**)

$$|l\, c_{k-1}^{(k-1)}| \leq (1-\Theta)^{-1} \cdot |y^{(k-1)}| \leq \frac{1}{5} 2^{-\frac{1}{2}n^2} \rho_{k-1} ; \quad \text{similarly}$$

$$x^{(k-1)} = (c_{k-1}^{(k-1)})^{l'} * x^{(k-1)^\sim} , \; l_{k-1} = l + l' \quad \text{and}$$

$$|l'\, c_{k-1}^{(k-1)}| \leq (1-\Theta)^{-1} |x^{(k-1)}| \leq (1+\Theta)^k |x| \leq \frac{1}{12} 2^{-\frac{1}{2}n^2} \cdot \rho_{k-1}$$

Therefore

$$(|l'|+|l|)\,|c_{k-1}^{(k-1)}| + \sum_{i=k}^{n} |l_i c_i^{(k-1)}| \leq \frac{1}{2} 2^{-\frac{1}{2}n^2} \rho_{k-1}$$

contains the bound (iv) for $T^{(k-1)}$. Since l and l' are unique and $y^{(k-1)^\sim} = x^{(k-1)^\sim}$ we have proved (iii) and (iv). The statements (v) about the commutators are now clear (recall $[c_i^{(k)}, c_j^{(k)}] = [c_i, c_j]^{(k)}$ $(k=1,\dots,n)$). In (vi) we observe $|y * x * y^{-1}| = |[y,x] * x| \leq R$ (4.2.1 (v)), hence $y * x * y^{-1}$ has the representation $c_1^{m_1} * \dots * c_n^{m_n}$. But (iii) implies $m_{j+1} = \dots = m_n = 0$ since

$$(y * x * y^{-1})^{(j+1)} = y^{(j+1)} * x^{(j+1)} * (y^{-1})^{(j+1)} = y^{(j+1)} * (y^{-1})^{(j+1)} = 0 .$$

4.6 The embedding of Γ_R into the fundamental group $\pi_1(M,p)$

We now apply theorem 4.5 to Γ_ρ (3.5, 4.2.2) using that the map t from Γ_ρ to the set T_ρ of translational parts is injective and preserves the $*$-product. We abbreviate $|\gamma| = |t(\gamma)|$. The representation with normal words will allow to prove that Γ_R is injective, i.e. that the natural map $\gamma \to \hat{\gamma}$ from Γ_R to its presented group $\hat{\Gamma}_R = W(\Gamma_R)/N(\Gamma_R)$ is an embedding (2.2.7). Then we use the group G of equivalence classes mod Γ_ρ of 3.6 to show that injectivity of Γ_R implies injectivity of π_R. The injectivity proofs are based on the following obvious fact:

4.6.1 Proposition. If Γ_r (any r) can be embedded in a product preserving way into a group - frequently the transformation group of a suitable set - then Γ_r is injective.

In our applications of 4.6.1 we prove that suitable embeddings preserve the product by using associativity for products which have loop length (much) larger than r. This is consistent: the information beyond Γ_r is used to know enough about the relations in Γ_r.

4.6.2 Proposition. (Injectivity of Γ_R)
Choose $\rho = \rho_o$ with 3.2 and assume $\sigma \cdot 2^{(n+1)^3} \leq R \leq 2^{-(n+2)^2}\rho$. Then Γ_R is injective and $\hat{\Gamma}_R$ is a torsion free nilpotent group with the generators $\gamma_1,\dots,\gamma_n$ and the commutator relations 4.5 (v): $[\gamma_i,\gamma_j] = \gamma_1^{k_1} * \dots * \gamma_{i-1}^{k_{i-1}}$ $(1 \leq i < j \leq n)$; in particular $\hat{\Gamma}_R$ is isomorphic to a nilpotent group structure on $\mathbb{Z}^n$. The condition on R is stronger than in 4.5 for convenience. It also suffices for section 5 and can be achieved by taking $m = 10$, $w \geq 2^{(n+2)^3} w(n)$ in 3.2 $(\sigma \leq 3w(n) \cdot d(M))$.

Proof. (i) The case $n \leq 3$ is trivial: All loops γ of length $|\gamma| \leq R$ can, with 4.5, be written as normal words of loops $\gamma = \gamma_1^{l_1} * \gamma_2^{l_2} * \gamma_3^{l_3}$. If the only nontrivial commutator of the generators is given by $\gamma_3 * \gamma_2 = \gamma_1^{\nu} * \gamma_2 * \gamma_3$, then we have sufficient associativity to prove

$$\gamma_3^{l} * \gamma_2^{m} = \gamma_1^{\nu l m} * \gamma_2^{m} * \gamma_3^{l} \quad \text{as long as} \quad |\gamma_3^{l}|, |\gamma_2^{m}| \leq 2^{\frac{1}{2} n^2} R .$$

This shows that the loops $\gamma = \gamma_1^{l_1} * \gamma_2^{l_2} * \gamma_3^{l_3} \in \Gamma_R$ multiply as the matrices $\begin{pmatrix} 1 & \nu l_3 & l_1 \\ 0 & 1 & l_2 \\ 0 & 0 & 1 \end{pmatrix}$ do and proves injectivity of Γ_R with 4.6.1 .

(ii) The case $n = 4$ illustrates a part of the proof in the general case. Let the nontrivial commutators be given by $[\gamma_3, \gamma_2] = \gamma_1^{\nu}$, $[\gamma_4, \gamma_2] = \gamma_1^{\mu}$, $[\gamma_4, \gamma_3] = \gamma_1^{\lambda} * \gamma_2^{\varkappa}$, call the above matrix group $G_3 (= \mathbb{Z}^3, .)$ and define (for $l \in \mathbb{Z}$)

$$g_4^l\{(m_1, m_2, m_3)\} :=$$

$$:= (m_1 + \mu l m_2 + \lambda l m_3 + \tfrac{1}{2}\nu\varkappa m_3(m_3-1) + \tfrac{1}{2}\mu\varkappa m_3 l(l-1), m_2 + \varkappa l m_3, m_3)$$

to obtain the l-th power of an automorphism $g_4 : G_3 \to G_3$. Then the product defined by

$$(\gamma, l) \cdot (\gamma', m) := (\gamma \cdot g_4^l\{\gamma'\} , l + m)$$

turns $G_3 \times \mathbb{Z}$ into a nilpotent torsion free group G_4 with the generators $\delta_i = (\gamma_i, 0)$, $i = 1,2,3$, $\delta_4 = (0,1)$. In this group g_4^l is conjugation:

$$g_4^l\{\gamma\} = \delta_4^l \, \gamma \, \delta_4^{-l} \quad (\gamma \in G_3 \subset G_4) ,$$

and the defining commutator relations are

$$(*) \qquad [\delta_3, \delta_2] = \delta_1^{\nu} , \quad [\delta_4, \delta_2] = \delta_1^{\mu} , \quad [\delta_4, \delta_3] = \delta_1^{\lambda} \cdot \delta_2^{\varkappa} .$$

Again each loop $\gamma = \gamma_1^{l_1} * \dots * \gamma_4^{l_4} \in \Gamma_R$ is mapped - injectively - onto $(l_1, \dots, l_4) \in G_4$; associativity - holding for products of eight loops of lengths $\leq 2^{n^2/2}$ - shows that the map is product preserving. Hence Γ_R is injective for $n = 4$.

Remark. The last step also showed that a torsion free nilpotent group with generators $\delta_1, \dots, \delta_4$ and the commutator relations $(*)$ exists for any values $\nu, \mu, \lambda, \varkappa \in \mathbb{Z}$. This conclusion fails for $n \geq 5$ (e.g. because of the Jacobi identity).

For the general case we define $(\gamma_1, \ldots, \gamma_j)_r$ to be the set of loops of length $\leq r \leq R$; they all have normal words $\gamma_1^{l_1} * \ldots * \gamma_j^{l_j}$; note that associativity in these products may involve loops of length up to $2^{\frac{1}{2}n^2} r$. The following induction step yields therefore injectivity only for $\Gamma_{R'}$ with $R' = 2^{-n^3} R$. This restriction is removed in 4.6.4 where $\Gamma_{R'}$ is known as subset of the fundamental group of M .

(iii) Assume by induction that $\Gamma_j := (\gamma_1, \ldots, \gamma_j)_r$ is injective and $\hat{\Gamma}_j \cong (\mathbb{Z}^j, \cdot)$ is nilpotent without torsion - which we know for $j \leq 4$ from (i), (ii). Then the same holds for $\Gamma_{j+1} := (\gamma_1, \ldots, \gamma_{j+1})_{cr}$ where cr is assumed $\geq 4 \max\{|\gamma_1|, \ldots, |\gamma_{j+1}|\}$ and $c := 2^{-(j+1)^2}$ comes from 4.1.4 .

<u>Proof</u>. We have to define "conjugation" $\gamma_{j+1}\{\ \cdot\ \} : \hat{\Gamma}_j \to \hat{\Gamma}_j$ via the presentation of $\hat{\Gamma}_j$. Observe that $\hat{\Gamma}_j$ can also be presented with elements much shorter than r ($\gamma \in \Gamma_j$ does not imply $\gamma_{j+1} * \gamma * \gamma_{j+1}^{-1} \in \Gamma_j$): In fact $\hat{\Gamma}_j$ is assumed torsion free and has therefore the presentation $\hat{\Gamma}_j = W(\{\gamma_1, \ldots, \gamma_j\})/N'$ = free group of words modulo the commutator relations 4.5 (v). Hence

$$\gamma_i \to \gamma_{j+1}\{\gamma_i\} := (\gamma_{j+1} * \gamma_i * \gamma_{j+1}^{-1})^{\wedge} \qquad (i=1,\ldots,j)$$

defines a homomorphism $W(\{\gamma_1, \ldots, \gamma_j\}) \to W(\Gamma_j)/N(\Gamma_j)$ which because of $[\gamma_{j+1} * \gamma_i * \gamma_{j+1}^{-1}, \gamma_{j+1} * \gamma_k * \gamma_{j+1}^{-1}] = \gamma_{j+1} * [\gamma_i, \gamma_k] * \gamma_{j+1}^{-1} \in (\gamma_1, \ldots, \gamma_j)_r$ (4.5 (v), (vi)) projects to an automorphism $\gamma_{j+1}\{\ \cdot\ \} : \hat{\Gamma}_j = W(\{\gamma_1, \ldots, \gamma_j\})/N' \to W(\Gamma_j)/N(\Gamma_j) = \hat{\Gamma}_j$ (The inverse operator comes similarly from γ_{j+1}^{-1}).

An induction first over the factors of the normal word decomposition of α (with 4.1.4 (ii), 4.2.3 (v)) and then on l shows that the <u>compatibility condition</u> (which computes the above automorphism loop-wise on rather long loops)

$$(*) \qquad (\gamma_{j+1}^{l} * \alpha * \gamma_{j+1}^{-l})^{\wedge} = \gamma_{j+1}^{l}\{\hat{\alpha}\}$$

holds for all $\alpha \in \Gamma_j$, $l \in \mathbb{Z}$ satisfying $|\alpha| \leq r \cdot c^{j}$, $|\gamma_{j+1}^{l}| \leq R$ (Note that under this hypothesis $\gamma_{j+1}^{l} * \alpha * \gamma_{j+1}^{-l}$ is indeed in Γ_j by 4.5 (v), (vi)).

Now decompose each $\gamma = \gamma_1^{l_1} * \ldots * \gamma_j^{l_j} * \gamma_{j+1}^{l} \in \Gamma_{j+1}$ as

$$\gamma =: \gamma_{(j)} * \gamma_{j+1}^{l}$$

and interpret γ as transformation $\gamma^T : \hat{\Gamma}_j \times \mathbb{Z} \to \hat{\Gamma}_j \times \mathbb{Z}$ by

$$\gamma^T(d,m) := (\hat{\gamma}_{(j)} * \gamma_{j+1}^{l}\{d\}, l + m) .$$

The action on the identity $\gamma^T(0,0) = (\hat{\gamma}_{(j)}, l)$ shows that the representation $\gamma \to \gamma^T$ is injective (Here we use the injectivity of Γ_j). Moreover for $\beta = \beta_{(j)} * \gamma_{j+1}^{k} \in \Gamma_{j+1}$ it follows from $|\beta_{(j)}|, |\gamma_{(j)}| \leq r \cdot c^{j}$ and from $(*)$ that

$$(\beta * \gamma)\hat{}_{(j)} = (\beta_{(j)} * \gamma_{j+1}^{k} * \gamma_{(j)} * \gamma_{j+1}^{-k})\hat{} = \hat{\beta}_{(j)} * \gamma_{j+1}^{k}\{\hat{\gamma}_{(j)}\} .$$

Hence

$$(\beta * \gamma)^T(d,m) = (\hat{\beta}_{(j)} * \gamma_{j+1}^{k}\{\hat{\gamma}_{(j)} * \gamma_{j+1}^{l}\{d\}\}, k + l + m)$$

$$= \beta^T \circ \gamma^T(d,m) .$$

Therefore $\gamma \to \gamma^T$ extends to an isomorphic embedding of $\hat{\Gamma}_{j+1}$ into the transformation group of $\hat{\Gamma}_j \times \mathbb{Z}$. In particular Γ_{j+1} is injective, and since $\hat{\Gamma}_j$ is assumed torsion free, it follows from

$$(\gamma_1^T)^{l_1} \circ \ldots \circ (\gamma_{j+1}^T)^{l} (0,0) = (\hat{\gamma}_1^{l_1} * \ldots * \hat{\gamma}_j^{l_j}, l)$$

that $\hat{\Gamma}_{j+1}$ acts without torsion. - Since $\prod_{j=3}^{n} 2^{-j^2} \geq 2^{-n^3}$ we have so far proved 4.6.2 for $R' = 2^{-n^3} R$; see 4.6.4.

We now turn to the part in Gromov's theorem 1.5 which concerns the nilpotent subgroup of the fundamental group. The important point is that the algebraic properties of Γ_R which have been detected so far remain the same if the Gromov product $*$ is replaced by the product of $\pi_1(M,p)$.

G denotes the group of equivalence classes mod Γ_ρ from 3.6 and Ω is a set of shortest representatives.

4.6.3 Proposition. (Injectivity of π_r)
Under the assumptions 4.6.2 and $2^{-n^3}R = R' \leq r \leq 7R$ holds: If Γ_r is injective then $\pi_{r/7}$ is injective and $\hat{\pi}_{r/7}$ is isomorphic to $\pi_1(M,p)$.

If moreover the representation $\gamma = \hat{\gamma}_1^{l_1} * \ldots * \gamma_n^{l_n}$ of $\gamma \in \Gamma_r$ as normal word extends to an isomorphism between Γ_r and $(Z^n,\cdot)$ then the group Γ generated by $\{\gamma_1,\ldots,\gamma_n\}$ in $\pi_1(M,p)$ is a normal subgroup isomorphic to $\hat{\Gamma}_{r/7}$.

Remark. The implication 4.6.3 can be used because of 4.6.2 for $r = R'$ and because of 4.6.4 for $r = 7R$.

Proof. To use 4.6.1 we represent each $\alpha \in \pi_{r/7}$ as a transformation

$$\alpha^T \ : \ \Omega \times \hat{\Gamma}_r \to \Omega \times \hat{\Gamma}_r$$

which we define - with the interpretation $\alpha^T(w,c) = \alpha * w * c$ in mind - as follows: Since $w \in \Omega$ implies $|w| \leq r/100$ (3.6.2) we have $\alpha * w \in \pi_{r/6}$ and therefore with 3.6.1, 3.6.2 the unique (!) decomposition

$$\alpha * w = \omega * \gamma \ , \ \omega \in \Omega \ , \ \gamma \in \Gamma_{r/3} \ ;$$

put $\qquad \alpha^T(w,c) := (\omega, \hat{\gamma * c})$.

The representation $\alpha \to \alpha^T$ is again injective - consider $\alpha^T(0,0)$ and use injectivity of Γ_r . To prove $(\alpha_1 * \alpha_2)^T = \alpha_1^T \circ \alpha_2^T$ write

$$\alpha_2 * w = \omega_2 * \gamma_2 \ , \ \alpha_1 * \omega_2 = \omega_{12} * \gamma_{12}$$

hence

$$(\alpha_1 * \alpha_2) * w = \alpha_1 * (\omega_2 * \gamma_2) = (\omega_{12} * \gamma_{12}) * \gamma_2$$

$$= \omega_{12} * (\gamma_{12} * \gamma_2) \qquad \text{(Associativity in } \pi_r \text{ !).}$$

Then $\quad (\alpha_1 * \alpha_2)^T(w,c) = (\omega_{12},(\gamma_{12} * \gamma_2)^{\wedge} * c) = (\omega_{12},\hat{\gamma}_{12} * \hat{\gamma}_2 * c)$

$$= \alpha_1^T\ (\omega_2,\hat{\gamma}_2 * C) = \alpha_1^T \circ \alpha_2^T\ (w,c)\ .$$

Thus $\alpha \to \alpha^T$ extends to an isomorphic embedding of $\hat{\pi}_{r/7}$ into the transformation group of the set $\Omega \times \hat{\Gamma}_r$. Now 4.6.1 proves injectivity of $\pi_{r/7}$ and 2.2.7 implies $\pi_{r/7} \subset \hat{\pi}_{r/7} \cong \pi_1(M,p)$! Next, Γ is because of 3.6.4 a normal subgroup of $\pi_1(M,p)$ and because of $\gamma_i \in \Gamma_{r/7} \subset \hat{\pi}_{r/7}$ a factor group of $\hat{\Gamma}_{r/7}$ - there might be more relations between the γ_i in $\pi_1(M,p)$ than in $\hat{\Gamma}_{r/7}$; however, the elements of $\pi_1(M,p)$ were constructed as transformations on $\Omega \times \hat{\Gamma}_r$ and since $\hat{\Gamma}_r$ is isomorphic to some group $(Z^n,\cdot)$ we see that no nontrivial normal word in the generators of Γ vanishes:

$$(\gamma_1^T)^{l_1} \circ \ldots \circ (\gamma_n^T)^{l_n}(0,0) = (0,\gamma_1^{l_1} * \ldots * \gamma_n^{l_n})$$

$$= (0,0) \text{ if and only if all } l_i = 0\ .$$

4.6.4 So far 4.6.2 and the usefulness of 4.6.3 have been proved only for $R' = 2^{-n^3}R$. This restriction will now be removed: Since $\hat{\Gamma}_{R'}$ is isomorphic to $\Gamma = \langle\gamma_1,\ldots,\gamma_n\rangle \subseteq \pi_1(M)$, each $\gamma \in \Gamma$ has the <u>unique</u> representation $\gamma_1^{l_1}\ldots\gamma_n^{l_n}$ with respect to the product of $\pi_1(M)$! On the other hand 4.5 (iii) and the assumptions on R in 4.6.2 imply unique normal word representation for $\gamma \in \Gamma_{7R}$. Therefore the map $\gamma_1^{l_1} * \ldots * \gamma_n^{l_n} \to \gamma_1^{l_n}\ldots\gamma_n^{l_n}$ is an embedding of Γ_{7R} into $\pi_1(M)$ which is also product preserving - trivially, because short homotopies are homotopies - hence Γ_{7R} is injective by 4.6.1 . Now - canonically as in 4.6.3 - Γ is a factor of $\hat{\Gamma}_{7R}$ and $\hat{\Gamma}_{7R}$ is a factor of $\hat{\Gamma}_{R'}$, therefore $\Gamma = \hat{\Gamma}_{R'}$ shows $\Gamma = \hat{\Gamma}_{7R}$. This completes 4.6.2, 4.6.3 .

4.6.5 <u>Theorem</u>. (Structure of the fundamental group, summary)
Choose R as in 4.6.2. Then

(i) π_R is injective and $\hat{\pi}_R \cong \pi_1(M,p)$.

(ii) $\hat{\Gamma}_R \cong (\mathbb{Z}^n,\cdot)$ is isomorphic to a nilpotent, torsionfree normal subgroup Γ of finite index in $\pi_1(M,p)$.

(iii) Γ is generated by n loops $\gamma_1,\ldots,\gamma_n$ such that each element $\gamma \in \Gamma$ can uniquely be written as a normal word $\gamma = \gamma_1^{l_1}\cdot\ldots\cdot\gamma_n^{l_n}$; these generators are adapted to the nilpotent structure, i.e.

$$\gamma_j \cdot \langle\gamma_1,\ldots,\gamma_i\rangle \cdot \gamma_j^{-1} = \langle\gamma_1,\ldots,\gamma_i\rangle \quad (1 \le i \le j \le n) .$$

(iv) Loops in $\pi_R \subset \pi_1(M,p)$ are equivalent mod Γ_ρ (3.6) if and only if they are equivalent mod Γ , i.e. $\Gamma\backslash\pi_1(M,p) = G$; here G (from 3.6.4) is isomorphic to a subgroup of $O(n)$ with $|G| \le 2\cdot 14^{\dim SO(n)}$ (3.6.2).

Proof. (i) - (iii) restate the results of 4.6.2 to 4.6.4. To prove (iv) consider the map $\Phi : \pi_R \to G$ which sends each α to its equivalence class mod Γ_ρ . This map is injective on Ω , the set of shortest representatives choosen before 4.6.3 . We also have $\Phi(\alpha_1 * \alpha_2) = \Phi(\alpha_1)\cdot\Phi(\alpha_2)$ if $\alpha_1,\alpha_2,\alpha_1 * \alpha_2 \in \pi_R$ (3.6.4), therefore the extension of Φ to a homomorphism of the free word group $W(\pi_R)$ onto G projects to an epimorphism

$$\hat{\Phi} : \pi_1(M,p) = \hat{\pi}_R \to G$$

which contains Γ in its kernel. However $\hat{\Phi}$ is still injective on Ω , therefore $\Gamma = \text{kern}\,\hat{\Phi}$.

5. The nilpotent Lie group N and the Γ-equivariant diffeomorphism $F : \tilde{M} \to N$.

This chapter finishes the proof of Gromov's theorem in the following steps:

5.1 embeds Γ (4.6) as a uniform discrete subgroup into an n-dimensional nilpotent Lie group N using that the multiplication in Γ is polynomial in the exponents of $\gamma = \gamma_1^{l_1} \bullet \dots \bullet \gamma_n^{l_n}$.

5.2 constructs a Γ-equivariant differentiable map F from the universal covering $\tilde{M}$ to the Lie group N by interpolating local maps with the averaging methods of 8.

5.3 estimates the curvature of a left invariant metric on N which is defined with the aim of making F as almost isometric as possible; the estimate is based on and similar to the commutator estimates 3.5 (ii).

5.4 proves with 5.3 that the map F has maximal rank and hence is a Γ-equivariant diffeomorphism (5.2.6).

5.5 proves the corollaries of 1.5.

5.1 The Malcev Polynomials

The nilpotent torsionfree subgroup $\Gamma \subset \pi_1(M,p)$ was obtained in 4.6 together with generators $\gamma_1,\dots,\gamma_n$; the nilpotent structure is determined by the commutators (compare 4.5 (v))

$$[\gamma_i,\gamma_j] = \gamma_1^{k_1} * \dots * \gamma_{i-1}^{k_{i-1}} \qquad (1 \le i \le j \le n) .$$

The unique representation of $\gamma \in \Gamma$ as normal word $\gamma = \gamma_1^{l_1} * \dots * \gamma_n^{l_n}$ allows to identify Γ with a nilpotent group structure on the integer lattice $\Gamma \cong (\mathbb{Z}^n,*)$; in this context we also denote γ by its exponent vector $(l_1,\dots,l_n)$. We therefore have integer valued functions

5.1.1 $\qquad P_i : \mathbb{Z}^n \times \mathbb{Z}^n \to \mathbb{Z}$, $Q_i : \mathbb{Z}^n \to \mathbb{Z}$ $\qquad$ such that

$$(l_1,\dots,l_n) * (m_1,\dots,m_n) = (P_1,\dots,P_n) ,$$

$$(l_1,\dots,l_n)^{-1} = (Q_1,\dots,Q_n) .$$

5.1.2 <u>Theorem</u>. (Malcev [22])
The functions P_i and Q_i in 5.1.1 are polynomials.

<u>Proof</u>. It suffices to show polynomiality of the P_i since then $(l_1,\dots,l_n)^{-1} = \gamma_n^{-l_n} \;\dots\; \gamma_1^{-l_1}$ shows the polynomiality of the Q_i .

(i) The proof is by induction and (as in 4.6.2) the case $n \le 3$ is trivial: If $[\gamma_3,\gamma_2] = \gamma_1^{\nu}$ then

$$(l_1,l_2,l_3) * (m_1,m_2,m_3) \;=\; (l_1 + m_1 + \nu\cdot l_3 m_2 \,,\; l_2 + m_2 \,,\; l_3 + m_3) \;.$$

(ii) The case $n = 4$ illustrates the induction step. If $[\gamma_4,\gamma_2] = \gamma_1^{\mu}$, $[\gamma_4,\gamma_3] = \gamma_1^{\lambda} * \gamma_2^{\kappa}$ then conjugation with γ_4^{l} is given explicitely in 4.6.2 (ii) and together with (i) yields (mixed notation for better readability)

$$(l_1,\dots,l_4) * (m_1,\dots,m_4) \;=$$

$$= \;(l_1,\dots,l_3) * \gamma_4^{l_4} * (m_1,\dots,m_3) * \gamma_4^{-l_4} * \gamma_4^{l_4+m_4} \;=$$

$$= \;(l_1 + m_1 + \nu\cdot l_3 m_2 + \mu\cdot l_4 m_2 + \lambda\cdot l_4 m_3 + \nu\kappa\cdot l_4 l_3 m_3 \,+$$

$$+ \tfrac{1}{2}\,\nu\kappa\cdot l_4 m_3 (m_3-1) + \tfrac{1}{2}\,\mu\kappa\cdot l_4 (l_4-1) m_3 \,,\; l_2 + m_2 + \kappa\cdot l_4 m_3 \,,$$

$$l_3 + m_3 \,,\; l_4 + m_4) \;.$$

(iii) Assume by induction that 5.1.2 holds for groups with $(n-1)$ generators. From this we first conclude that powers are given by polynomials and then reduce the induction step to this fact.

<u>Lemma</u>. For each $(m_1,\dots,m_{n-1})$ there exist polynomials $F_i(l)$ $(i=1,\dots,n-1)$ sucht that

$$(m_1,\dots,m_{n-1})^{l} \;=\; (F_1(l),\dots,F_{n-2}(l),m_{n-1}\cdot l) \;.$$

Moreover, the $F_i(l)$ are for fixed l polynomials in $m_1,\dots,m_{n-1}$ by the induction hypothesis (iii) .

<u>Proof</u>. Consider the factors $\langle\gamma_1,\dots,\gamma_{n-1}\rangle/\langle\gamma_1,\dots,\gamma_j\rangle$ $(j=n-2,\dots,1)$. Clearly $F_{n-1}(l) = m_{n-1}\cdot l$, $F_{n-2}(l) = m_{n-2}\cdot l$ (see part (i)). By induction we assume that $F_2,\dots,F_{n-2}$ are polynomials, and it remains to show that F_1 is a polynomial. For each $l \in \mathbb{Z}$ holds

$$(m_1,\dots,m_{n-1})^{l+1} = \gamma_1^{F_1(l)} * (0,F_2(l),\dots,F_{n-1}(l)) * (m_1,\dots,m_{n-1})$$

$$= (F_1(l) + \tilde{P}(l), F_2(l+1),\dots,F_{n-1}(l+1)) ,$$

hence $\qquad F_1(l+1) = F_1(l) + \tilde{P}(l)$

where $\qquad \tilde{P}(l) = P_1(0,F_2(l),\dots,F_{n-1}(l),m_1,\dots,m_{n-1})$

is the polynomial obtained from the first Malcev polynomial P_1 of the group $\langle\gamma_1,\dots,\gamma_{n-1}\rangle$ (using the induction hypothesis (iii) above).

Now we complete the proof of 5.1.2. Decompose

$$(*) \quad (l_1,\dots,l_n) * (m_1,\dots,m_n) =$$

$$(l_1,\dots,l_{n-1}) * \gamma_n^{l_n} * (m_1,\dots,m_{n-2}) * \gamma_n^{-l_n} * \gamma_n^{l_n} * \gamma_{n-1}^{m_{n-1}} * \gamma_n^{-l_n} * \gamma_n^{l_n+m_n} .$$

We can apply the induction hypothesis (iii) to the groups $\langle\gamma_1,\dots,\gamma_{n-1}\rangle$ and $\langle\gamma_1,\dots,\gamma_{n-2},\gamma_n\rangle$; therefore, if we show that

$$\gamma_n^{l_n} * \gamma_{n-1}^{m_{n-1}} * \gamma_n^{-l_n} \in \langle\gamma_1,\dots,\gamma_{n-1}\rangle$$

is given by polynomials in l_n, m_{n-1} then the whole product $(*)$ is polynomial in $l_1,\dots,l_n,m_1,\dots,m_n$.

Abbreviate $[\gamma_{n-1}^{-1},\gamma_n] = (\alpha_1,\dots,\alpha_{n-2})$ then

$$\gamma_n^{l} * \gamma_{n-1} * \gamma_n^{-l} = \gamma_{n-1} * (\gamma_{n-1}^{-1} * \gamma_n * \gamma_{n-1})^{l} * \gamma_n^{-l}$$

$$= \gamma_{n-1} * (\gamma_1^{\alpha_1} * \dots * \gamma_{n-2}^{\alpha_{n-2}} * \gamma_n)^{l} * \gamma_n^{-l} ,$$

with the power lemma applied to $\langle \gamma_1,\dots,\gamma_{n-2},\gamma_n \rangle$:

$$= \gamma_{n-1} * (\gamma_1^{F_1(l)} * \dots * \gamma_{n-2}^{F_{n-2}(l)} * \gamma_n^l) * \gamma_n^{-1}$$

with induction hypothesis (iii) applied to $\gamma_{n-1} * (F_1(l),\dots,F_{n-2}(l)) \in \langle \gamma_1,\dots,\gamma_{n-1} \rangle$:

$$= (A_1(l),\dots,A_{n-1}(l)) \ , \ A_i(l) \ \text{polynomial} \ .$$

Finally

$$\gamma_n^l * \gamma_{n-1}^m * \gamma_n^{-1} = (A_1(l),\dots,A_{n-1}(l))^m$$

is a polynomial in l and m because of the lemma applied to powers in $\langle \gamma_1,\dots,\gamma_{n-1} \rangle$.

5.1.3 Corollary. (Embedding of Γ into N)
Use the polynomials P_i of 5.1.2 to extend the multiplication $*$ from $\mathbb{Z}^n$ to $\mathbb{R}^n$. Then $N := (\mathbb{R}^n,*)$ is a torsionfree nilpotent Lie group with $(\mathbb{R}^j,*)$ $(j=1,\dots,n)$ as normal subgroups. $(\mathbb{Z}^j,*)$ is a discrete subgroup of $(\mathbb{R}^j,*)$ with the unit cube $\{(t_1,\dots,t_j) \ ; \ 0 \le t_i \le 1\}$ as compact fundamental domain.

Proof. All claims (e.g inverses, associativity, vanishing of commutators) are statements about polynomials on $\mathbb{R}^n, \mathbb{R}^n \times \mathbb{R}^n$ etc. which are known to be true on $\mathbb{Z}^n, \mathbb{Z}^n \times \mathbb{Z}^n$ etc. .

5.1.4 Definition. (Left invariant metric on N)
The λ-normal generators $\gamma_1,\dots,\gamma_n$ for Γ determine a basis $X_i := t(\gamma_i) \in T_pM$, $i=1,\dots,n$ (2.3), they also determine a basis $Y_i := \exp^{-1}\gamma_i \in T_eN = L$ for the Lie algebra L of N . We introduce a scalar product on L such that the linear map given by $X_i \to Y_i$ is an isometry between T_pM and L , and extend this product by left translation to a Riemannian metric on N . - Clearly this definition is designed to make the "lattices" $\Gamma \cdot \tilde{p} \subset \tilde{M}$ and $\Gamma \subset N$ as nearly isometric as possible; 5.4 will show that this aim is indeed achieved.

5.1.5 Remark. For any simply connected nilpotent Lie group one can use 7.7.4 and the terminating power series of the Campbell-Hausdorff formula (7.2.7) to describe the multiplication by polynomials of degree $\leq n$ in the Lie algebra. From this description one sees immediately that any homomorphism of nilpotent Lie algebras extends to a homomorphism of the corresponding simply connected nilpotent Lie groups. However this description is not well adapted to uniform discrete subgroups.

The Malcev polynomials (5.1.2) on the other hand show immediately that any endomorphism of Γ is polynomial and consequently extends to a polynomial endomorphism of the Lie group N constructed in 5.1.3. We show in 5.1.8 that Malcev polynomials are available on any simply connected nilpotent Lie group which possesses a uniform discrete subgroup. Hence we can rephrase the preceding statement as

5.1.6 Proposition. (Γ determines N)
Any isomorphism $\Gamma \to \hat{\Gamma}$ between uniform discrete subgroups of the simply connected nilpotent Lie groups N , $\hat{N}$ respectively extends to a unique isomorphism $N \to \hat{N}$. From this we immediately deduce

5.1.7 Corollary. ($\pi_1(M)$ determines N)
Let M , $\hat{M}$ be compact n-dimensional Riemannian manifolds which are finitely covered by nilmanifolds $\Gamma \backslash N$ resp. $\Delta \backslash \hat{N}$. If M and $\hat{M}$ have the same fundamental group then $N = \hat{N}$.

Proof. Observe that the intersection $\Delta' := \Gamma \cap \Delta$ of the two finite index subgroups of π_1 is again a finite index subgroup. Consequently Δ' is uniform in N as well as $\hat{N}$ and we can apply 5.1.8, 5.1.6.

5.1.8 Proposition. (Malcev [22]; uniform discrete subgroups)
Let Δ be a uniform discrete subgroup of an n-dimensional simply connected nilpotent Lie group $\hat{N}$. Then Δ has a "triangular" set of generators $\delta_1, \ldots, \delta_n$ (i.e. $[\delta_i, \delta_j] \in \langle \delta_1, \ldots, \delta_{i-1} \rangle$) . Each $y \in \hat{N}$ can uniquely be written as $\delta_1^{t_1} \cdot \ldots \cdot \delta_n^{t_n}$; $y \in \Delta$ if and only if $(t_1, \ldots, t_n) \in \mathbb{Z}^n$ and the multiplication of $\hat{N}$ is given by the Malcev polynomials of $\Delta = \langle \delta_1, \ldots, \delta_n \rangle$ (with the exponents as coordinates).

Proof. The fact that exp is a diffeomorphism (7.7.4) is used extensively.

(i) Consider a left invariant metric on $\hat{N}$ such that $\|\mathrm{ad}_{\hat{}}\| \le \frac{1}{100r}$ where r is the diameter of a compact fundamental domain of $\Delta \backslash \hat{N}$ (7.7.3). Then Δ is r-dense (2.1.4) in the ball of radius $10r$ around e, hence $\exp^{-1}\Delta$ is still $(2r)$-dense in the ball of radius $10r$ around the origin of $L = T_e\hat{N}$ (7.4.4). In particular

$$B_{10r}(0) \cap \exp^{-1}\Delta \text{ spans the Lie algebra } L \text{ of } \hat{N} .$$

(ii) The triangular set is obtained inductively. The case $\dim \hat{N} = 1$ is trivial. For the induction step take $\delta_1 \neq e$ from the center of Δ such that

$$Y_1 := \exp^{-1}\delta_1 \text{ has minimal length; put } N_1 := \exp \mathbb{R}\cdot Y_1 .$$

Then we have for all $Y \in \exp^{-1}\Delta$:

$$\exp((\mathrm{Exp}\ \mathrm{ad}\ Y_1)\cdot Y) = \delta_1(\exp Y)\delta_1^{-1} = \exp Y .$$

Since $\exp^{-1}\Delta$ spans L, it follows that the linear map $\mathrm{Exp}\ \mathrm{ad}\ Y_1$ is the identity on L, and - applying the (terminating) power series Log to $\mathrm{Exp}\ \mathrm{ad}\ Y_1$ - we obtain that Y_1 is in the center of L, hence

$$N_1 \text{ is in the center of } \hat{N} .$$

Put $\Delta_1 := N_1 \cap \Delta$ and observe that δ_1 is a generator for Δ_1. We claim:

> $\Delta_1 \backslash \Delta$ is a uniform discrete subgroup of the simply connected nilpotent Lie group $N_1 \backslash \hat{N}$.

Proof. The projection $\hat{N} \to N_1 \backslash \hat{N}$ maps the compact Dirichlet fundamental domain

$$D := \{n \in \hat{N} ;\ d(n,e) \le d(n,\Delta - \{e\})\} \qquad (\text{for } \hat{N} \bmod \Delta)$$

onto the set of orbits $\{N_1\cdot n, n \in D\}$ which therefore contains a set of representatives of $(N_1 \backslash \hat{N}) \bmod (\Delta_1 \backslash \Delta)$. The orbits $\Delta_1\cdot\delta\ (\delta \in \Delta)$ lie on the "lines" $N_1\delta\ (\delta \in \Delta)$ in such a way that the segments between consecutive points of Δ_1 have length $\|Y_1\|$; also, there are only finitely many $\delta \in \Delta$ in the compact set $\{n \in \hat{N} ; d(n,D) \leq \|Y_1\|\}$, so that only finitely many "segments" $\exp([0,1] \cdot Y_1) \cdot \delta$ $(\delta \in \Delta)$ meet D . This proves discreteness of $\Delta_1 \backslash \Delta$ in $\hat{N}_1 \backslash \hat{N}$. Now the projection $\hat{N} \to N_1 \backslash \hat{N}$ is continuous, hence the image compact.

Now take a triangular set of generators $\delta_2', \ldots, \delta_n'$ of Δ/Δ_1 in $\hat{N}/N_1$. Then δ_1 together with a set $\delta_2, \ldots, \delta_n$ of preimages in $N \cap \Delta$ is triangular for Δ and the unique normal word decomposition carries over from $\hat{N}/N_1$ to $\hat{N}$ (with all $\delta \in \Delta$ having integer exponents since $N_1 \cap \Delta = \langle\delta_1\rangle$); as in 5.1.2 the product in Δ is given by the Malcev polynomials with respect to these generators. The differential of the projection $\hat{N} \to \hat{N}/N_1$ has the kernel $\mathbb{R}Y_1$ in L , hence

$$Y_1 \; := \; \exp^{-1}\delta_i \quad (i = 1,\ldots,n) \quad \text{is a triangular basis of } L .$$

(iii) The (terminating) Campbell-Hausdorff formula shows that the transition between Malcev-coordinates $(t_1,\ldots,t_n)$ and exponential coordinates $(x_1,\ldots,x_n)$ is polynomial, namely:

$$\exp^{-1}(\delta_1^{t_1} \cdot \ldots \cdot \delta_n^{t_n}) = x_1 Y_1 + \ldots + x_n Y_n$$

$$x_n = t_n \, , \; x_{n-1} = t_{n-1}$$

$$x_j = t_j + R_j(t_{j+1},\ldots,t_n) \quad (j=n-2,\ldots,1)$$

where each R_j is polynomial by 7.2.7 .

In particular, the multiplication of $\hat{N}$ is polynomial (5.1.5) also in Malcev coordinates and given by the Malcev polynomials on the integer lattice (ii) , hence everywhere.

5.2 The local identifications and their Γ-equivariant average

5.2.1 Actions of Γ and preferred bases.

Let $\tilde{M}$ be the universal Riemannian covering of M and $\tilde{p} \in \tilde{M}$ a point above $p \in M$. The group Γ being a finite index subgroup of the decktransformation group $\pi_1(M,p)$ (4.6.5) acts uniformly on $\tilde{M}$ by isometries. Also, Γ acts uniformly on the simply connected nilpotent Lie group N by left translations (5.1.3). We use the two actions of Γ to transport the bases $\{X_1,\ldots,X_n\} \subset T_{\tilde{p}}\tilde{M}$ (identified with T_pM) and $\{Y_1,\ldots,Y_n\} \subset T_eN = L$ of 5.1.4 to the tangent spaces $T_{\gamma\tilde{p}}\tilde{M}$ resp. $T_\gamma N$ and consider all these tangent spaces identified via these preferred bases.

5.2.2 <u>The injectivity radius of $\tilde{M}$</u>.

The number $1/2\ R$ of 4.6.2, 4.6.3 is a lower bound for the injectivity radius of $\exp_{\tilde{p}} : T_{\tilde{p}}\tilde{M} \to \tilde{M}$.

<u>Proof</u>. Since the maximal rank radius of $\exp_{\tilde{p}}$ is (much) larger than R , an injectivity radius $< \frac{1}{2} R$ would imply the existence of a geodesic loop $\tilde{\alpha}$ at $\tilde{p}$ in $\tilde{M}$ of length $< R$. This loop projects to a loop α at p in M of the same length as $\tilde{\alpha}$. Because of 4.6.5, the loop α is the shortest geodesic loop in a nontrivial homotopy class $[\alpha] \in \pi_1(M,p)$. Now $[\alpha]$ considered as element of the (fixed point free) deckgroup of the covering $\tilde{M} \to M$, maps $\tilde{p}$ to the endpoint $\neq \tilde{p}$ of the lift of α to $\tilde{M}$ - a contradiction since the lift of α is the closed loop $\tilde{\alpha}$.

5.2.3 <u>Local diffeomorphisms</u>.

We define for all $\gamma \in \Gamma$ local maps

$$F_\gamma \ : \ B_{R/2}(\gamma\tilde{p}) \to N \quad , \quad F_\gamma(x) \ := \ \exp_\gamma \circ \exp^{-1}_{\gamma\tilde{p}}(x) \ ,$$

where $B_{R/2}(\gamma\tilde{p})$ is the ball of radius $R/2$ around $\gamma\tilde{p}$ on which the exponential map $\exp_{\gamma\tilde{p}} : T_{\gamma\tilde{p}}\tilde{M} \to \tilde{M}$ has an inverse (5.2.2), and $\exp_\gamma = L_\gamma \circ \exp \circ (dL_\gamma)^{-1} : T_\gamma N \to N$ is the Lie exponential left translated to $T_\gamma N$. (The identifications $T_{\gamma\tilde{p}}\tilde{M} \leftrightarrow T_\gamma N$ from 5.2.1 via the preferred bases are suppressed in the notation).

5.2.4 <u>Local equivariance</u>.

The compatibility of the local maps with the two actions of Γ is obvious from 5.2.3:

$$\gamma_1 * F_\gamma(x) \ = \ F_{\gamma_1\gamma}(\gamma_1 x) \quad \text{for all} \quad \gamma,\gamma_1 \in \Gamma \quad \text{and} \quad x \in B_{R/2}(\gamma\tilde{p}) \ .$$

5.2.5 <u>Average of the local maps</u>.

We now interpolate the maps F_γ with the averaging technique of 8.3 to obtain a Γ equivariant differentiable map $F : \tilde{M} \to N$. For $X \in \tilde{M}$ the image $F(x)$ will be obtained as the center of mass of the various weighted image points $F_\gamma(x)$, $\gamma \in \Gamma$. Modifying 8.3.1 slightly by working with the diameter D of $\Gamma \backslash \tilde{M}$,

$$D/2 \leq \sigma \leq 2^{-(n+1)^3} \cdot R \qquad (3.6.3,\ 4.6.2)$$

rather than the injectivity radius of $\tilde{M}$ we take as weights

$$\phi_\gamma(x) := \psi\left(\frac{d(x,\gamma\tilde{p})}{D}\right) \cdot \left(\sum_{\gamma' \in \Gamma} \psi\left(\frac{d(x,\gamma'\tilde{p})}{D}\right)\right)^{-1} = \phi_{\gamma_1\gamma}(\gamma_1 x)$$

where ψ is the smooth cut off function from 8.3.1 and the sum over Γ is finite since $\Gamma \subset \pi_1(M,p)$ acts properly discontinuous on $\tilde{M}$ and ψ has compact support. Now the map

$$F : \tilde{M} \to N \ , \quad F(x) := \mathcal{C}\{F_\gamma(x), \phi_\gamma(x)\} \qquad (8.3.1)$$

is Γ-equivariant: $F(\gamma x) = \gamma * F(x)$ (5.2.4, 8.3.2) and differentiable (8.3.3).

5.2.6 <u>End of the proof of Gromov's theorem</u>.

So far the metric 5.1.4 played no role. It will be used in 5.4 to show that dF_X has maximal rank for all $x \in \tilde{M}$. If we anticipate this now, then

$$\Gamma\backslash F : \Gamma\backslash\tilde{M} \to \Gamma\backslash N$$

is a differentiable map of maximal rank between compact manifolds, hence a finite covering. Its lift

$$F : \tilde{M} \to N$$

is therefore a Γ-equivariant finite covering of the simply connected space N - hence a Γ-equivariant diffeomorphism. Now $\Gamma\backslash F$ is the desired diffeomorphism between the finite covering $\Gamma\backslash\tilde{M}$ of M (with deckgroup $\simeq G$ (3.6)) and the nilmanifold $\Gamma\backslash N$.

5.3 The left invariant metric and its curvature.

In this section we derive from the commutator estimate (in terms of loop lengths)

$$|t[\alpha,\beta]| \leq \frac{4\theta}{\rho} |t(\alpha)| \; |t(\beta)| \qquad (3.5\ (iii))$$

a similar bound for the Lie bracket of L. In order to restate 3.5 (iii) in terms of the Lie algebra L we need

5.3.1 Definition.

$$G(X) := \sum_{i=1}^{n} x_i Y_i \quad \text{if} \quad \exp X = \gamma_1^{x_1} * \dots * \gamma_n^{x_n} .$$

Note that each $n \in N$ has the decomposition $n = \gamma_1^{t_1} * \dots * \gamma_1^{t_n}$ as normal word, and (with 7.7.4) $G : L_i := \text{span}\{Y_1,\dots,Y_i\} \to L_i$ is a diffeomorphism. Moreover

$$G(x_1Y_1 + x_iY_i) = x_1Y_1 + x_iY_i \quad (i = 2,\dots,n) .$$

We have from theorem 4.5 (iii) that every loop γ of length $\leq R$ can be written as $\gamma = \gamma_1^{k_1} * \dots * \gamma_n^{k_n}$ with

$$|t(\gamma) - \Sigma\, k_i\, t(\gamma_i)| \leq \theta\, |\Sigma\, k_i\, t(\gamma_i)| \qquad (4.5\ (iv)),$$

and the norm in L has been defined such that

$$|\Sigma\, k_i\, t(\gamma_i)|_{T_p\tilde{M}} = |\Sigma\, k_i\, Y_i|_L .$$

Finally recall from 7.1.2, 7.2.7, 7.7.4

5.3.2 $$\exp H((\operatorname{Exp} \operatorname{ad} X)\cdot Y,-Y) = [\exp X, \exp Y]_N .$$

Therefore we can rewrite 3.5 (iii) (with a correction factor $(1-\theta)^{-1}(1+\theta)^2$, $\theta \leq 1/50$) as

5.3.3 $$|G(H((\mathrm{Exp\ ad}\ X)\cdot Y,-Y))| \leq \frac{4.3\Theta}{\rho}\,|G(X)|\cdot|G(Y)|$$

if $\exp X$, $\exp Y \in \Gamma_R$.

In order to obtain a bound for $|[X,Y]|$ from 5.3.3 and 5.3.2 we need in the inductive proof and again in 5.4 also a comparison between the vectors $X \in \exp^{-1}\Gamma_R$ and the corresponding lattice vector $G(X)$ which is similar to 4.5 (iv).

5.3.4 <u>Proposition</u>. (Norm of the Lie bracket)
Same assumption as in 4.6.2. For $i = 1,\dots,n$ holds

(i) $$\|\mathrm{ad}\ X|_{L_i}\| \leq \frac{5\Theta}{\rho}\,|X| \qquad \text{(see also 7.7.1)}$$

if $X \in L_i$ or $X = Y_{i+1},\dots,Y_n$.

(ii) $$|G(Z) - Z| \leq \frac{4.9\Theta}{\rho}\,2^{\frac{1}{2}n^2}\min\{|Z|,|G(Z)|\}\cdot\textstyle\sum_{j=1}^{i+1}|z_jY_j|$$

$$\leq \frac{5\Theta}{\rho}\,2^{n^2}\min\{|Z|^2,|G(Z)|^2\},$$

if $Z = \sum_{j=1}^{i+1} z_jY_j \in L_{i+1}$ with $\sum|z_jY_j| \leq 2^{\frac{1}{2}n^2}R$

$(L_{n+1} := L_n = L)$.

Note that $Y_1,\dots,Y_n$ is a normal basis (5.1.4, 4.5 (i)); therefore (i) implies (with 4.1.4 (iii))

(iii) $$\|\mathrm{ad}\ X|_{L_i}\| \leq \frac{5\Theta}{\rho}\,2^{n^2/2}|X| \quad \text{for all } X \in L .$$

<u>Remark</u>. The induction is needed to control error terms whereas the main estimate (5.3.3) is induction free. Therefore the inequalities have the same constants for all i .

To interpolate 5.3.3 from the discrete set $G(\exp^{-1}\Gamma_R)$ we also need the following

5.3.5 Lemma. Let $D \subset L$ be a set which is δ-dense (2.1.4) in the ball $B_r(0) \subset L$. Let $A : L \to W$ be a linear map from L into some metric vector space W. Put

$$|A|_D := \max\{ \frac{1}{r} |AX|_W ; X \in D \cap B_r\} .$$

Then
$$||A|| \leq \frac{r}{r-2\delta} |A|_D .$$

Proof. $|A|_D$ is also the maximum of $\frac{1}{r} |AX|$ for X from the convex hull of $D \cap B_r$. However

$$\mathrm{con}(D \cap B_r) \supset B_{r-2\delta} .$$

(Otherwise a supporting hyperplane of $\mathrm{con}(D \cap B_r)$ would meet $B_{r-2\delta}$ and cut off so big a piece of B_r that D could not be δ-dense in B_r). Therefore

$$|A|_D \geq \max\{ \frac{1}{r} |AY|_W ; Y \in B_{r-2\delta}\} = \frac{r-2\delta}{r} ||A|| .$$

Proof of proposition 5.3.4. The start of the induction is trivial: Since L_1 is in the center of L, we have $\mathrm{ad}\, X|_{L_1} = 0$ for all $X \in L$ and $G(Z) = Z$ on L_2. Now assume 5.3.4 (i) and (ii) for i.

To prove (i) on L_{i+1} we first consider preimages of lattice points:

$$X = G^{-1}(\sum_{j=1}^{i+1} l_j Y_j), l_j \in \mathbb{Z}, |X| \leq \frac{1}{2} R , \text{ or } X = Y_{i+2}, \dots Y_n ;$$
$$Y = G^{-1}(\sum_{j=1}^{i+1} k_j Y_j), k_j \in \mathbb{Z}, |Y| \leq \frac{1}{2} R .$$

The assumption $R < \rho \cdot 2^{-(n+2)^2}$ (4.6.2) and 5.3.4 (ii) implies that

$$|G(Z) - Z| < 10^{-3}|Z| \quad \text{for all} \quad Z \in L_{i+1} \quad \text{satisfying} \quad |Z| < R ;$$

therefore $\exp X$, $\exp Y \in \Gamma_R$ and we can rewrite 5.3.3 as

$$|H((\mathrm{Exp}\, \mathrm{ad}\, X)Y,-Y)| \leq \frac{4.4\Theta}{\rho} |X||Y| .$$

Next $\|\mathrm{ad}\, X|_{L_i}\|$, $\|\mathrm{ad}\, Y|_{L_i}\| \leq 10^{-3}$ by 5.3.4 (iii), therefore we can apply 7.7.5 (iii) to obtain

$$|[X,Y]| \leq 1.01\, |H((\mathrm{Exp}\ \mathrm{ad}\ X)Y,-Y)| ,$$

hence

$$|[X,Y]| \leq \frac{4.5\Theta}{\rho} |X||Y| .$$

Finally we recall that the lattice points $\sum_{j=1}^{i+1} m_j Y_j$ are $(2^{-(n+1)^3}R)$-dense in the ball $B_{R/2}(0)$ (5.1.4, 4.5 (i)), so that $\exp^{-1}(\Gamma_{R/2})$ is $10^{-4}R$-dense in $B_{R/2}$ by (ii). Now we apply 5.3.5 twice to obtain

$$(*) \qquad |[X,Y]| \leq \frac{4.6\Theta}{\rho} |X||Y|$$

for all $X,Y \in L_{i+1}$ or $X = Y_{i+2},\ldots,Y_n$.

To prove (ii) on L_{i+2} we consider $Z = \sum_{j=1}^{i+2} z_j Y_j$ with $\sum_{j=1}^{i+2} |z_j Y_j| \leq 2^{n^2/2}R$ and define

$$Z_{(j)} := G^{-1}(\textstyle\sum_{k=1}^{j} z_k Y_k) \qquad (j = 1,\ldots,i+2) .$$

$$\text{Then} \qquad Z_{(j+1)} - \sum_{k=1}^{j+1} z_k Y_k = H(Z_{(j)}, z_{j+1}Y_{j+1}) - \sum_{k=1}^{j+1} z_k Y_k$$

$$= H(Z_{(j)}, z_{j+1}Y_{j+1}) - (Z_{(j)} + z_{j+1}Y_{j+1}) + (Z_{(j)} - \textstyle\sum_{k=1}^{j} z_k Y_k) ,$$

hence

$$|G^{-1}(Z) - Z| = |\textstyle\sum_{j=1}^{i+1} H(Z_{(j)}, z_{j+1}Y_{j+1}) - Z_{(j)} - z_{j+1}Y_{j+1}|$$

$$(7.8.2\ (ii):) \qquad \leq \sum_{j=1}^{i+1} \left|[Z_{(j)}, z_{j+1}Y_{j+1}]\right| \cdot f(\|z_{j+1}\ \mathrm{ad}\ Y_{j+1}|_{L_{i+1}}\|)$$

$$((*)\ \text{above}:) \qquad \leq \frac{4.7\Theta}{\rho} \sum_{j=1}^{i+1} |z_{j+1}Y_{j+1}| \cdot |Z_{(j)}|$$

$$(5.3.4\ (ii):) \qquad \leq \frac{4.8\Theta}{\rho} \sum_{j=1}^{i+1} |z_{j+1}Y_{j+1}| \cdot |\textstyle\sum_{k=1}^{j} z_k Y_k|$$

$$(4.1.4\ :) \qquad \leq \frac{4.8\Theta}{\rho} 2^{1/2n^2} |Z| \cdot \sum_{k=1}^{j+2} |z_k Y_k| .$$

This proves 5.3.4 (observe $|Z| \leq 1.01|G^{-1}(Z)|$) .

5.4 Maximal rank of F

5.4.1 Outline. In this section we complete 5.2.6 by proving that dF has maximal rank. Fix $x_o \in \tilde{M}$, abbreviate $y_o := F(x_o)$ and let $t \to x_t$ be an arbitrary curve through x_o with $|\dot{x}| \neq 0$. It is shown in 8.3.3 and 8.3.4 that dF_{x_o} has maximal rank if and only if $(\frac{d}{dt} \upsilon(x_t,y_o))_{t=o} \neq 0$, where

$$\upsilon(x_t,y_o) = -\sum_{\gamma \in \Gamma} \phi_\gamma(x_t) \cdot \exp_{y_o}^{-1} (F_\gamma(x_t)) \in T_{y_o} N , \quad \phi_\gamma \text{ from 5.2.5.}$$

Since $\sum_{\gamma \in \Gamma} \phi_\gamma(x_o)\exp_{y_o}^{-1} F_\gamma(x_o) = 0$ (8.1.4) it therefore suffices to show

$$(*) \qquad 0 \neq \sum_{\gamma \in \Gamma} (\frac{d}{dt} \Psi(d(x_t,\gamma\tilde{p})/D)_{t=o} \cdot \exp_{y_o}^{-1} F_\gamma(x_o)$$
$$+ \sum_{\gamma \in \Gamma} \Psi(d(x_o,\gamma\tilde{p})/D) \cdot (\frac{d}{dt} \exp_{y_o}^{-1} \circ F_\gamma(x_t))_{t=o} .$$

The proof of $(*)$ proceeds in the following steps: The number of terms in both sums of $(*)$ is compared (5.4.2). The distance between the different local images $F_\gamma(x_o)$ is shown to be small (5.4.3). The differentials of the different local maps are close to each other (5.4.4). These details are taken together in 5.4.5.

5.4.2 Lemma. (Number of terms in $(*)$)
Define

$$N_1 := \# \{\gamma;\psi(d(x_o,\gamma\tilde{p})/D) = 1\} ,$$

$$N_2 := \# \{\gamma;\psi(d(x_o,\gamma\tilde{p})/D) \neq 0\} \qquad \text{then}$$

$$N_2 \leq N_1 \cdot 1.6^n .$$

Proof. Around each $\gamma\tilde{p} \in \tilde{M}$ we have the fundamental domain

$$\{y \in \tilde{M} ; d(y,\gamma\tilde{p}) \leq d(y,\gamma_1\gamma\tilde{p}) \text{ for all } \gamma_1 \in \Gamma\}$$

with volume $V := \mathrm{vol}(\Gamma \backslash \tilde{M})$ and contained in a ball of radius $D := \mathrm{diam}(\Gamma \backslash \tilde{M})$. Clearly

$$N_1 \geq \mathrm{vol}(B_{7D}(x_o))/V \ , \quad N_2 \leq \mathrm{vol}(B_{11D}(x_o))/V \ .$$

Now 6.4.1 gives (in view of the strong curvature assumptions)

$$N_2/N_1 \leq \left(\frac{\sinh \Lambda r}{\Lambda}\right)^n \Bigg|_{r = 11D} \cdot \left(\frac{\sin \Lambda r}{\Lambda}\right)^{-n} \Bigg|_{r = 7D} \leq 1.6^n \ .$$

5.4.3 Proposition. (Distances between the local images)
Same assumptions as in 5.2.3. If $x_o \in \tilde{M}$ and $\gamma, \delta \in \Gamma$ satisfy $d(x_o, \gamma\tilde{p}) \leq 10D$ and $\delta\tilde{p} \in \Gamma\tilde{p}$ is a point closest to x_o (closer than D) then

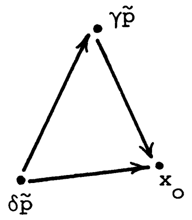
$$d(F_\gamma(x_o), F_\delta(x_o)) \leq 1.6^{-n} D/11 \ ;$$

hence also

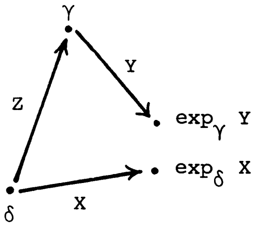
$$d(F(x_o), F_\delta(x_o)) \leq 1.6^{-n} D/11 \qquad (8.1.4,\ 8.1.8),$$

$$|\exp^{-1}_{F(x_o)} F_\gamma(x_o) \leq 1.6^{-n} D/5 \qquad (7.4.4,\ 8.1.9)\ .$$

Proof.



in $\tilde{M}$



in N

The local maps F_γ (5.2.3) are defined with exponential maps from different tangent spaces; to prove 5.4.3, 5.4.4 we therefore need fairly canonical maps between these tangent spaces, namely the following

(i) Identifications. Recall that all tangent spaces $T_{\gamma\tilde{p}}\tilde{M}, T_\gamma N$ ($\gamma \in \Gamma$) were isometrically identified in 5.2.1 in a Γ-equivariant way via the preferred bases. Using left translations $dL_n : T_e N \to T_n N$ we extend this canonically to an identification of all tangent spaces of N. Such a canonical procedure is not available on $\tilde{M}$, but since x_o and δ are fixed throughout the proof we will map all tangent spaces $T_x\tilde{M}$ ($x \in B_{10D}(x_o)$) via radial parallel translation P_{rad} (from x to $\delta\tilde{p}$) onto $T_{\delta\tilde{p}}\tilde{M}$. Note that $P_{rad} : T_{\gamma\tilde{p}}\tilde{M} \to T_{\delta\tilde{p}}\tilde{M}$ does not agree with the above Γ-equivariant identification if the rotational part (2.3) of the loop $\delta^{-1}\gamma$ at $p \in M$ is nontrivial; however 3.5 shows for all γ with $d(x_o,\gamma\tilde{p}) \le 10D$ (i.e. $|t(\delta^{-1}\gamma)| \le 11\,D$):

$$\|\text{identification error of } P_{rad} : T_{\gamma\tilde{p}}\tilde{M} \to T_{\delta\tilde{p}}\tilde{M}\| \le 18\,\Theta\,D/\rho\ .$$

(ii) Errors caused by the nonlinearity of the exponential maps of $\tilde{M}$ and N.

On $\tilde{M}$ we have from 6.6.1 and 6.4.1 and 3.2 (K)

$$(ii)_M \qquad |\exp^{-1}_{\delta\tilde{p}} x_o - (\exp^{-1}_{\delta\tilde{p}}\gamma\tilde{p} + P_{rad}\exp^{-1}_{\gamma\tilde{p}} x_o)| \le$$

$$\le \frac{1}{3}\Lambda \cdot 11\,D^2 \sinh 12\,\Lambda D \cdot \frac{12\,\Lambda\,D}{\sin 12\,\Lambda D} \le 45\,\Lambda^2 D^3$$

On N consicer vectors $Y \in T_\gamma N$, $X \in T_\delta N$, $Z := \exp^{-1}_\delta \gamma$ (figure above) and observe $\exp_\delta X = \exp_\gamma H(-Z,X)$ (7.2.7). Then 7.4.4, 7.8.2 and 5.3.4 imply - assuming $|X| \le D$, $|Y|,|Z| \le 12\,D$ -

$$(ii)_N \qquad d(\exp_\delta X, \exp_\gamma Y) = d(\exp_\gamma H(-Z,X), \exp_\gamma Y)$$

$$\le f(14\,D \cdot \|ad\|) \cdot |H(-Z,X)-Y|$$

$$\le f(70\,\Theta\,D/\rho)(|X-Z-Y| + |[Z,X]| \cdot f(5\,\Theta\,D/\rho))$$

$$\le 1.01\,(|X-(Z+Y)| + 5\,\frac{\Theta}{\rho}\,|Z|\cdot|X|)\ .$$

If there would be no further errors, then $(ii)_N$ - with $|X-(Z+Y)| \le 45\,\Lambda^2 D^3$ from $(ii)_M$ - would complete the proof.

(iii) Errors caused since <u>Γ is not abelian.</u>
In the definition 5.2.1 of the local maps the vectors $\exp^{-1}_{\delta\tilde{p}} x_o$ and $\exp^{-1}_{\gamma\tilde{p}} x_o$ (figure above) are Γ-equivariantly transported from $\tilde{M}$ to N via the preferred bases. Therefore these vectors should be the vectors X and Y in $(ii)_N$. In $(ii)_M$ we have $P_{rad} \exp^{-1}_{\gamma\tilde{p}} x_o \in T_{\delta\tilde{p}}\tilde{M}$ instead of the Γ-transport of $\exp^{-1}_{\gamma\tilde{p}} x_o$, i.e. we made the identification error (i) $\le$
$\le (18\ \Theta\ D/\rho) \cdot |\exp^{-1}_{\gamma\tilde{p}} x_o|$.

Next, to use $(ii)_{M_1}$ in $(ii)_N$ we have to be able to compare the vectors $\exp^{-1}_{\delta\tilde{p}} \gamma\tilde{p}$ and $\exp^{-1}_{\delta} \gamma$. Since $\delta^{-1}\gamma$ is a loop (in Γ) of length $\le 11\ D$ we have the normal word decomposition $\delta^{-1}\gamma = \gamma_1^{k_1} * \ldots * \gamma_n^{k_n}$ as a product of loops (4.5) as well as in Γ (4.6). Now 4.2.3 (v) implies

$$(iii)_M \qquad |t(\delta^{-1}\gamma) - \Sigma\, k_i\ t(\gamma_i)| \le \frac{\Theta}{\rho} 2^{n^2}\ |\Sigma\, k_i\ t(\gamma_i)|^2$$

and 5.3.4 gives $(Y_i = \exp^{-1} \gamma_i)$

$$(iii)_N \qquad |\exp^{-1}(\delta^{-1}\gamma) - \Sigma\, k_i Y_i| \le \frac{5\Theta}{\rho} 2^{n^2}\ |\Sigma\, k_i Y_i|^2 .$$

Note that $\Sigma\, k_i\ t(\gamma_i)$ and $\Sigma\, k_i Y_i$ are canonically identified in (i) and not longer than $11.1\ D$. Another small error is caused since $\exp^{-1}_{\delta\tilde{p}} \gamma\tilde{p}$ is the initial tangent and $t(\delta^{-1}\gamma)$ the final tangent vector of the loop, but 3.5 (i) bounds the difference by $1.6\ \frac{\Theta}{\rho}\ (11\ D)^2$.

(iv) We insert all errors in the described way in $(ii)_N$ to obtain

$$d(F_\gamma(x_o), F_\delta(x_o)) \le$$

$$1.01(45\ \Lambda^2 D^3 + 180\ \Theta\ D^2/\rho + (6 \cdot 2^{n^2} + 1.6)\ \frac{\Theta}{\rho}\ 11.1^2 D^2 + 55.5\ \Theta\ D^2/\rho)$$

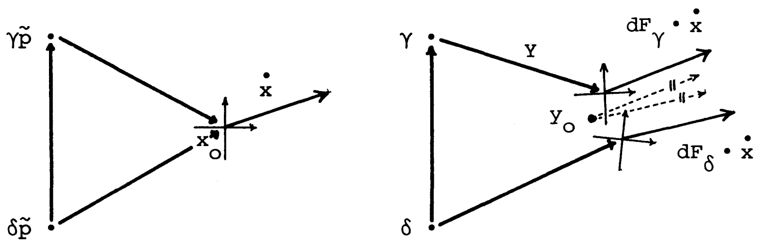
$$\le 1.6^{-n}\ D/11 ,$$

where $D \le 2\sigma$ (5.2.5) and $D/\rho \le 2^{-n^2} \cdot 40^{-n}$ (4.5) were used. (Obviously we have much smaller error estimates than needed in the present proof.)

5.4.4 Proposition. (The differentials of the F_γ almost agree)
Assumptions as in 5.4.3, $y_o = F(x_o)$, compare 5.4.1 (*) .

$$|d \exp_{y_o}^{-1} \circ dF_\gamma \cdot \dot{x} - d \exp_{y_o}^{-1} \circ dF_\delta \cdot \dot{x}| \leq 0.1 |\dot{x}| .$$

Proof.



Recall $F_\gamma(x) = \exp_\gamma \circ \text{identification}(T_{\gamma\tilde{p}}\tilde{M} \to T_\gamma N) \circ \exp_{\gamma\tilde{p}}^{-1} x$. We have to compare the differentials of the exponential maps involved with respect to the identifications 5.4.3 (i). On N occurs only one type of error controlled by 7.4.4 and 5.3.4:

For $y \in T_\gamma N$, $|Y| \leq 11 D$, $\eta := \exp_\gamma Y$ holds

$$\|(d \exp_\gamma)_Y - dL_{\eta\gamma^{-1}}\| \leq f(11 D\|ad\|) - 1 \leq 55 \Theta D/\rho ;$$

and with 5.4.3 and $W_\gamma := \exp_{y_o}^{-1} F_\gamma(x_o)$

$$\|(d \exp_{y_o}^{-1})_{W_\gamma} - dL_{y_o \cdot F_\gamma(x_o)^{-1}}\| \leq \frac{f(D\cdot\|ad\|/10)-1}{2-f(D\cdot\|ad\|/10)} \leq \Theta D/\rho .$$

On $\tilde{M}$ we first have the same type of error just discussed on N , it is controlled by 6.4.2 ($\kappa = 0$, $|K| \leq \Lambda^2$) : For $Y \in T_{\gamma\tilde{p}}\tilde{M}$, $|Y| \leq 10 D$, $x := \exp_{\gamma\tilde{p}} Y$ holds

$$|(d \exp)_Y - P_{\gamma\tilde{p} \to x}| \leq \frac{\sinh 10 \Lambda D}{10 \Lambda D} - 1 \leq 20 \Lambda^2 D^2 .$$

For $\gamma = \delta$ this is the only error; for $\gamma \neq \delta$ there are two more:

Since parallel translation in $\tilde{M}$ is path depending, we decided in 5.4.3 (i) to use radial parallel translation to $\delta\tilde{p}$ where tangent spaces had to be identified by parallel translation (here for the use of 6.4.2). The error between this radial identification and $P_{\gamma\tilde{p} \to x}$ is controlled by 6.1.1, 6.2.1, 6.7, 6.7.1 (for $d(\delta\tilde{p},x) \le 1.1\ D$) :

$$|P_{rad}(T_{\gamma\tilde{p}}\tilde{M} \to T_{\delta\tilde{p}}\tilde{M} \to T_x\tilde{M}) - P_{\gamma\tilde{p} \to x}| \le 10\ \Lambda^2 D^2 .$$

Finally, we have to come back to the Γ-equivariant transport $T_{\gamma\tilde{p}}\tilde{M} \to T_{\delta\tilde{p}}\tilde{M}$; this adds the identification error 5.4.3 (i).

Now add the error on N and the three errors on $\tilde{M}$ to prove 5.4.4:

$$|d\ \exp_{y_o}^{-1} \circ dF_\gamma \cdot \dot{x} - d\ \exp_{y_o}^{-1} \circ dF_\delta \cdot \dot{x}| \le (74\ \Theta\ D/\rho + 30\ \Lambda^2 D^2) \cdot |\dot{x}| .$$

(Again the estimates are much better than needed here.)

5.4.5 Proof of 5.4.1 (*).
Since $|\frac{d}{dt}\psi(d(x_t,\gamma\tilde{p})/D)| \le |\dot{x}|/D$ we have from 5.4.2, 5.4.3

$$|\sum_{\gamma \in \Gamma} \frac{d}{dt}\psi(d(x_t,\gamma\tilde{p})/D)|_{t=o} \cdot \exp_{y_o}^{-1} F_\gamma(x_o)| \le 0.2\ N_1 \cdot |\dot{x}|$$

and from 5.4.2, 5.4.4

$$|\sum_{\gamma \in \Gamma} \psi(d(x_o,\gamma\tilde{p})/D) \cdot \frac{d}{dt} \exp_{y_o}^{-1} \circ F_\gamma(x_t)| \ge 0.8\ N_1 \cdot |\dot{x}| .$$

5.5 Applications under more specific assumptions

5.5.1 Proof of Corollary 1.5.1

We have to find an upper bound for the injectivity radius at $p \in M^n$ under the assumption $|K|d^2(M) \le \varepsilon \le \varepsilon_n$ if $\Gamma \subset \pi_1(M)$ is not abelian. Take $w = 2^{(n+2)^3}(\frac{\varepsilon_n}{\varepsilon}) \cdot w(n)$ in 4.6.2 and 3.2. Then we find with 3.2 $\rho_o \ge 2wd(M)m^4$ $(m = 10)$ such that $|[\alpha,\beta]| \le \frac{4\Theta}{\rho_o}|\alpha|\cdot|\beta|$ for all loops $\alpha,\beta \in \Gamma$ with lengths $|\alpha|,|\beta| \le \frac{1}{3}\rho_o$ (3.5 (iii)). By 4.5 (i) there exist two noncommuting loops $\alpha,\beta \in \Gamma$ of length $\le 2^n \cdot \sigma = 2^n \cdot 3w(n)\ d(M)$ hence $[\alpha,\beta]$ is a nontrivial geodesic loop at p of length $\leq 2^{-n^3}(\varepsilon/\varepsilon_n)\ d(M)$.

5.5.2 Proof of Corollary 1.5.2

Consider the family g_t of ε-flat metrics on $\Gamma\backslash\mathcal{H}_o$, $e^{-2bt} \leq g_t \leq e^{-2at}$ (1.4. (iii)). If Γ has nilpotency degree p , then we find loops $\alpha_1,\ldots,\alpha_p$ in Γ such that $\beta := [\ldots[\alpha_1,\alpha_2],\ldots,\alpha_p]$ is a nontrivial geodesic loop. As $t \to \infty$ (i.e. $\varepsilon \to 0$) we obtain from 3.5 (iii)

$$|\beta| \ := \ |[\ldots[\alpha_1,\alpha_2],\ldots,\alpha_p]| \ \leq \ \text{const}\ |\alpha_1|\cdot\ldots\cdot|\alpha_p| \ ,$$

where the constant is independent of t and where the $|\alpha_i|$ are of order $\leq$ $\leq e^{-at}$, and $|\beta|$ is of order $\geq e^{-bt}$ as $t \to \infty$. This is only possible for $b \geq pa$.

6. Curvature controlled constructions

6.1 Curvature. The covariant derivative of a vectorfield $t \to X(t)$ along a curve $t \to c(t)$ is denoted $\frac{D}{dt} X$; if $X(t) = Y(c(t))$, Y a vectorfield defined near the image of c , then $\frac{D}{dt} X = D_{\dot c} Y$, where DY is called covariant differential of Y . The curvature tensor is the skewsymmetric part of the second covariant differential

$$R(u,v)X := D^2_{u,v} X - D^2_{v,u} X \quad , \quad D^2_{u,v} X := D_u(D_v X) - D_{D_u v} X \; .$$

If D is the Levi-Civita derivative of a Riemannian metric, we have the sectional curvature

$$K = \langle R(u,v)v,u \rangle \quad , \qquad u,v \in T_pM \quad \text{orthonormal.}$$

Constant sectional curvature κ implies

$$R_\kappa(u,v)X = \kappa \cdot (\langle v,x\rangle \cdot u - \langle u,x\rangle \cdot v) \; .$$

Sectional curvature bounds $\delta \le K \le \Delta$ imply (since $b(u,w) := \langle R(u,v)v,w \rangle$ is a symmetric bilinear form and since

$$6\, R(u,v)w = R(u,v+w)(v+w) - R(v,u+w)(u+w)$$

$$- R(u,v-w)(v-w) + R(v,u-w)(u-w) \; ,$$

see [5]).

6.1.1
$$|R(u,v)v - R_{\frac{\Delta+\delta}{2}}(u,v)v| \le \frac{\Delta-\delta}{2} |u|\,|v|^2 \; ,$$

$$|\langle R(u_1,u_2)u_3,u_4\rangle| \le \frac{2}{3}(\Delta-\delta) \quad \text{for orthonormal } u_i \; ,$$

$$||R|| := \max\{|R(u_1,u_2)u_3| \; ; \; |u_i|=1\} = \max(\Delta,-\delta,\frac{2}{3}(\Delta-\delta))$$

$$\le \frac{4}{3}\max|K| \; .$$

Because of these inequalities sectional curvature bounds give frequently sufficient control of the curvature tensor.

6.2 Parallel Translation. We consider linear and affine parallel translations; the definition of the curvature tensor immediately implies in both cases curvature controlled dependence on the path. This is basic for the proof of Gromov's and other pinching theorems, see 2.3, 2.4.

Assumptions. Let $c_i : [o,1] \to M$ be curves from p to q and c_t $(1 \leq t \leq 2)$ a homotopy between c_1 and c_2. Assume for the area of the homotopy

$$\iint \left|\frac{\partial}{\partial t} c_t(s) \wedge \frac{\partial}{\partial s} c_t(s)\right| dsdt \leq F$$

and for all $t \in [1,2]$ for the lengths of the paths c_t

$$\int_o^1 \left|\frac{\partial}{\partial s} c_t\right| ds \leq L .$$

Let E be a vector bundle over M with a fibre norm and a covariant derivative such that parallel translation is a norm-isometry. Let

$$s \to X_i(s) \ , \ \frac{D}{ds} X_i = o \ , \ X_1(o) = X_2(o) \in E_p$$

be parallel vectorfields along c_i with the same initial vector. Then:

6.2.1 Path dependence of linear translations.

$$|\sphericalangle(X_1(1) \ , \ X_2(1))| \leq \|R\| \cdot F \qquad \text{(Riemannian case)}$$

$$|X_1(1) - X_2(1)| \leq \|R\| \cdot F \qquad \text{(general case)} .$$

Let in the Riemannian case

$$s \to Y_i(s) \ , \ \frac{D}{ds} Y_i = \frac{\partial}{\partial s} c_i \ , \ Y_1(o) = Y_2(o) \in T_pM$$

be vectorfields which are called "affine parallel". Then:

6.2.2 Path dependence of affine translations.

$$|Y_1(1) - Y_2(1)| \leq (|Y_i(o)|+L) \cdot ||R|| \cdot F$$

Remark. The affine translation is closely related to the old concept of the "development of a curve": Let $Y(o) = O$ and Levi-Civita-translate $Y(s)$ back to $c(o)$, then the image in T_pM is called development of c ; it has in the euclidean geometry of T_pM the same Frenet invariants as the curve c in the Riemannian geometry of M .

Proof. Let $X(s,t)$ be the unit vectorfield along the homotopy which is parallel along the paths c_t , i.e.

$$X(o,t) = X_1(o) = X_2(o) \in E_p , \frac{D}{ds} X(s,t) = O .$$

Then, since $t \to X(1,t) \in E_q$ joins $X_1(1)$ to $X_2(1)$:

$$\left.\begin{array}{r} |\sphericalangle(X_1(1), X_2(1))_{T_qM}| \\ |X_1(1) - X_2(1)|_{E_q} \end{array}\right\} \leq \int_1^2 |\frac{D}{dt} X(1,t)| \, dt$$

$$\leq \int_1^2 \int_o^1 |\frac{D}{ds}\frac{D}{dt} X(s,t)| \, ds \, dt \qquad \text{(since } \frac{D}{dt} X(o,t) = O)$$

$$= \int_1^2 \int_o^1 |R(\frac{\partial}{\partial s} c, \frac{\partial}{\partial t} c)X(s,t)| \, ds \, dt \quad \text{(since } \frac{D}{dt}\frac{D}{ds} X = O)$$

$$\leq ||R|| \cdot |X| \cdot \int_1^2 \int_o^1 |\frac{\partial}{\partial s} c \wedge \frac{\partial}{\partial t} c| \, ds \, dt = ||R|| \cdot F .$$

Similarly let $Y(s,t)$ be the vectorfield along the homotopy which is affine parallel along the paths, i.e.

$$\frac{D}{ds} Y(s,t) = \frac{\partial}{\partial s} c_t .$$

Since the difference of two affine parallel vectorfields is a linear parallel field it suffices to assume special initial conditions $Y(o,t) = O \in T M$.

Observe

$$|Y(s,t)| \le |Y(o,t)| + \int_o^1 |\frac{D}{ds} Y(s,t)| ds \le \text{length } (c_t) \le L .$$

Then, since $t \to Y(1,t)$ joins $Y_1(1)$ to $Y_2(1)$ in $T_q M$

$$|Y_2(1) - Y_1(1)| \le \int_1^2 |\frac{D}{dt} Y(1,t)| dt$$

$$\le \int_1^2 \int_o^1 |\frac{D}{\partial s} (\frac{D}{dt} Y - \frac{\partial}{\partial t} c_t)(s,t)| ds\, dt$$

$$\text{(since } \frac{\partial}{\partial t} c_t(1) = 0 , \frac{D}{dt} Y(o,t) - \frac{\partial}{\partial t} c_t(o) = 0)$$

$$= \int_1^2 \int_o^1 |R(\frac{\partial}{\partial s} c , \frac{\partial}{\partial t} c) Y(s,t)| \, ds\, dt$$

$$\text{(since } \frac{D}{dt} (\frac{D}{ds} Y(s,t) - \frac{\partial}{ds} c_t) = 0)$$

$$\le L \cdot \|R\| \cdot F .$$

6.3 Jacobi fields and geodesic constructions. Any construction with families of geodesics $c_t(s)$ gives rise to Jacobi fields $J_t(s) = \frac{\partial}{\partial t} c_t(s)$ along these geodesics. They are determined by the Jacobi equation

$$\text{6.3.1} \qquad \frac{D}{ds}\frac{D}{ds} J + R(J, \frac{\partial}{\partial s} c) \frac{\partial}{\partial s} c = 0$$

together with the initial conditions

$$J_t(o) = \frac{\partial}{\partial t} c_t(o) , \frac{D}{ds} J_t(o) = \frac{D}{ds}\frac{\partial}{\partial t} c_t(o) = \frac{D}{dt}\frac{\partial}{\partial s} c_t(o) .$$

These initial conditions are specified by the geodesic construction in question; if they are such that one has good estimates for the solutions of the Jacobi-equation, then one further integration translates the estimates into information about the geodesic construction. The best known example of this procedure are the Rauch estimates for the radial exponential map: By differentiating the radial family of geodesics at $t = o$

$$c_t(r) = \exp_p r \cdot (v+tw)$$

one obtains a Jacobi field with simple initial data whose values give the differential of $\exp_p$:

6.3.2 $$J(o) = 0\ ,\ \frac{D}{dr} J(o) = w \quad \text{and} \quad J(r) = (d\,\exp_p)_{rv} \cdot rw\ .$$

Under the curvature assumption $\delta \leq K \leq \Delta$ one can show that $|J(r)|$ lies between the values of the constant curvature solutions with the final conclusion that the length-distortion of $\exp_p$ is between the constant curvature (δ resp. Δ) cases, see 6.4 for a more explicit statement.

6.3.3 Observation. The tangential component of a Jacobi field is independent of the metric; use $\langle J,c'\rangle'' = 0$ to obtain

$$J^{\tan}(s) = (a+s\cdot b)\cdot c'(s) \quad (a,b \in \mathbb{R})\ .$$

Therefore all further estimates will be stated for normal Jacobi fields $J(s) \perp c'(s)$.

6.3.4 Abbreviation. In the case of constant sectional curvature κ the Jacobi equation reduces to the scalar equation

$$f'' + \kappa \cdot f = 0\ .$$

If κ is constant or a continuous function we use special symbols to denote those solutions which have the same initial data as cos or sin :

If $f(o) = 1,\ f'(o) = 0$ denote the solution by $\mathbf{c}_\kappa$,

if $f(o) = 0,\ f'(o) = 1$ denote the solution by $\mathbf{s}_\kappa$.

6.3.5 Lemma. (Estimates in terms of upper curvature bounds.)
Assume $K \leq \Delta$ and $|c'| = r$. Let $f_\Delta = |J(o)|\cdot \mathbf{c}_{\Delta r^2} + |J|'(o)\cdot \mathbf{s}_{\Delta r^2}$ be the solution of $f'' + \Delta r^2 f = 0$ with the same initial conditions as $|J|$. Assume $f_\Delta(s) > 0$ for $s \in (o,t)$. Then:

(i) $$\langle J,J'\rangle \cdot f_\Delta \geq \langle J,J\rangle \cdot f'_\Delta \qquad \text{on } (o,t)\ ,$$

(ii) $$1 \leq \frac{|J(s)|}{f(s)} \leq \frac{|J(t)|}{f(t)} \qquad s \in (o,t)\ ,$$

(iii) $$|J(o)| \cdot C_{\Delta r^2}(s) + |J|'(o) \cdot S_{\Delta r^2}(s) \leq |J(s)|\ ,\ s \in [o,t]$$

("Rauch's lower bound").
Equality holds if and only if $K = \Delta$ and $|J|'(o) = |J'(o)|$.

Proof. Recall that $J \perp c'$ is assumed after 6.3.3 .

$$|J|'' + \Delta r^2 |J| =$$

$$= |J|^{-1}(-\langle R(J,c')c',J\rangle + \Delta r^2 \langle J,J\rangle) + |J|^{-3}(|J'|^2|J|^2 - \langle J,J'\rangle^2) \geq 0$$

therefore

$$(|J|' \cdot f_\Delta - |J| \cdot f')' = |J|'' \cdot f_\Delta + |J| \cdot f_\Delta'' \geq 0\ .$$

Now $\lim_{s \to o}(|J|' f_\Delta - |J| \cdot f_\Delta') = 0$ gives the first inequality. Next

$$\left(\frac{|J|}{f_\Delta}\right)' = \frac{1}{f_\Delta^2} \cdot (|J|' \cdot f_\Delta - |J| \cdot f_\Delta') \geq 0 \qquad \text{and}$$

$\lim_{s \to o} \frac{|J|}{f_\Delta}(s) = 1$ proves the second. Rauch's lower bound is a special case.

6.3.6 Lemma. (Comparison between J and J')
Assume $K \leq \Delta$, $J(o) = 0$, $|c'| = r$. Let t be such that $S_{\Delta r^2}$ is increasing on (o,t) . Then

$$|J(s) - s \cdot J'(s)| \leq |J(t)| \cdot \frac{1}{2} \max|K| \cdot s^2 r^2$$

This inequality is sharp only in the flat case; a more complicated one is sharp if the curvature is constant; a weaker one holds as long as $S_{\Delta r^2}$ is positive; proofs are similar.)

Proof. For every unit parallel field P and $s \in (o,t)$ holds

$$|\langle J(s) - s\cdot J'(s),\ P(s)\rangle'| = |s\cdot\langle R(J,c')c',P\rangle(s)|$$

$$\leq \max|K| \cdot |J(s)|r^2 s$$

(with 6.3.5:)

$$\leq |J(t)| \frac{\mathbf{s}_{\Delta r}(s)}{\mathbf{s}_{\Delta r}(t)} \max|K|r^2 s .$$

Use the crude estimate $\mathbf{s}_{\Delta r^2}(s) \leq \mathbf{s}_{\Delta r^2}(t)$ and integrate $\int_o^s \ldots\ldots ds$.

6.3.7 Lemma. (Vectorvalued estimates)
Assume twosided curvature bounds $\delta \leq K \leq \Delta$ and for simplicity $|c'| = 1$.
Let κ be a parameter and put $\lambda = \max(\Delta-\kappa,\kappa-\delta)$. (The inequalities will be sharper for larger $\kappa \leq \frac{1}{2}(\Delta+\delta)$, but they hold on a larger interval if κ is smaller.)
Let A be the vectorfield along c which satisfies

$$\frac{D}{ds}\frac{D}{ds} A + \kappa\cdot A = o , \quad A(o) = J(o) , \quad A'(o) = J'(o) .$$

Let a be the solution of

$$a'' + (\kappa-\lambda)\, a = \lambda\cdot|A| , \quad a(o) = a'(o) = o .$$

Let t be such that $\mathbf{s}_\kappa$ is positive on (o,t) . Then

$$|J-A|(s) \leq a(s) \quad \text{on} \quad [o,t] .$$

6.3.8 Corollary. (Special initial data, lower curvature bounds)
Assume in addition that $J(o)$ and $J'(o)$ are linearly dependent and that t is such that

$$f_\kappa = |A(o)|\cdot \mathbf{c}_\kappa + |A|'(o)\cdot \mathbf{s}_\kappa \quad \text{is positive on} \quad (o,t) .$$

Then:

$$|A| = f_\kappa \ , \ |A|^{-1}\cdot A \ \text{ is a parallel field,} \quad a = f_{\kappa-\lambda} - f_\kappa \ ,$$

(i) $|J-A| \leq (f_{\kappa-\lambda} - f_\kappa)$ on $[o,t]$;

with $\kappa = \frac{1}{2}(\Delta+\delta)$ this gives "Rauch's upper bound":

(ii) $|J(s)| \leq f_\delta(s) = |J(o)|\cdot \mathbf{c}_\delta(s) + |J|'(o)\cdot \mathbf{s}_\delta(s)$;

with $\kappa = 0$, $\Lambda^2 = \lambda = \max|K|$ and P_s parallel translation along c one has a good comparison with the flat case:

(iii) $|J(s) - P_s(J(o)+s\cdot J'(o))| \leq |J(o)|(\cosh \Lambda s-1) + |J|'(o)\,\frac{1}{\Lambda}\,(\sinh \Lambda s - \Lambda s)$.

Proof. For every unit parallel field we have with 6.1.1

$$|\langle J-A,P\rangle'' + \kappa\cdot\langle J-A,P\rangle| = |\langle J''-\kappa J,P\rangle| \leq \lambda\cdot|J| \ .$$

If b satisfies

$$b''+\kappa b = \lambda\cdot|J| \ , \ b(o)=b'(o) = 0 \qquad \text{then}$$

$$\frac{1}{\mathbf{s}_\kappa}\cdot\{\langle J-A,P\rangle - b\}$$

vanishes at 0 and is nonincreasing, its derivative being

$$\mathbf{s}_\kappa^{-2}\cdot\int_0 (\{\quad\}''\cdot\mathbf{s}_\kappa - \{\quad\}\cdot\mathbf{s}_\kappa'') \leq 0 \ .$$

This gives $|J-A| \leq b$ as long as $\mathbf{s}_\kappa$ is nonnegative and implies

$$b'' + \kappa b \leq \lambda b + \lambda|A| \ .$$

As before $\frac{1}{\mathbf{s}_\kappa}\cdot\{b-a\}$ vanishes at 0 and is nonincreasing, which proves $b \leq a$ and therefore the lemma. All claims in the corollary are then obvious.

Applications to geodesic constructions.

6.4 Proposition. (Rauch's and other estimates on $\exp_p$)
Assume curvature bounds $\delta \le K \le \Delta$. Then: The radial exponential map $\exp_p : T_pM \to M$ is a radial isometry:

$$|(d\exp)_v \cdot v| = |v| ;$$

perpendicular to the radial direction one has the Rauch estimates along $\exp rv$ $(|v|=1)$:

6.4.1 $$|w| \cdot \frac{\mathbf{s}_\Delta(r)}{r} \le |(d\exp_p)_{rv} \cdot w| \le |w| \cdot \frac{\mathbf{s}_\delta(r)}{r} .$$

The bounds are sharp in the respective constant curvature cases. The left inequality holds as long as the lower bound is nonnegative, the right inequality holds as long as $\mathbf{s}_{\frac{1}{2}(\Delta+\delta)}$ is positive (or with a different proof: up to the first conjugate point). The differential can be compared with parallel translation P_r along $\exp rv$ $(|v| = 1 , w \perp v)$:

6.4.2 $$|d\exp_{rv}(w) - P_r(\frac{\mathbf{s}_\kappa(r)}{r} \cdot w)| \le |w| \cdot \frac{\mathbf{s}_{\kappa-\lambda}(r) - \mathbf{s}_\kappa(r)}{r} ,$$

which holds as long as $\mathbf{s}_\kappa$ is nonnegative. (κ arbitrary, $\lambda = \max(\Delta-\kappa, \kappa-\delta)$, see 6.3.8 .)

Proof. We saw $(d\exp)_{rv} \cdot rw = J(r)$ with $J(0) = 0$, $J'(0) = w$ in 6.3.2 . Apply 6.3.5, 6.3.8 (i), (ii) with these initial conditions.

6.4.3 Corollary. (Angle comparisons for small geodesic triangles)
Assume curvature bounds $\delta \le K \le \Delta$. Consider a geodesic triangle $T = pqr \subset M$ with circumference u less than the radius of injectivity of M and if $\Delta > 0$ also $u < 2\pi\Delta^{-\frac{1}{2}}$. Then we have triangles T_δ , T_Δ in the planes M_δ , M_Δ of constant curvature δ, Δ which have the same edge-lengths as T ("Aleksandrow triangles"). For corresponding angles one has

$$\alpha_\delta \le \alpha \le \alpha_\Delta , \quad \beta_\delta \le \beta \le \beta_\Delta , \quad \gamma_\delta \le \gamma \le \gamma_\Delta .$$

Proof. To prove $\alpha_\delta \le \alpha \le \alpha_\Delta$ consider also triangles T_δ^*, T_Δ^* which have the angle α and the two adjacent edgelengths the same as T. Then use the distance decreasing maps $\exp_p \circ \exp_{p_\delta}^{-1}$ (resp. $\exp_{p_\Delta} \circ \exp_p^{-1}$) (6.4.1) to map the third geodesic edge of T_δ^* (resp. T) onto a curve joining those vertices of T (resp. T_Δ^*) which are different from p (resp. p_Δ). Since the geodesic edge is not longer than this curve we have for the third edgelengths of T_δ^*, T, T_Δ^* the inequalities

6.4.4 $$a_\delta \ge a \ge a_\Delta .$$

This shows that the third edge of T_δ^* (resp. T_Δ^*) has to be decreased (resp. increased) to give the third edge of T_δ (resp. T_Δ); this decreases the angle α of T_δ^* to the angle α_δ of T_δ (resp. increases the angle α of T_Δ^* to the angle α_Δ of T_Δ).

6.4.5 Remark. To prove $\alpha \le \alpha_\Delta$ we only used that u is less than twice the radius of injectivity of exp. The lower bound $\alpha_\delta \le \alpha$ holds in fact without any restrictions on the triangle ("Toponogow's theorem"). The above proof shows this if $\delta + \Delta \le 0$. If more positive curvature occurs, in particular if $\delta > 0$, then the proof is more complicated and not just a consequence of Jacobi field estimates [7].

6.4.6 Proposition. (Convexity, injectivity and derivatives of exp)
Let r be smaller than the injectivity radius of $\exp_p$; let Δ be an upper curvature bound on the compact ball $B_r(p)$, and if $\Delta > 0$ assume $r < \frac{1}{2}\pi\Delta^{-\frac{1}{2}}$.

(i) The function $f : B_r \to R$, $f(q) := \frac{1}{2} d(p,q)^2$ has the first derivative

$$\operatorname{grad} f(q) = -\exp_q^{-1} p \ ;$$

its second derivative is expressible in terms of the Jacobi field $J(s)$ along the geodesic from p to q determined by $J(o) = o$, $J(1) = X \in T_qM$:

$$D_X \operatorname{grad} f(q) = J'(1) .$$

In particular 6.3.5 (i) and 6.3.6 imply

$$r \cdot \frac{\mathbf{s}'_\Delta}{\mathbf{s}_\Delta}(r) \cdot |X|^2 \le |D^2_{X,X} f| \le (1 + \tfrac{1}{2} \max|K| \cdot r^2) \cdot |X|^2 ,$$

so that f is <u>convex</u> on $B_r(p)$ $(r \cdot \frac{\mathbf{s}'_\Delta}{\mathbf{s}_\Delta}(r) > 0)$.

(ii) Any two points in B_r have <u>in</u> B_r a unique and shortest goedesic connection of length $\le 2r$; in other words, $\exp_q$ is for each $q \in B_r$ a diffeomorphism onto B_r from a suitable preimage of B_r in T_qM .

<u>Proof</u>. (i) For any geodesic $q(t)$ in B_r with $\dot{q}(0) = x$ define the family of geodesics from p to $q(t)$

$$c(s,t) := \exp_p s \cdot \exp_p^{-1} q(t) = \exp_{q(t)}(1-s) \cdot \exp_{q(t)}^{-1} p .$$

Then

$$f(q(t)) = \frac{1}{2} \int_0^1 \left\langle \frac{\partial}{\partial s} c , \frac{\partial}{\partial s} c \right\rangle ds ,$$

$$\frac{d}{dt} f(q(t)) = \int_0^1 \left\langle \frac{D}{\partial t} \frac{\partial}{\partial s} c , \frac{\partial}{\partial s} c \right\rangle ds = \left\langle \frac{\partial}{\partial t} c , \frac{\partial}{\partial s} c \right\rangle (1,t) ,$$

since $\frac{D}{\partial t} \frac{\partial}{\partial s} c = \frac{D}{\partial s} \frac{\partial}{\partial t} c$ and $\frac{\partial}{\partial t} c\,(0,t) = 0$. Now

$$- \exp_{q(t)}^{-1} p = \frac{\partial}{\partial s} c\,(1,t) = \operatorname{grad} f\,(q(t)) .$$

To the family $c(s,t)$ of geodesics corresponds the family of Jacobi fields along these geodesics

$$J_t(s) := \frac{\partial}{\partial t} c\,(s,t)$$

which are determined by the boundary data

$$J_t(o) = 0 , J_t(1) = \dot{q}(t) .$$

Clearly $-\frac{D}{dt} \exp^{-1}_{q(t)} (p) = \frac{D}{\partial t}\frac{\partial}{\partial s} c\ (1,t) = \frac{D}{\partial s}\frac{\partial}{\partial t} c\ (1,t) = J_t'(1)$.

The bounds of the second derivative now follow directly from 6.3.5 (i) and 6.3.6 .

(ii) First let $q_o, q_1 \in B_r(p)$ be interior points with $r_1 := d(q_1,p) \leq d(q_o,p) =: r_o < r$. In the compact metric space $B_r(p)$ we have (by Arzela-Ascoli) a shortest curve $q(t)$ from q_o to q_1 which we can write as

$$q(t) = \exp_p v(t) \quad \text{with} \quad |v(t)| \leq r \ (< \tfrac{1}{2}\pi\,\Delta^{-\frac{1}{2}}) .$$

Because of 6.3.5 we can change $q(t)$ to the following curve $\tilde{q}(t)$ which is not longer than $q(t)$ but in the <u>interior</u> of $B_r(p)$, hence a shortest geodesic of length $\leq r_o + r_1 < 2r$:

$$\tilde{q}(t) = \exp_p \tilde{v}(t) , \tilde{v}(t) := v(t) \cdot r_o / \max(r_o, |v(t)|) .$$

For boundary points we take limits to obtain a shortest geodesic connection of length $\leq 2r$ in $B_r(p)$.

The exponential map has along any geodesic of length $\leq 2r$ maximal rank (6.4.1) so that these geodesics are locally unique. In particular, the set of pairs $(q_o,q_1) \in B_r \times B_r$ with at least two geodesic connections of length $\leq 2r$ in B_r is compact. If this set were not empty we could take a pair (q_o,q_1) with minimal distance and two joining geodesics c_1,c_2 in B_r , $d(q_o,q_1) = \text{length}\ (c_1) \leq \text{length}\ (c_2) \leq r$. Now, if c_1 and c_2 would not form a closed geodesic but have an angle $< \pi$ at, say, q_1 , then we could find points $\tilde{q}_1$ closer to q_o than q_1 and still with two geodesic connections from q_o to $\tilde{q}_1$, contradicting the choice of (q_o,q_1) . But c_1 and c_2 cannot form a closed goedesic either, since the lower bound of D^2f excludes that f has a maximum on it. This proves (ii) .

The following example is not used in the proof of the almost flat manifold theorem but in the application 2.5.3 which Gromov used to illustrate the power of the results in section 2 . It is also an example where the tangential components of Jacobi fields (6.3.3) need special attention.

6.5 Example. (The normal exponential map of a geodesic)
Because special initial conditions were important for sharp Jacobi field estimates one cannot always give bounds which can be interpreted in constant curvature goemetries (of the same dimension). For certain applications it is enough to have some explicit bound.

Assume: M is complete and simply connected with curvature bound $K \leq -\lambda^2 < 0$. Let $\alpha : [0,1] \to M$ be a geodesic segment and $\beta, \gamma : [0,\infty) \to M$ unit speed geodesic rays perpendicular to α with $\beta(0) = \alpha(0)$, $\gamma(0) = \alpha(1)$.
Then:

$$\cosh \lambda d(\beta(t),\gamma(t)) \geq (\cosh \lambda d(\beta(0),\gamma(0)) - 1)\cdot\cosh^2(\lambda t) + 1$$

$$=: \cosh \lambda f(t) ,$$

hence

6.5.1 $$d(\beta(t),\gamma(t)) \geq f(t) \geq d(\beta(0),\gamma(0)) ,$$

with f strictly increasing and $\lim_{t\to\infty} f(t) = \infty$.

Proof. Let $c(r) = \exp_{\alpha(r)} v(r)$ be the shortest geodesic from $\beta(t)$ to $\gamma(t)$ where $v(r)$ is a vectorfield along α and $\perp \alpha'$. To estimate $c'(r)$ introduce the family of geodesics $s \to c(r,s) = \exp_{\alpha(r)} s\cdot v(r)$ and the corresponding Jacobi fields

$$s \to J_r(s) = \frac{\partial}{\partial r} c(r,s) \quad \text{with} \quad J_r(0) = \alpha'(r),\ J_r(1) = c'(r),\ \frac{D}{ds} J_r(0) = \frac{D}{dr} v(r) .$$

From $\langle\alpha'(r), v(r)\rangle = 0$ we have $0 = \langle\alpha'(r), \frac{D}{dr} v(r)\rangle = \langle J,J'\rangle(0)$; since $J_r(0) \perp v(r)$ we have $J_r(0) = J_r^{norm}(0)$ (6.3.3) so that 6.3.5 can be used to give

$$|J_r^{norm}(1)| \geq |\alpha'(r)| \cdot \cosh \lambda \, |v(r)| .$$

Since

$$J_r^{tan}(o) = 0 \quad \text{and} \quad \frac{D}{ds} J_r^{tan}(o) = \frac{\langle v(r), \frac{D}{dr} v(r)\rangle}{\langle v(r), v(r)\rangle} \cdot v(r)$$

we find

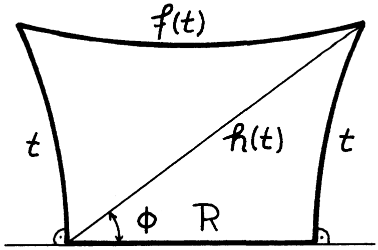
$$|J_r^{tan}(1)| = \left|\frac{d}{dr} |v(r)|\right| ,$$

hence

$$\int_0^1 |c'(r)| dr \geq \int_0^1 \left(|\alpha'(r)|^2 \cdot \cosh^2 \lambda \, |v(r)| + \left|\frac{d}{dr} |v(r)|\right|^2\right)^{\frac{1}{2}} dr .$$

The right hand side of this inequality depends only on $|v(r)|$ and a lower bound can be obtained by interpreting that expression in the 2-dimensional hyperbolic plane of curvature $-\lambda^2$; in the 2-dimensional case the inequality is sharp since

$$\frac{D}{ds} J_r^{tan}(o) = \frac{D}{ds} J_r(o) , \text{ hence } |J_r^{norm}(1)| = |\alpha'(r)| \cdot \cosh \lambda \, |v(r)| .$$



The minimum is obtained from the hyperbolic quadrilateral in the figure and the formulas

$$\cosh \lambda h = \cosh \lambda t \cdot \cosh \lambda R$$

$$\sin \phi = \sinh \lambda t / \sinh \lambda h$$

$$\cosh \lambda f = \cosh \lambda t \cdot \cosh \lambda h - \sinh \lambda t \sinh \lambda h \cdot \sin \phi$$

$$= \cosh \lambda R \cdot \cosh^2 \lambda t - \sinh^2 \lambda t .$$

The next result does not appear in Gromov's manuscript. We use it to prove the commutator estimates in section 2 and to construct the equivariant interpolation in section 5.

6.6 <u>Proposition</u>. (Comparison of riemannian und euclidean parallel translation.)

Assume $|K| \leq \Lambda^2$.

Let $v \in T_pM$ be fixed; for $w \in T_pM$ let $w(t)$ be the parallel vectorfield along $c(t) = \exp tv$; $c(1) = q$. We compare the following two maps from T_pM into M which agree in the flat case (!):

$$F : w \to \exp_p(v+w)$$
$$G : w \to \exp_q(w(1))$$

6.6.1 $$d(F(w), G(w)) \leq \frac{1}{3} |v| \cdot |w| \cdot \Lambda \sinh \Lambda(|v| + |w|) \cdot \sin \measuredangle (v,w) .$$

If we use parallel translations along the geodesics involved to identify tangent spaces, then we also control the differentials:

6.6.2
$$|dF_w \cdot X - X| \leq |X| \cdot (\frac{\sinh \Lambda\rho}{\Lambda\rho} - 1) , \quad \text{where} \quad \rho = |v+w| ,$$
$$|dG_w \cdot X - X| \leq |X| \cdot (\frac{\sinh \Lambda\rho}{\Lambda\rho} - 1) , \quad \text{where} \quad \rho = |w| .$$

<u>Proof</u>. The control of the differentials is given by 6.4.2 $(\kappa = 0)$ when applied to geodesics of speed ρ rather than 1 . To estimate the distance, consider the family of geodesics $s \to c(s,t) = \exp_{c(t)}(s\cdot(w(t)+(1-t)\cdot \dot{c}(t)))$.

Then $s \to \frac{\partial c}{\partial t}(s,t)$ is a family of Jacobi fields with linearly dependent initial conditions:

$$J_t(o) = \dot{c}(t) , \frac{D}{\partial s} J_t(o) = \frac{D}{\partial t}\frac{\partial}{\partial s} c_{|s=o} = -\dot{c}(t) .$$

Since the curve $t \to c(1,t)$ joins $F(w)$ to $G(w)$ and has the tangent vectors $J_t(1)$ we can apply 6.3.8 (iii) to get the desired estimate. In this case the Jacobi fields involved clearly are not perpendicular to their geodesics, but the initial conditions are such that $J_t^{\tan}(1) = 0$ for all t . With

$$\cosh x - \frac{\sinh x}{x} \leq \frac{1}{3} x \cdot \sinh x \quad \text{and} \quad \rho = \max(|w|, |v+w|)$$

we obtain from 6.3.8 applied to geodesics of speed $|\frac{\partial c}{\partial s}|$:

$$d(F(w), G(w)) \leq \int_0^1 |J_t(1)^{\text{norm}}| \, dt$$

$$\leq \int_0^1 |J_t(0)^{\text{norm}}| \cdot |\frac{\partial c}{\partial s}| \cdot \frac{1}{3} \Lambda \cdot \sinh \Lambda\rho \, dt .$$

Finally

$$|J_t(0)^{\text{norm}}|^2 \cdot |\frac{\partial c}{\partial s}|^2 = |\frac{\partial c}{\partial t}|^2 |\frac{\partial c}{\partial s}|^2 - \langle \frac{\partial c}{\partial t}, \frac{\partial c}{\partial s} \rangle^2 \Big|_{s=0}$$

$$= |v|^2 \cdot |w+(1-t)v|^2 - \langle v, w+(1-t)v\rangle^2$$

$$= |v|^2 |w|^2 - \langle v,w\rangle^2 ,$$

which ends the proof.

The following area estimate of A. D. Aleksandrow is used to bound the area of certain homotopies (compare 6.2, 2.2) in the proof of Gromov's theorem. Note that this estimate from above depends only on upper curvature bounds, whereas usually lower curvature bounds are employed for such a purpose. The result is rather sharp, sharper than worthwhile for our present application. We use it because it is geometrically elegant and a typical (though little known) pinching result.

6.7 Proposition. (A. D. Aleksandrow's area estimate for geodesic triangles.)

Let $T = pqr$ be a small geodesic triangle, i.e. make the assumptions of 6.4.3 . We span a 2-dimensional "ruled" surface S into the triangle by choosing C^3-differentiable monotone maps $f : [0,1] \to pq$ (= segment from p to q), $g : [0,1] \to pr$ and by joining $f(t)$ to $g(t)$ by the unique minimizing geodesic. The intrinsic curvature of S is also $\leq \Delta$, while a lower curvature bound depends on f and g . Then:

> The area of the ruled surface S is not larger than the area of T_Δ (6.4.3).

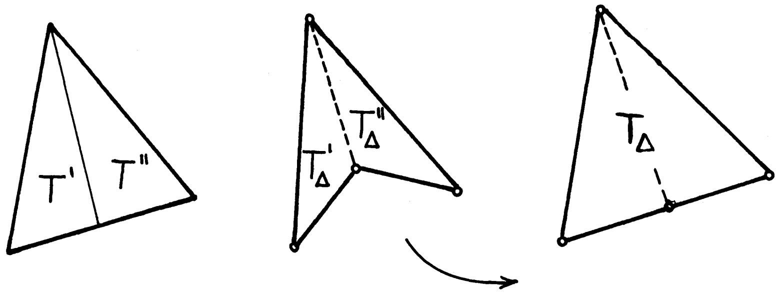
<u>Proof</u>. The curvature bound for S follows either from the Gauß equation or by observing that the goedesics from f(t) to g(t) are also geodesics on S ; differentiation gives vectorfields J which are Jacobi fields on M as well as on S , therefore

$$\langle J,J\rangle'' = 2\left\langle \frac{D}{ds} J, \frac{D}{ds} J\right\rangle - 2\left\langle R(J,c')c',J\right\rangle$$

and $\quad \left|\frac{D^M}{ds} J\right| \geq \left|\frac{D^S}{ds} J\right| \quad$ imply $\quad K^S \leq K^M \leq \Delta$.

Now ist suffices to consider T as a convex triangle on the surface S with curvature bound $K^S \leq \Delta$. It is technically simpler to work also with a lower curvature bound, say δ , although it may not exist if $\dot{f}$ or $\dot{g}$ vanish somewhere. Aleksandrow handles the general case [2]; we assume the lower bound and omit a final limit argument.

<u>Step 1</u>.



Let $T = T' \cup T''$ be a division of T into subtriangles. 6.4.3 shows that $T'_\Delta \cup T''_\Delta$ is a (usually) nonconvex quadrilateral in the plane M_Δ and, with some spherical or hyperbolic geometry, we get

$$\text{area } (T'_\Delta \cup T''_\Delta) \leq \text{area } (T_\Delta) .$$

Therefore, if the proposition is true for T' and T'' it follows for $T' \cup T'' = T$.

Repetition of this argument shows that it is enough to prove the claim for very small triangles. The dividing segment in T is not longer than the dividing segment in T_Δ (6.4.4); therefore the triangles become small, under subdivision, at least as fast as in the constant curvature plane M_Δ .

<u>Step 2</u>. It suffices to prove for small triangles with circumference u

$$(*) \qquad \text{area}\ (T) \leq \text{area}\ (T_\Delta) \cdot (1 + \text{const} \cdot u^2)\ ,$$

since T can be subdivided, by joining the midpoints of its edges, into four triangles with circumferences roughly $\frac{1}{2}u(T)$. Application of step 1 improves $(*)$ to

$$\text{area}\ (T) \leq \text{area}\ (T_\Delta) \cdot (1 + \tfrac{1}{4} \cdot \text{const} \cdot u^2)\ .$$

By repetition $\text{area}\ (T) \leq \text{area}\ (T_\Delta)$ follows.

<u>Step 3</u>. Use $\exp_{p_\Delta} \circ \exp_p^{-1}$ to map the triangle T into M_Δ ; the map is a radial isometry which, in a ball of radius $\frac{1}{2}u(T)$, decreases lengths but not more than by a factor $\mathbf{s}_\Delta(\frac{1}{2}u) \cdot \mathbf{s}_\delta(\frac{1}{2}u)^{-1} = (1 - \text{const} \cdot u^2)$. Consequently also areas are decreased but not by more than a factor $(1 - \text{const} \cdot u^2)$!

To complete the proof we show that the area of the image of T in M_Δ is $\leq \text{area}\ (T_\Delta)$; this then gives

$$\text{area}\ (T_\Delta) \geq \text{area}\ (\text{image}\ T) \geq \text{area}\ (T) \cdot (1 - \text{const} \cdot u^2)$$

which was needed in step 2 .

We describe the image of T as follows: Subdivide the edge qr of T to obtain a division $T = T_1 \cup \ldots \cup T_n$. The trigonometry of step 1 gives

$$\text{area}\ (T_{1\Delta} \cup \ldots \cup T_{n\Delta}) \leq \text{area}\ (T_\Delta)\ .$$

As the subdivision is made finer the polygon $T_{1\Delta} \cup \ldots \cup T_{n\Delta}$ converges to $\exp_{p_\Delta} \circ \exp_p^{-1}(T)$ since $\alpha_{i\delta} \leq \alpha_i \leq \alpha_{i\Delta}$ (6.4.3) and $\frac{\alpha_{i\Delta}}{\alpha_{i\delta}}$ converges to 1 (by trigonometry).

6.7.1 Remark. Among all triangles with two edgelengths a,b on the unit sphere the maximal area is realized by the "Thales triangle", where the angle between a and b equals the sum of the two others. The area is $2 \arcsin(\mathrm{tg}\,\frac{1}{2} a \cdot \mathrm{tg}\,\frac{1}{2} b)$. If $a,b \leq \frac{\pi}{3}$ then area $\leq 1.25 \cdot \frac{ab}{2}$.

7. Lie groups

In this section we treat metric properties of Lie groups in a fairly self-contained way. In the compact case we emphasize Finsler metrics since they offer advantages in applications without requiring more complicated proofs. The results are needed for $SO(n)$, but if one works with the left invariant connection the specialization from compact groups to $SO(n)$ does not simplify proofs. In the noncompact case we need

(i) results on the group of motions of $\mathbb{R}^n$ which follow from properties of $SO(n)$ and

(ii) results on left invariant metrics on nilpotent Lie groups, some of which we can prove only with detailed estimates of remainder terms in the Campbell-Hausdorff formula. Almost flat metrics are given in 7.7.2.

7.1 Basic notions. The tangent space T_eG of a Lie group G at the identity e is made into a Lie algebra as follows:

Extend tangent vectors $X, Y \in T_eG$ via left translations $L_a : G \to G$, $L_a(g) = a \cdot g$, to left invariant vectorfields

$$X^*(a) := (dL_a)_e \cdot X ,$$

observe that the bracket of left invariant vectorfields is left invariant, $[X^*,Y^*](a) = dL_a \cdot [X^*,Y^*](e)$, and put

$$[X,Y] := [X^*,Y^*](e) .$$

The differential of the conjugation $K_a : G \to G$, $K_a(g) = a\,g\,a^{-1}$ is the adjoint representation

$$Ad : G \to GL(T_eG) , \; Ad(a) := d(K_a)_e .$$

The 1-parameter subgroup with initial tangent X is denoted $\exp tX$ ($= e^{tX}$ for brevity); for example if S is a skewsymmetric matrix then $e^{tS} = id + \sum_{k=1}^{\infty} \frac{(tS)^k}{k!}$ is a 1-parameter subgroup in $SO(n)$.

It is fundamental that the differential of Ad at e is the bracket:

7.1.1
$$\frac{d}{dt}\,(d(K_{\exp tX})_e)\big|_{t=0} = d(\mathrm{Ad})_e\cdot X =: \mathrm{ad}\,X$$
$$\mathrm{ad}\,X(Y) = [X,Y] .$$

Immediate consequences are

7.1.2
$$\mathrm{Exp}(t\,\mathrm{ad}\,X) := \mathrm{id} + \sum_{k=1}^{\infty} \frac{1}{k!}(t\,\mathrm{ad}\,X)^k = d(K_{\exp tX})_e ,$$
$$dK_a\cdot[X,Y] = [dK_a\cdot X,\ dK_a\cdot Y] ,$$
$$a\cdot\exp tX = \exp(t\cdot dK_a\cdot X)\cdot a ,$$
$$[\exp X,\ \exp Y]_G = \exp((\mathrm{Exp}\,\mathrm{ad}\,X)\cdot Y)\cdot\exp(-Y) .$$

7.2 Lie bracket and curvature. On each Lie group we have the left invariant connection D^L for which left invariant vectorfields are parallel. The curvature tensorfield vanishes, since on left invariant vectorfields obviously

$$R^L(X^*,Y^*)Z^* := D^L_{X^*}D^L_{Y^*}Z^* - D^L_{Y^*}D^L_{X^*}Z^* - D^L_{[X^*,Y^*]}Z^* = 0 .$$

The torsion field T is obtained by left translating the bracket from T_eG :

7.2.1
$$T(X^*,Y^*) := D^L_{X^*}Y^* - D^L_{Y^*}X^* - [X^*,Y^*] = -[X,Y]^* .$$

The following torsion free connection 7.2.2 is the Levi-Civita connection of biinvariant Riemannian metrics; it is also useful for other metrics.

7.2.2
$$D_X Y := D^L_X Y - \frac{1}{2}T(X,Y) .$$

Its left invariant curvature tensor field is expressible by the Lie bracket and also D-parallel:

7.2.3 $$R(X^*,Y^*)Z^* = \frac{1}{4}\left[Z^*,[X^*,Y^*]\right];$$

$$DR = 0 \text{ (from } [u,[v,w]] + [v,[w,u]] + [w,[u,v]] = 0) .$$

Both connections have the 1-parameter subgroups and their translates (left or right) as geodesics; therefore they have the same Jacobi fields obtained by differentiating families of geodesics.

7.2.4 Let $J(t)$ be a Jacobi field along the geodesic $c(t) = a\cdot\exp tX$. Using left translations we describe it as

$$J(t) = dL_{c(t)}\cdot k(t) \ , \ k : \mathbb{R} \to T_eG .$$

The Jacobi equation $D^L_{\dot c}(D^L_{\dot c} J + T(J,\dot c)) = 0$ is then reduced to the ordinary differential equation $\ddot k + [X,\dot k] = 0$ with the explicit solution

$$\dot k(t) = \mathrm{Exp}(t \operatorname{ad} X)\cdot\dot k(0) ,$$

$$k(t) = k(0) + f(-t \operatorname{ad} X)\cdot t\dot k(0) , \text{ where } f(z) = \frac{1}{z}(e^z-1) .$$

7.2.5 If one works with the biinvariant connection D , then one can translate these solutions, since D-parallel translation along c can be written as

$$P_t = dL_{c(t)}\cdot\mathrm{Exp}(-\frac{t}{2}\operatorname{ad} X) ,$$

into formulas closely resembling the Riemannian case:

$$J(t) = P_t\left(\cosh(\frac{t}{2}\operatorname{ad} X)\cdot J(0) + \frac{\sinh}{\mathrm{id}}(\frac{t}{2}\operatorname{ad} X)\cdot t\frac{D}{dt}J(0)\right) ,$$

$$\frac{D}{dt}J(t) = P_t\left(\sinh(\frac{t}{2}\operatorname{ad} X)\cdot\left[\frac{1}{2}X,J(0)\right] + \cosh(\frac{t}{2}\operatorname{ad} X)\cdot\frac{D}{dt}J(0)\right).$$

7.2.6 The arguments which gave 6.3.2 or 6.6.3 carry over:

(i) $$(d\exp)_{tX}\cdot Y = \frac{1}{t}J(t) ,$$

with J a Jacobi field along the geodesic $t \to \exp tX$ satisfying

$$J(o) = 0 \, , \; \frac{D}{dt} J(o) = \frac{D^L}{dt} J(o) = Y \, ;$$

(ii)
$$\frac{d}{dt} (\exp^{-1}_{q(t)} p) = \frac{D}{ds} J_t(1) \, ,$$

with J_t a Jacobi field along $s \to \exp_{q(t)} ((1-s) \cdot \exp^{-1}_{q(t)} p)$ satisfying

$$J_t(o) = 0 \, , \; J_t(1) = \dot{q}(t) \, .$$

This explicit knowledge of derivatives of exp is the key for more careful studies of the relation between a Lie group and its Lie algebra. One immediate application which we need is the

7.2.7 Proposition. (Differential equation for the Campbell-Hausdorff-formula)

Define $H : [0,1] \to T_eG$ by

$$\exp H(t) = \exp X \cdot \exp t y \, ,$$

where X and Y are such that we stay in a domain where exp is injective. Then

$$\frac{d}{dt} H(t) = g(-\mathrm{ad}\, H(t)) \cdot Y \, , \quad H(o) = X \, ,$$

where
$$g(Z) = Z \cdot (e^Z - 1)^{-1} = 1 - \frac{1}{2} Z + \sum_{k=1}^{\infty} B_{2k} \frac{Z^{2k}}{(2k)!} \, , \; [21] .$$

The solution as a power series in X and tY is the Campbell-Hausdorff-formula; its first few terms are

$$H(X,tY) := H(t) = X + tY + \frac{t}{2} [X,Y] + \frac{t}{12} [X,[X,Y]] + \frac{t^2}{12} [Y,[Y,X]] \ldots$$

Proof. Define a family of geodesics $c_t(s) = \exp(s \cdot H(t))$ with $\frac{\partial}{\partial s} c_t(o) = H(t)$. The induced Jacobi fields $J_t(s) = \frac{\partial}{\partial t} c_t(s)$ have the initial data $J_t(o) = o$, $\frac{D}{ds} J_t(o) = \frac{d}{dt} H(t)$. Since in this case $J_t(1) = dL_{c_t(1)} \cdot Y$ we obtain from 7.2.4 $Y = f(-\operatorname{ad} H(t)) \cdot \frac{d}{dt} H(t)$, where $f(-\operatorname{ad} H)$ is the inverse of $g(-\operatorname{ad} H)$. The analytic differential equation can be solved with the Picard-Lindelöf iteration in any Lie subalgebra; the first few terms are easily obtained starting from

$$H_o(t) = X+tY , \quad H_1(t) = X + \int_o^t g(-\operatorname{ad} H_o(t))Y dt .$$

We now turn to metric considerations.

7.3 <u>Norms</u>. Any norm $\|\ .\ \|$ on T_eG can be left translated to all other tangent spaces of G and thus gives a Finsler metric on G . We use this in particular for the linear and affine isometry group of $\mathbb{R}^n = T_pM$; to our curvature control of parallel translations in 6.2 the following norms are well adapted:

7.3.1 <u>Example</u>. For $S \in so(n)$ put

$$\|S\| := \max\{|Sv|_{\mathbb{R}^n} ; v \in \mathbb{R}^n , |v| = 1\} .$$

One has

$$\|[S,T]\| = \|ST - TS\| \leq 2\|T\| \cdot \|S\| .$$

The corresponding (biinvariant) metric on $SO(n)$ (and $O(n)$) is

$$d(A,B) := \max\{|\sphericalangle(Av,Bv)| ; v \in \mathbb{R}^n , |v| = 1\} .$$

Diameter and injectivity radius (!) of exp are $= \pi$.

From this metric on $O(n)$ we derive a left invariant metric on the group of motions of T_pM . We write motions as

$$\tilde{A}_i(v) = A_i \cdot v + a_i \quad (A_i \in O(n) , a_i \in \mathbb{R}^n)$$

and define

7.3.2 $$\tilde{d}(\tilde{A}_1,\tilde{A}_2) := \max(d(A_1,A_2)\ ,\ c\Lambda|a_1 - a_2|) = \tilde{d}(\tilde{A}_2^{-1}\circ\tilde{A}_1\ ,\ \mathrm{id})\ ,$$

where c is an adjustable parameter, the decrease of which decreases the influence of the translational parts on the definition. The additional factor Λ is a curvature bound for $M(|K| \leq \Lambda^2)$; its inclusion makes the distance on the group of motions of T_pM independent of trivial changes (by constant factors) of the Riemannian metric which is desirable in pinching situations.

If we describe the tangent space of the group of motions by $so(n) \times \mathbb{R}^n$ then the corresponding norm is

7.3.3 $$||(S,a)|| = \max(\ ||S||,\ c\Lambda|a|)\ .$$

Abbreviate $||A|| := d(A,\mathrm{id})$, $||\tilde{A}|| := \tilde{d}(\tilde{A},\mathrm{id})$, abusing notations.

The use of these Finsler metrics simplifies many estimates in the proof of Gromov's theorem; various bounds become independent of the dimension. The following estimates are just as easily proved for Finsler- as for Riemannian metrics. The only extra work is the proof of 7.5.1, namely that the 1-parameter subgroups are in the (biinvariant) Finsler case distance minimizing curves. (This is trivial in the Riemannian case, since the connection 7.2.2 is torsion free and metric, hence the Levi-Civita connection of any biinvariant Riemannian metric.)

7.3.4 In the biinvariant case conjugation is a (Finsler) isometry, it follows that the power series

$$\mathrm{Exp}(\mathrm{ad}\ X)\ :\ T_eG \to T_eG$$

is a norm-isometry: $||\mathrm{Exp}(\mathrm{ad}\ X)\cdot Y|| = ||Y||$. This makes many results look as in the Riemannian case; 7.4.1 and 7.4.2 are Rauch type estimates.

7.3.5 We also make the convention, that we always use the maximum norm for linear maps; given any norm $||\ .\ ||$ on T_eG this leads to the (semi-)

norm (which is biinvariant if $\|\ .\ \|$ is biinvariant)

$$|||X||| := \|\mathrm{ad}\ X\| := \max\{\ \|\mathrm{ad}\ X\cdot Y\|\ ;\ \|Y\| = 1\}\ .$$

It is useful since $\mathrm{ad}[X,Y] = \mathrm{ad}\ X \cdot \mathrm{ad}\ Y - \mathrm{ad}\ Y \cdot \mathrm{ad}\ X$ implies a dimension independent bound for the bracket:

$$|||\ [X,Y]\ ||| \leq 2\,|||X|||\cdot|||Y||| \quad \text{or} \quad |||\mathrm{ad}\ ||| \leq 2\ .$$

7.4 Jacobi-field estimates. We treat the biinvariant and the general case seperately. Proofs are immediate because of the explicit solutions. On the convex ball $B_r = \{X \in T_eG\ ;\ \|\mathrm{ad}\ X\| \leq r\}$ we will have uniform estimates; they resemble the Riemannian estimates.

7.4.1 Lemma. (Upper bounds, compare 6.3.8 (ii)). Let $\|\ .\ \|$ be a biinvariant norm and $J(t) = dL_{c(t)}\cdot k(t)$ a Jacobi field along c . Then 7.2.4, 7.3.4 imply $\|\dot{k}(t)\| = \|\dot{k}(o)\|$, hence

$$\|J(t) - dL_{c(t)}\ J(o)\| \leq t\cdot\|\frac{D^L}{dt} J(o)\|\ .$$

In particular 7.2.6 (i) gives

$$\|(d\ \exp)_{tX}\cdot Y\| \leq \|Y\|\ ,$$

i.e. exp does not increase lengths.

7.4.2 Lemma. (Lower bounds, compare 6.3.5 (iii)). Let $\|\ .\ \|$ be a biinvariant norm and $J(t) = P_t(\frac{\sinh}{\mathrm{id}}(\frac{t}{2}\ \mathrm{ad}\ X)\cdot(t\ \frac{D}{dt}J(o)))$ a Jacobi field (7.2.5) along $c(t) = \exp tX$. Then, if $\|t\ \mathrm{ad}\ X\| < 2\pi$,

$$\frac{1}{t}\ \|J(t)\| \geq \frac{\sin}{\mathrm{id}}\ (\ \|\frac{t}{2}\ \mathrm{ad}\ X\|)\cdot\|\frac{D}{dt}J(o)\|\ ,$$

i.e. $(d\ \exp)_{tX}$ decreases lengths, but not more than by a factor

$$\frac{\sin}{\mathrm{id}}\ (\ \|\frac{t}{2}\ \mathrm{ad}\ X\|)\ .$$

Proof. The power series $\frac{z}{\sinh z}$ is an alternating power series in even powers with radius of convergence π [21 S. 121]. Therefore

$$\left\|\left(\frac{\sinh}{\mathrm{id}}\left(\frac{t}{2}\operatorname{ad} X\right)\right)^{-1}\right\| = \left\|\frac{\mathrm{id}}{\sinh}\left(\frac{t}{2}\operatorname{ad} X\right)\right\| \le \frac{\mathrm{id}}{\sin}\left(\left\|\frac{t}{2}\operatorname{ad} X\right\|\right)$$

proves the estimate.

7.4.3 Lemma. (Comparison of J and $\frac{D}{dt}J$, compare 6.3.6). Under the assumptions of 7.4.2 we have with a similar proof

$$\left\|J(t) - t\frac{D}{dt}J(t)\right\| \le \left(1 - \frac{\mathrm{id}}{\tan}\left(\left\|\frac{t}{2}\operatorname{ad} X\right\|\right)\right)\cdot\|J(t)\| .$$

In the left invariant case one has to decide whether it is more convenient to work with the exponential map of the group or the metric. The metric exponential is always surjective while the other one may not; on the other hand, the relations between the metric exponential and the group structure are not so easily exploited. If the group is simply connected and nilpotent then the group exponential is a diffeomorphism (7.7.4) while the metric exponential may not. In this case the choice seems clear and the following estimates will be needed.

7.4.4 Lemma. (Lipschitz estimates for the group exponential in terms of a left invariant norm $\|\,.\,\|$.) Recall from 7.2.4, 7.2.6:

$$(d\exp)_{tX}\cdot Y = dL_{\exp tX}\cdot f(-t\operatorname{ad} X)\cdot Y ,$$

where $f(z) = \frac{1}{z}(e^z-1)$. Therefore

$$\|(d\exp)_{tX}\cdot Y - dL_{\exp tX}\cdot Y\| \le (f(\|t\operatorname{ad} X\|)-1)\cdot\|Y\|,$$

$$(2 - f(\|t\operatorname{ad} X\|))\cdot\|Y\| \le \|(d\exp)_{tX}\cdot Y\| \le f(\|t\operatorname{ad} X\|)\cdot\|Y\| .$$

Proof. Since the power series f has positive coefficients

$$\|f(-t\operatorname{ad} X)-\mathrm{id}\| \le f(\|t\operatorname{ad} X\|) - 1 .$$

7.4.5 Lemma. (Estimates for $q \to \exp_q^{-1} p$, compare 6.4.6, 7.2.6, 7.4.3) Assume that on some neighborhood $B(p)$ the vectorfield

$$v : B \to TB \quad , \quad v(q) := \exp_q^{-1} p$$

is defined. Let D be the biinvariant, torsion free connection 7.2.2 and abbreviate $r(q) := ||\mathrm{ad}\ v(q)||$ (after left translation of $v(q)$ to the identity and using any left invariant norm $||\ .\ ||$). Then

$$||D_X v - X|| \leq (1 - \frac{\mathrm{id}}{\tan}(\frac{1}{2} r(q))) \cdot ||X|| \ .$$

Proof. Because of 7.2.6 (ii) we only have to estimate $||\frac{D}{\partial s} J(1) - J(1)||$ for which we use again explicite formulas (7.2.5):

$$\frac{D}{ds} J(1) - J(1) = P_s\Big((\frac{\mathrm{id}}{\tanh} - \mathrm{id}) \cdot \frac{\sinh}{\mathrm{id}} (\frac{1}{2}\ \mathrm{ad}\ \exp_p^{-1} q) \cdot \frac{D}{ds} J(o)\Big) \ ,$$

hence

$$||\frac{D}{ds} J(1) - J(1)|| \leq ||(\frac{\mathrm{id}}{\tanh} - \mathrm{id})(\frac{1}{2}\ \mathrm{ad}\ \exp_p^{-1} q)|| \cdot ||J(1)|| \ .$$

As in 7.4.2, if we change the coefficients of the power series $(\frac{\mathrm{id}}{\tanh} - \mathrm{id})$ to their absolute values we obtain the power series $(\mathrm{id} - \frac{\mathrm{id}}{\tan})$, this and $||\mathrm{ad}\ \exp_p^{-1} q|| = ||\mathrm{ad}\ \exp_q^{-1} p||$ complete the proof.

7.5 Metric results in the biinvariant case

7.5.1 Proposition. (Shortest curves) The one-parameter subgroups and their translates locally minimize Finsler distances in the biinvariant (!) case.

Proof. (i) For any (left) invariant metric one has $d(\mathrm{id}, \exp X) \leq$ length $(t \to \exp tX)|_o^1 = ||X||$.
(ii) Arbitrary differentiable curves can be approximated by piecewise geodesic curves such that also the length is approximated.
(iii) Therefore it suffices to prove the triangle inequality for the lengths of geodesic connections (not for distances!), that is we have to prove (see 7.2.7)

7.5.2 $$\|H(X,Y)\| \leq \|X\| + \|Y\| .$$

We use the curve $H : [o,1] \to T_eG$ given by $\exp H(t) = \exp X \cdot \exp tY$, subdivide $[o,1]$ into intervalls $[t_i, t_{i+1}]$ $(i=o,\dots,n)$ of length $\frac{1}{n}$ and use

$$H(t_{i+1}) = H(t_i) + \int_{t_i}^{t_{i+1}} g(-\operatorname{ad} H(t)) \cdot Y \, dt$$
$$= H(t_i) + \frac{1}{n} g(-\operatorname{ad} H(t_i)) \cdot Y + O(n^{-2}) .$$

To have the remainder term as $O(n^{-2})$ one needs $\|\operatorname{ad} H(t)\| < \pi$ for all $t \in [o,1]$ since the radius of convergence of g is π. To prove 7.5.2 we may multiply the norm by a constant and assume without loss of generality $\|[X,Y]\| \leq \|X\| \cdot \|Y\|$ or $\|\operatorname{ad} X\| \leq \|X\|$. We first prove 7.5.2 for all those $X,Y \in T_eG$ for which $\|H(t)\| < \pi$ holds, then we remove this restriction as follows:
The set $A := \{(X,Y) \in T_eG \times T_eG \; ; \; \|X\| + \|Y\| < \pi \text{ and } \|H(t)\| < \pi\}$ is clearly nonempty and open, and if $(X,Y) \in A$ then $(X,tY) \in A$ $(o \leq t \leq 1)$; also, if we have 7.5.2 on A, then A is closed in the convex set $B = \{(X,Y) \in T_eG \times T_eG \; ; \; \|X\| + \|Y\| < \pi\}$. Hence $A = B$.

Finally

$$\|H(1)\| - \|X\| = \sum_{i=o}^{n-1} \|H(t_{i+1})\| - \|H(t_i)\| ,$$
$$\|H(t_i)\| = \|\operatorname{Exp}(\tfrac{1}{n} \operatorname{ad} V_i) \cdot H(t_i)\| \quad (7.3.4) .$$

If we choose $V_i = -\frac{1}{2} Y - \left(\sum_{k=1}^{\infty} B_{2k} \frac{(\operatorname{ad} H(t_i))^{2k-1}}{(2k)!} \right) \cdot Y$ (7.2.7), then

$$g(-\operatorname{ad} H(t_i)) \cdot Y - [V_i, H(t_i)] = \frac{1}{n} \cdot Y , \quad \text{hence}$$
$$\|H(t_{i+1}) - \operatorname{Exp}(\tfrac{1}{n} \operatorname{ad} V_i) \cdot H(t_i)\| = O(n^{-2}) ,$$
$$\|H(1)\| - \|X\| \leq \sum_{i=o}^{n-1} \|H(t_{i+1}) - \operatorname{Exp}(\tfrac{1}{n} \operatorname{ad} V_i) \cdot H(t_i)\|$$
$$\leq \|Y\| + O(\tfrac{1}{n}) .$$

7.5.3 Corollary. (Comparison of distances in T_eG and G). Let $\|\ \bullet\ \|$ be any biinvariant norm or seminorm such that $\|\mathrm{ad}\,X\| \le \|X\|$. Assume that exp is injective on $B_{2r} = \{X \in T_eG\ ;\ \|X\| \le 2r\}$ and assume $r < \pi$. Then

(i) $\exp(B_r)$ is geodesically convex in G.

(ii) For $X,Y \in B_r$ holds $\frac{2\sin r/2}{r}\cdot\|X-Y\| \le d(e^X,e^Y) \le \|X-Y\|$.

Proof. (ii) follows from 7.4.1, 7.4.2 and (i). For (i) it suffices to show (because of continuity): If $X,Y \in B_r$ then the midpoint of a minimizing geodesic from $\exp X$ to $\exp Y$ is also in $\exp B_r$. Let $\exp Z$ be this midpoint, then e is midpoint of $\exp(-Z)\cdot\exp X$ and $\exp(-Z)\cdot\exp Y$, i.e. $\exp(-Z)\cdot\exp X = (\exp(-Z)\cdot\exp Y)^{-1}$ or $\exp 2Z = \exp Z\,\exp Y\,\exp(-Z)\,\exp X =$ $= \exp(\mathrm{Exp}(\mathrm{ad}\,Z)\cdot Y)\cdot\exp X$. Since exp is injective on B_{2r} we get from 7.2.7, 7.5.2, 7.3.4 $2Z = H(\mathrm{Exp}(\mathrm{ad}\,Z)\cdot Y,X)$,

$$\|2Z\| \le \|\mathrm{Exp}(\mathrm{ad}\,Z)\cdot Y\| + \|X\| = \|Y\| + \|X\|, \text{ i.e. } Z \in B_r .$$

7.5.4 Proposition. (Commutators in the group and in its Lie algebra)

(i) $d(e^Xe^Y,e^Ye^X) \le \|[X,Y]\|$,

(ii) $d(e^Xe^Y,e^{X+Y}) \le \frac{1}{2}\|[X,Y]\|$,

(iii) Under the assumptions of 7.5.3 the first remainder of the Cambell-Hausdorff power series satisfies (compare 6.6)

$$\|H(X,Y) - (X+Y)\| \le \frac{r}{4\cdot\sin r/2}\cdot\|[X,Y]\|.$$

(iv) If $\|Y\| \le r$ (as in 7.5.3) then

$$d(e^Xe^Ye^{-X}e^{-Y},e^{[X,Y]}) \le \|[X,Y]\|\cdot(\|\mathrm{ad}\,X\| + \|\mathrm{ad}\,Y\|)\frac{r}{4\sin r/2}\,e^{\|\mathrm{ad}\,X\|}.$$

Proof. (i) The curve $t \to c(t) = \exp(\mathrm{Exp}(t\,\mathrm{ad}\,Y)\cdot X)$ from e^X to $e^Ye^Xe^{-Y}$ has length $\le \|[X,Y]\|$ since

$$\|\tfrac{d}{dt}\mathrm{Exp}(t\,\mathrm{ad}\,Y)\cdot X\| = \|\mathrm{Exp}(t\,\mathrm{ad}\,Y)\cdot[Y,X]\| = \|[Y,X]\|$$

and since exp does not increase distances, 7.4.1 .

(ii) $d\left((e^{X/n}\cdot e^{Y/n})^n, (e^{X/n})^n\cdot(e^{Y/n})^n\right) \leq \frac{n(n-1)}{2} d(e^{X/n}e^{Y/n}, e^{Y/n}e^{X/n})$

$\leq \frac{1}{2} \|[X,Y]\|$ (i) ,

$d((e^{X/n}\cdot e^{Y/n})^n, (e^{(X+Y)/n})^n) \leq n\cdot d(e^{H(X/n,Y/n)}, e^{(X+Y)/n})$

$\leq n\cdot \|H(X/n,Y/n) - (X+Y)/n\|$ (7.4.1)

$\leq O(1/n)$ (Campbell-Hausdorff) .

(iii) follows from (ii) and 7.5.3 .

(iv) Since $e^X e^Y e^{-X} e^{-Y} = \exp(\mathrm{Exp}(\mathrm{ad}\cdot X)\cdot Y)\ e^{-Y}$ we have from 7.4.1 and (iii)

$$d(e^X e^Y e^{-X} e^{-Y}, e^{[X,Y]}) \leq \|H(\mathrm{Exp}(\mathrm{ad}\ X)\cdot Y,-Y) - [X,Y]\|$$

$$\leq \|H(\mathrm{Exp}(\mathrm{ad}\ X)\cdot Y,-Y) - (e^{\mathrm{ad}\ X}\cdot Y-Y)\| + \|(e^{\mathrm{ad}\ X}\cdot Y-Y) - [X,Y]\|$$

$$\leq \frac{r}{4\sin r/2}\|[X,Y]\|(\|\mathrm{ad}\ X\|+\|\mathrm{ad}\ Y\|)\cdot f(\|\mathrm{ad}\ X\|)$$

Remarks. Biinvariance was used heavily in the proofs of 7.4.1 to 7.5.4 ; the results are close to what one would expect from the first terms of the Campbell-Hausdorff formula. If one has only left invariance then all estimates contain additional exponental factors (compare 7.4.4 to 7.4.1 and 7.4.2); we do not need these in general, instead we will come back to left invariant metrics on simply connected nilpotent Lie groups with rather special additional assumptions.

7.6 Applications to O(n) and to the motions of $\mathbb{R}^n$.

In the following two propositions which are important tools in the proof of Gromov's theorem we illustrate how the use of Finsler mectrics improves the dimension dependence of results (without other changes of the arguments).

7.6.1 Proposition. (Pairwise distances in $O(n)$)

(i) For given $\phi \in (0,\pi)$ there are at most $2 \cdot (2\pi/\phi)^{\dim SO(n)}$ elements in $O(n)$ with pairwise distance $\geq \phi$.

(ii) For given $A \in O(n)$ there are at most $m = 2 \cdot (2\pi/\phi)^{[n/2]}$ iterates A^j $(j = 0,\dots,m-1)$ with pairwise distance $\geq \phi$; in particular there exists $k \leq m$ such that $d(A^k, id) < \phi$.

Proof. Since open balls $B_{\phi/2}$ of radius $\phi/2$ around elements of pairwise distance $\geq \phi$ are disjoint and since exp maps a ball of radius π onto $SO(n)$ (use the metric 7.3.1!) we have the volume ratio

$$2\cdot \mathrm{vol}(SO(n)) \,/\, \mathrm{vol}(B_{\phi/2}) \;=\; 2\cdot \mathrm{vol}(B_\pi) \,/\, \mathrm{vol}(B_{\phi/2})$$

as an obvious bound for the number of elements with pairwise distance $\geq \phi$. The iterates A^j are not arbitrary in $O(n)$ but lie on two tori of dimension $\leq [n/2]$ which are flat and totally geodesic in the two components of $O(n)$. Since the exponential map restricted to the tangent spaces of these tori is length- (and therefore volume-) preserving we have immediately the explicit bound $2\cdot(2\pi/\phi)^{\dim \text{torus}}$. If $d(A^{j_1}, A^{j_2}) < \phi$ with $0 \leq j_1 < j_2 \leq m$, put $k = j_2 - j_1$. - To get the explicit bound in (i) we estimate $\det(d\exp)_{tX}$ by using pairwise orthogonal (in the standard Riemannian metric) Jacobi fields which start in eigenspaces of the nonpositive symmetric operator $(\mathrm{ad}\, X)^2$. Because of $DR = 0$ and $R(J,X)X = \frac{-1}{4}(\mathrm{ad}\, X)^2 \cdot J$ (7.2.3) we have:

$|J|^{-1}\cdot J$ is parallel (where $|J| \neq 0$);

$\|J(t)\| = \|\frac{D}{dt} J(0)\| \cdot \frac{\sin}{id}(\frac{t}{2}\lambda_i)$, where λ_i is eigenvalue of $-(\mathrm{ad}\, X)^2$;

$\lambda_i \leq 2\pi$ since $\|\mathrm{ad}\, X\| \leq 2\|X\| \leq 2\pi$ (with 7.3.1).

Therefore

$$\det(d\exp)_{tX} = \prod_{i=1}^{\dim SO(n)-1} \frac{\sin}{id}\left(\frac{t}{2}\lambda_i\right) \quad \text{decreases with } t \,,$$

hence $$\mathrm{vol}(B_\pi) \leq \mathrm{vol}(B_{\phi/2}) \cdot (2\pi/\phi)^{\dim SO(n)} .$$

The following modification to a noncompact (!) situation is used in section 3 to establish a crucial a priori bound.

7.6.2 Proposition. (Pairwise distances in the group of motions)
Let $\tilde{A}_i \neq \mathrm{id}$ $(i = 1,\dots,k)$ be motions with rotational parts in $SO(n)$, i.e.

$$\tilde{A}_i(v) = \exp(S_i)\cdot v + a_i .$$

Assume that the $\tilde{A}_i$ pairwise satisfy

$$\tilde{d}(\tilde{A}_i^{-1}\tilde{A}_j,\mathrm{id}) \geq \max(\tilde{d}(\tilde{A}_i,\mathrm{id})-\varepsilon\tilde{d}(\tilde{A}_j,\mathrm{id}),\ \tilde{d}(\tilde{A}_j,\mathrm{id})-\varepsilon\tilde{d}(\tilde{A}_i,\mathrm{id}))$$

then
$$k \leq \left(\frac{3-\varepsilon}{1-\varepsilon}\right)^{n + \dim SO(n)} .$$

Proof. By definition (7.3.2) $\tilde{d}(\tilde{A}_i,\mathrm{id}) = \max(\|S_i\|, c\Lambda|a_i|)$. Since the exponential map of $SO(n)$ does not increase distances (7.4.1) we also have

$$\tilde{d}(\tilde{A}_i,\tilde{A}_j) \leq \max(\|S_i - S_j\|,\ c\Lambda|a_i - a_j|) .$$

This implies that on the vector space $W = so(n) \times \mathbb{R}^n$ with $\|w\| := \max(\|S\|, c\Lambda|a|)$ we find at least as many w_i satisfying the system of inequalities

$$\|w_i - w_j\| \geq \max(\|w_i\|-\varepsilon\|w_j\|\ ,\ \|w_j\|-\varepsilon\|w_i\|)$$

as we had motions $\tilde{A}_i$. For a fixed pair i,j we may assume (using the homogeneity of the inequalities on the vector space) $1 = \|w_i\| \leq \|w_j\|$. Then

$$\|(\|w_j\|^{-1}\cdot w_j - w_i)\| \geq \|w_j - w_i\| - \|(\|w_j\|^{-1}\cdot w_j - w_j)\|$$

$$\geq \|w_j\|-\varepsilon\|w_i\| - (\|w_j\|-1)$$

$$= \|\,\|w_j\|^{-1}\cdot w_j\| - \varepsilon\|w_i\| .$$

In other words, the unit vectors $w_i / \|w_i\|$ still satisfy the same system of inequalities. Therefore the open balls of radius $(1-\varepsilon)/2$ around these unit vectors are disjoint and contained in the ball of radius $1 + (1-\varepsilon)/2$ around 0 ; the volume ratio $(\frac{3-\varepsilon}{1-\varepsilon})^{\dim W}$ is then a bound on the number of vectors w_i , hence also on the number of motions $\tilde{A}_i$.

7.7 Left invariant metrics.

7.7.1 Proposition. (Lie bracket and curvature of the Levi-Civita connection.) Let

$$D^U_X Y = D^L_X Y + U(X,Y)$$

be the Levi-Civita connection of a left invariant scalar product $\langle\ ,\ \rangle$ on a Lie group G . (D^L is the left invariant connection from 7.2 whose torsion field T is essentially the Lie bracket; U is a left invariant tensor field.)

Then

$$\langle U(X,Y),Z\rangle = \frac{1}{2}\{-\langle T(X,Y),Z\rangle+\langle Y,T(X,Z)\rangle+\langle X,T(Y,Z)\rangle\} ,$$

$$R^U(X,Y)Z = U(X,U(Y,Z))-U(Y,U(X,Z))-U(T(X,Y),Z) ,$$

$$\|R^U\| \leq 6\,\|T\|^2 .$$

(This curvature bound is 24 times the bound in the biinvariant case.)

Proof. Since D^U is torsion free and D^L has torsion T we have

$$U(X,Y) - U(Y,X) = -T(X,Y) .$$

Since D^U and D^L are metric connections, i.e. product differentiations for the scalar product, we have

$$\langle U(X,Y),Z\rangle + \langle Y,U(X,Z)\rangle = 0 .$$

This determines U in terms of T and $<\,,\,>$ as above. The formula for R^U is immediate from the definition if we use $D^U_{Y^*}Z^* = U(Y^*,Z^*)$ for left invariant vectorfields. The curvature bound is then trivial.

7.7.2 Proposition. (Nilmanifolds are almost flat)
On any nilpotent Lie algebra L there exist families of metrics $\|\ \|_\mu$ (7.7.3 (ii)) with the following properties:

(i) $\|\mathrm{ad}\|_\mu$ - and with 7.7.1 also $\|R_\mu\|_\mu$ - is bounded independent of μ.

(ii) The unit balls $D_\mu := \{X \in L\ ;\ \|X\|_\mu \le 1\}$ give an expanding exhaustion of L.

In particular, on any compact quotient of the corresponding nilpotent Lie group there exist ε-flat metrics (1.3) for any $\varepsilon > 0$.

Proof. Choose a "triangular" basis $\{X_1,\dots,X_n\}$ for L, i.e. a basis such that $L_i = \mathrm{span}(X_1,\dots,X_i)$ is for each i an ideal of L and $[L,L_{i+1}] \subset L_i$. Let

$$[X_i,X_j] = \sum_{k<\min(i,j)} \gamma_{ijk}\cdot X_k\ ,\quad \gamma := \max|\gamma_{ijk}|\ .$$

Define the length of $X = \Sigma a_i X_i$ by

7.7.3 (i) $$\|X\|_\mu^2 := \sum_{i=1}^n \mu_i^2 a_i^2\ ,$$

where the μ_i are such that $\sum_{k=1}^{i-1}\mu_k^2 \le \mu_i^4 \le \mu_{i+1}^4$, which is implied by the simpler condition $\mu_{i+1}^4 \ge \mu_i^2 + \mu_i^4$; explicit examples are obtained for any $q \in (0,1)$ with

(ii) $$\mu_i(q) := q(q/\sqrt{2}\,)^{n-i}(i = 1,\dots,n)\ .$$

No matter how small q is taken we conclude from

$$\|[X_i,X_j]\|_\mu^2 \le \gamma^2\cdot \sum_{k<\min(i,j)} \mu_k^2 \le \gamma^2\cdot\mu_i^2\cdot\mu_j^2$$

that the norm of the Lie bracket is $\leq \gamma n$:

$$\|[\Sigma a_i X_i \ , \ \Sigma b_j X_j]\|_\mu \leq \|\sum_{i,j} \mu_i a_i \mu_j b_j \ \frac{[X_i, X_j]}{\mu_i \mu_j}\|_\mu \leq \gamma \sum_{i,j} |\mu_i a_i| \cdot |\mu_j b_j|$$

$$\leq \gamma n \ \|\Sigma a_i X_i\|_\mu \cdot \|\Sigma b_j X_j\|_\mu \ .$$

7.7.4 Proposition. Let N be the nilpotent Lie group with Lie algebra L. The group exponential map

$$\exp : L \to N$$

is a diffeomorphism which, because of 7.2.7, maps subalgebras to subgroups. The difference between left translation and the differential of $\exp$ can be bounded with 7.4.4 using any left invariant norm $(f(z) = z^{-1}(e^z - 1))$.

$$\|(d \exp)_{tX} \cdot Y \ - \ dL_{\exp tX} \cdot Y\| \leq (f(\|t \text{ ad } X\|)-1) \cdot \|Y\|.$$

Proof. $\exp$ has maximal rank on L because of the lower bound 7.4.4 and the existence of left invariant metrics for which the bracket has arbitrarily small norm (7.7.3). This implies that $\exp$ is surjective: $\gamma \in N$ can be joined to $e \in N$ by a curve $\gamma(t)$ $(\gamma(0) = e \ , \ \gamma(1) = \gamma)$; then choose with 7.7.2 a left invariant metric such that $\|X\| \leq 2 \cdot \text{length}(\gamma) =: 2l$ implies $\|\text{ad } X\| \leq \log 1.4$; now γ can be lifted to a curve $\tilde{\gamma}$ in the compact ball of radius $2l$ in L such that $\exp \tilde{\gamma}(t) = \gamma(t)$. (Any lift of $\gamma|_{[0,t)}$ to that ball has length $\leq (2-1.4)^{-1} \cdot l < 2l$, so that the lift exists on a larger interval.) Finally, $\exp_e$ is injective, since the closed curve formed by two geodesics with the same endpoint is homotopic to zero; this homotopy can be lifted to L giving the contradiction of two different rays with the same endpoint.

7.8 Remainder estimates for the Campbell-Hausdorff formula

The following estimates are needed in 5.3 to bound the curvature of a certain left invariant metric on a nilpotent Lie group in terms of the commutator estimates 3.5 .

7.8.1 Lemma. (Bound for the Campbell-Hausdorff formula). Let N be a Lie group with Lie algebra L and a left invariant metric coming from a norm $\|\ .\ \|$ on L. If $\|\mathrm{ad}\ X\| + \|\mathrm{ad}\ Y\| \leq \log(1+\varepsilon) < \log(2-e^{-\pi}) \approx 0.67$ and if $\exp H = \exp X \cdot \exp Y$, then

$$\|\mathrm{ad}\ H\| \leq -\log(2-e^{\|\mathrm{ad}\ X\| + \|\mathrm{ad}\ Y\|}) \leq \log(1+\frac{\varepsilon}{1-\varepsilon}) < \pi ,$$

$$\|\mathrm{ad}\ H\| \leq \frac{e^{\|\mathrm{ad}\ X\| + \|\mathrm{ad}\ Y\|}}{2-e^{\|\mathrm{ad}\ X\| + \|\mathrm{ad}\ Y\|}} \cdot (\|\mathrm{ad}\ X\| + \|\mathrm{ad}\ Y\|) \leq \frac{1+\varepsilon}{1-\varepsilon}(\|\mathrm{ad}\ X\| + \|\mathrm{ad}\ Y\|).$$

(Compare to 7.5.2 and 7.3.5 ; the loss of the factor $\frac{1+\varepsilon}{1-\varepsilon}$ destroys the proof that the 1-parameter subgroups are distance minimizing.)

Proof. $\exp H = \exp X \cdot \exp Y$ implies $\mathrm{Exp}(\mathrm{ad}\ H) = \mathrm{Exp}(\mathrm{ad}\ X) \circ \mathrm{Exp}(\mathrm{ad}\ Y)$. From

$$\|\mathrm{Exp}(\mathrm{ad}\ H) - \mathrm{id}\| =$$

$$\|(\mathrm{Exp}(\mathrm{ad}\ X)-\mathrm{id}) + (\mathrm{Exp}(\mathrm{ad}\ Y)-\mathrm{id}) + (\mathrm{Exp}(\mathrm{ad}\ X)-\mathrm{id}) \circ (\mathrm{Exp}(\mathrm{ad}\ Y)-\mathrm{id})\|$$

$$\leq e^{\|\mathrm{ad}\ X\|} - 1 + e^{\|\mathrm{ad}\ Y\|} - 1 + (e^{\|\mathrm{ad}\ X\|} - 1) \cdot (e^{\|\mathrm{ad}\ Y\|} - 1)$$

$$= e^{\|\mathrm{ad}\ X\| + \|\mathrm{ad}\ Y\|} - 1 \leq \varepsilon < 1$$

we have the convergent power series

$$\mathrm{ad}\ H = \mathrm{Log}(\mathrm{Exp}(\mathrm{ad}\ H)) = \sum_{k=1}^{\infty} (-1)^{k-1} \frac{(\mathrm{Exp}(\mathrm{ad}\ H)-\mathrm{id})^k}{k} ,$$

hence

$$\|\mathrm{ad}\ H\| \leq \sum_{k=1}^{\infty} \frac{\|\mathrm{Exp}(\mathrm{ad}\ H)-\mathrm{id}\|^k}{k} = -\log(1-\|\mathrm{Exp}(\mathrm{ad}\ H)-\mathrm{id}\|) .$$

7.8.2 Lemma. (Remainder estimates for the Campbell-Hausdorff formula.) Let N be a nilpotent Lie group with normed $(\|\ .\ \|)$ Lie algebra L. Let $\{X_1,\dots,X_n\}$ be a triangular basis (as in 7.7.2) for L, $L_i = \mathrm{span}\{X_1,\dots,X_i\}$ the descending sequence of ideals $([L,L_{i+1}] \subset L_i)$.

Let $X \in L$, $Y \in L_{i+1}$, $\|\mathrm{ad}\, X_{|L_i}\| + \|\mathrm{ad}\, Y_{|L_i}\| < \log(2-e^{-\pi}) \approx 0.67$.

Determine $\exp H(t) = \exp X \cdot \exp tY$ with 7.2.7 and 7.7.4 from

$$\dot{H}(t) = g(-\mathrm{ad}\, H(t))\cdot Y, \quad H(o) = X, \quad g(z) = z\cdot(e^z-1)^{-1} = \frac{1}{f(z)}. \quad \text{Then}$$

(i) $\|[H(t)-X,Y]\| \le \|[X,Y]\| \cdot (e^{t\|\mathrm{ad}\, Y_{|L_i}\|}-1)$;

(ii) $\|H(t)-X-tY\| \le \|[X,tY]\| \cdot f(\|t\,\mathrm{ad}\, Y_{|L_i}\|)$,

$\|H(t)\| \le \|tY\| + e^{\|t\,\mathrm{ad}\, Y_{|L_i}\|}\cdot\|X\|$;

(iii) $\|H(e^{\mathrm{ad}\, X}\cdot Y,-Y) - [X,Y]\| \le$

$$\|[X,Y]\| \cdot (\tfrac{1}{2}\|\mathrm{ad}\, X_{|L_i}\| + \|\mathrm{ad}\, Y_{|L_i}\| \cdot f(\|\mathrm{ad}\, Y_{|L_i}\|)) \cdot f(\|\mathrm{ad}\, X_{|L_i}\|).$$

Remark. Note that for $Y \in L_{i+1}$ the estimates are in terms of $\mathrm{ad}\, Y_{|L_i}$, $\mathrm{ad}\, X_{|L_i}$ (compare 7.5.4 (iv)). This is the particular feature of the nilpotent case which allows an induction; $\mathrm{ad}\, X_{|L_1} = 0$.

Proof. We estimate the higher order terms in the power series:

$$\dot{H} = g(-\mathrm{ad}\, H)\cdot Y = Y + \frac{1}{2}[Y,H] + g_2(-\mathrm{ad}\, H)\cdot Y,$$

where $$g_2(z) = \sum_{k=1}^{\infty} B_{2k} \frac{z^{2k}}{(2k)!} = -1 + \frac{z}{2}\cdot\frac{e^z+1}{e^z-1},$$

$$\sum_{k=1}^{\infty} |B_{2k}| \frac{r^{2k}}{(2k)!} = -g_2(ir) = 1 - \frac{r}{2}\,\mathrm{tg}\,\frac{r}{2} \le r^2/8. \quad (\text{if } r \le \pi)$$

From 7.8.1 we have $\|\mathrm{ad}\, H_{|L_i}\| < \pi$, hence with

$$\|\dot{H} - Y - \frac{1}{2}[Y,H]\| = \|g_2(-\mathrm{ad}\, H)\cdot Y\| \le \frac{1}{8}\|\mathrm{ad}\, H_{|L_i}\|\cdot\|[H,Y]\|$$

(a) $$\|\dot{H} - Y\| \le \|[H,Y]\|.$$

Consequently $\|[H,Y]^{\bullet}\| \le \|\mathrm{ad}\, Y_{|L_i}\| \cdot \|[H,Y]\|$, hence $\| [H-X,Y](t) \| \leq \| [X,Y] \| (e^{t \| \mathrm{ad}\, Y|_{L_i} \|} - 1)$, which is (i) .

($\| \dot{f} \| \leq \alpha \cdot \| f \|$ implies

$$\| f(t) - f(o) \| \leq \int_o^t \alpha \| f(\tau) \| d\tau \leq \| f(o) \| \cdot \alpha t + \alpha^2 \int_o^t \int_o^\tau \| f \|$$

$$\leq \| f(o) \| \cdot (e^{\alpha t} - 1) ,$$ by induction.)

The bound $\| [H,Y] \| \leq \| [X,Y] \| \cdot e^{t \| \mathrm{ad}\, Y|_{L_i} \|}$ is inserted in (a) , integration gives (ii) .
For (iii) we use the trivial power series estimates

$$\| [e^{\mathrm{ad}\, X} \cdot Y, -Y] \| \leq \| \mathrm{ad}\, Y|_{L_i} \| \cdot f(\| \mathrm{ad}\, X|_{L_i} \|) \cdot \| [X,Y] \| ,$$

$$\| (e^{\mathrm{ad}\, X} \cdot Y - Y) - [X,Y] \| \leq \frac{1}{2} \| \mathrm{ad}\, X|_{L_i} \| \cdot f(\| \mathrm{ad}\, X|_{L_i} \|) \cdot \| [X,Y] \| ,$$

the triangle inequality and (ii) :

$$\| H(e^{\mathrm{ad}\, X} \cdot Y, -Y) - (e^{\mathrm{ad}\, X} \cdot Y - Y) \| \underset{(ii)}{\leq} \| [e^{\mathrm{ad}\, X} \cdot Y, -Y] \| \cdot f(\| \mathrm{ad}\, Y|_{L_i} \|) .$$

8. Nonlinear averages

In section 3.6.4 we need the result 8.2 on almost homomorphisms. In section 5.4 we need the result 8.3 on the interpolation of locally defined maps. Both tools are based on a generalization of the euclidean center of mass [12], [18].

8.1 The nonlinear center of mass

8.1.1 Mass distribution. Let A be a measure space of volume 1 — more specifically let A either be a finite set of weighted points or a compact subset of a Riemannian manifold, with volume normalized to 1 . B denotes a convex set diffeomorphic to a ball in a manifold M with affine connection D . The center of mass $\mathcal{C}_f$ is to be defined for a measurable map $f : A \to B$; we call f a normalized mass distribution in B .

8.1.2 The euclidean case. If $M = \mathbb{R}^n$ then one can define the center $\mathcal{C}_f$

(i) either as the unique minimum point of the convex function on B

$$P_f : B \to \mathbb{R} \;, \quad P_f(x) := \frac{1}{2} \int_A |x - f(a)|^2 \, da \;,$$

(ii) or as the unique zero of the vectorfield on B

$$\operatorname{grad} P_f(x) = \int_A (x - f(a)) \, da \;,$$

$$\mathcal{C}_f = \int_A f(a) \, da = x - \operatorname{grad} P_f(x) \;.$$

Definition (i) immediately generalizes to the Riemannian case (8.1.3); definition (ii) shows that the center of mass is an affine rather than a metric notion, it can be made to work under more general circumstances (8.1.4) than (8.1.3), for example on a simply connected nilpotent Lie group without a metric.

8.1.3 Definition. (Using convex functions)
Let $B_\rho(p)$ be a ball in a Riemannian manifold M and assume

$$\rho < \text{injectivity radius of } \exp_p$$

$$\rho < \frac{1}{4}\cdot\pi\Delta^{-\frac{1}{2}}, \text{ if } \Delta \text{ is an upper curvature bound on } B_\rho \text{ and if } \Delta > 0 .$$

Let $f : A \to B_\rho$ be a normalized mass distribution (8.1.1). Then, the function

$$P_f : B_\rho \to \mathbb{R} \;, \quad P_f(x) := \frac{1}{2}\int_A d(x,f(a))^2\, da \qquad \text{(compare 8.1.2)}$$

is because of 6.4.6, 6.3.6 an average of convex functions with

(i) $\quad \operatorname{grad} P_f(x) = -\int_A \exp_x^{-1} f(a)\, da \;,$

(ii) $\quad |D_Y \operatorname{grad} P_f - Y| \leq 2(\Lambda\rho)^2 \cdot |Y| \quad (\text{if } |K| \leq \Lambda^2) \;.$

Its unique minimum point $\mathcal{C}_f$ is called the center of the massdistribution $f : A \to B_\rho$ (adapted to the Levi-Civita connection).

The next definition (generalized from 8.1.2 (ii)) aims to define the center as an affine notion. It does not fully succeed since some metric information is used to prove the uniqueness of the center. However the center is independent of the auxilliary metric. See 8.1.6, application to compact Lie groups, and 8.1.8, application to nilpotent Lie groups.

8.1.4 Definition. (Using indices of vectorfields)
Let D be a (symmetric) connection on a manifold M with exponential map $\exp$ and let B be a convex set for the geodesics of D, B diffeomorphic to a ball. For a normalized mass distribution $f : A \to B$ define the vectorfield (compare 8.1.2 (ii))

$$v_f : B \to TB \;, \quad v_f(x) := -\int_A \exp_x^{-1} f(a)\, da \;.$$

The convexity of B implies that the vectorfield v_f points outward along ∂B, therefore v_f has <u>total index 1</u> and consequently <u>at least one zero</u> in B.

Assume that one can prove an estimate for the derivative Dv_f as in 8.1.3 (ii) or 7.4.5 which implies that <u>every zero</u> of v_f has <u>index 1</u>.

Then: The unique zero of v_f in B is called the center $\mathcal{C}_f$ of the mass distribution f adapted to the connection D.
The center of finitely many points $f_i \in B$ with weigths ϕ_i will also be denoted $\mathcal{C}\{f_i, \phi_i\}$.

As in the euclidean case we have

8.1.5 <u>Proposition</u>. (Compatibility properties of the center)
Let $\phi : A \to A$ be a volume preserving map and let $U : M \to M$ be an isometry in case 8.1.3 resp. an affine diffeomorphism in case 8.1.4, then we have for all $x \in B$ directly from the definitions

$$P_{U \circ f \circ \phi}(Ux) = P_f(x) \qquad \text{(in case 8.1.3)},$$

$$v_{U \circ f \circ \phi}(Ux) = dU_x v_f(x) \qquad \text{(in case 8.1.4)},$$

hence
$$\mathcal{C}_{U \circ f \circ \phi} = U(\mathcal{C}_f) \,,$$

i.e. the center commutes with isometries (resp. connection preserving maps) and is unaffected by volume preserving "permutations" of the masses.

8.1.6 <u>Example</u>. (Center on compact Lie groups)
Given any biinvariant Finsler metric on a compact Lie group G (assumed normalized so that $||\mathrm{ad}\ X|| \leq ||X||$), we use the biinvariant (and metric) connection (7.2.2) and have

(i) Convexity of Finsler balls B_ρ of radius $\rho < \pi$ (7.5.3).

(ii) $$||D_X v_f - X|| \le (1 - \frac{\rho}{\tan\rho}) \cdot ||X|| \qquad (7.4.5)$$

$$< ||X|| \quad \text{if} \quad \rho < \frac{\pi}{2} .$$

Therefore 8.1.4 gives a center of mass $\mathcal{C}_f$ (adapted to the biinvariant connection) for any normalized mass distribution

$$f : A \to B = B_\rho \ , \quad \rho < \frac{\pi}{2} .$$

Note that the present balls B_ρ are in general much larger convex sets than balls which satisfy 8.1.3 for Riemannian metrics on G .

It may be helpful to see in a simple application how easily the center can be used:

8.1.7 Proposition. (Conjugation of isomorphic subgroups)
Let H and G be compact Lie groups, H with volume normalized to 1 , G with biinvariant Finsler metric normalized to $||\mathrm{ad}\, X|| \le ||X||$. Let $H' \subset G$ and $H'' \subset G$ be two homomorphic images of H and assume

(*) $$d(h',h'') < \frac{\pi}{2} \quad \text{for all } h \in H .$$

Define $f : H \to B_{\pi/2}(e) \subset G$, $f(h) := (h')^{-1} \cdot h''$.

Then

$$\mathcal{C}_f^{-1} \cdot h' \cdot \mathcal{C}_f = h'' \quad \text{for all } h \in H ,$$

i.e. the center of f conjugates H' and H'' .

Proof. Fix $k \in H$. Left translation by k' and right translation by k'' are isometries of G , right translation by k^{-1} is volume preserving on H , therefore 8.1.5 implies

$$k' \cdot \mathcal{C}_f = \mathcal{C}_{k' \cdot f} = \mathcal{C}_{h \to k' \cdot (h')^{-1} \cdot h''}$$

$$= \mathcal{C}_{h \to ((h \cdot k^{-1})')^{-1} \cdot (h \cdot k^{-1})'' \cdot k''} = \mathcal{C}_f \cdot k'' .$$

Corollary. For a finite group H there are at most $N := (\mathrm{vol}(G)/\mathrm{vol}(B_{\pi/2}))^{|H|}$ maps form H into G such that no two of them satisfy 8.1.7 (✱) . Therefore there are at most N nonconjugate homomorphisms from H into G .

Next we apply 8.1.4 to nilpotent Lie groups where the center can be defined as an affine notion.

8.1.8 Example. (Center on nilpotent Lie groups)
Let C be a compact set in a simply connected nilpotent Lie group N and let D be the biinvariant connection (7.2.2) which has the (left) translates of 1-parameter subgroups as its geodesics.
Any normalized mass distribution $f : A \to C$ (8.1.1) has a center of mass $\mathcal{C}_f$ adapted to the connection D .

Proof. Choose in the following lemma some metric $\|\ \|_\mu$ such that C is contained in the D-convex unit ball $B := \exp D_\mu$. Geodesics, which join two points of B have a μ-length $\leq 1+e^{1/9}$ (7.8.2 (ii)), therefore we have from 7.4.5

$$(✱) \qquad \|D_X v_f - X\|_\mu \leq \left(1 - \frac{\mathrm{id}}{\tan}\left((1+e^{1/9})/18\right) \cdot \|X\|_\mu \leq \frac{1}{90}\|X\|_\mu .$$

The assumptions of 8.1.4 to define the center of f are satistied.

8.1.9 Lemma. (Convex exhaustion of nilpotent Lie groups)
Normalize the almost flat metrics 7.7.3 so that

$$\|\mathrm{ad}\, X\|_\mu \leq \frac{1}{9}\|X\|_\mu \quad \text{by putting} \quad \mu_i := \mu_i(q) \cdot 9\gamma n .$$

Then

(i) The compact unit balls

$$D_\mu := \{X \in L ; \|X\|_\mu \leq 1\}$$

are an exhaustion of the Lie algebra L (as $q \to 0$).

(ii) The balls $B_{\mu,r} := \exp(r \cdot D_\mu)$ are D-convex for $r \leq 2$; the balls $B_{\mu,1}$ are a convex exhaustion of the Lie group N .

Proof. (i) restates 7.7.3. To prove (ii) recall that $\exp_e : L \to N$ is a diffeomorphism (7.7.4) and define

$$h : N \to \mathbb{R} \quad , \quad h(x) := \frac{1}{2} \langle \exp_e^{-1} x \, , \, \exp_e^{-1} x \rangle_\mu \, .$$

Clearly $B_{\mu,r} = h^{-1}([0,\frac{1}{2} r^2])$; we show that h is convex on $B_{\mu,2}$. Consider the set

$$A := \{(x,y) \in B_{\mu,2} \times B_{\mu,2} \text{ ; the geodesic from } x \text{ to } y \text{ is in } B_{\mu,4}\} \, .$$

Obviously A is nonempty and closed. Let $x(t) = \exp X(t)$, $0 \leq t \leq 1$, be any nonconstant geodesic in $B_{\mu,4}$. We shall show $\frac{d^2}{dt^2} h(x(t)) > 0$; then, if the endpoints are in $B_{\mu,r}$ $(r < 4)$, the whole geodesic is also in $B_{\mu,r}$. This proves that A is also open in $B_{\mu,2} \times B_{\mu,2}$ hence $A = B_{\mu,2} \times B_{\mu,2}$. In other words $B_{\mu,2}$ is convex and h is a convex function on $B_{\mu,2}$ (with convex sublevels $B_{\mu,r}$). By definition we have

$$h(x(t)) = \frac{1}{2} \langle X(t), X(t) \rangle_\mu \, , \quad \frac{d}{dt} h \circ x = \langle X, \dot{X} \rangle_\mu \, ,$$

$$\frac{d^2}{dt^2} h \circ x = \langle \dot{X}, \dot{X} \rangle_\mu + \langle X, \ddot{X} \rangle_\mu \, .$$

We shall prove $|\langle X, \ddot{X} \rangle|_\mu \leq 0.93 |\dot{X}|_\mu^2$, implying $\frac{d^2}{dt^2} h \circ x > 0$. From $\dot{x} = (d \exp)_X \cdot \dot{X} = dL_x \cdot f(-\mathrm{ad}\, X) \cdot \dot{X}$ (7.2.4, 7.2.6) and $\frac{D^L}{dt} \dot{x} = 0$ we get

$$(*) \qquad (\frac{d}{dt} f(-\mathrm{ad}\, X)) \cdot \dot{X} + f(-\mathrm{ad}\, X) \cdot \ddot{X} = 0 \, .$$

Termwise differentiation of the power series f and 7.4.4 give

$$\| \frac{d}{dt} f(-\mathrm{ad}\, X(t)) \| \leq f'(\|\mathrm{ad}\, X\|) \cdot \|\mathrm{ad}\, \dot{X}\| \, ,$$

$$\| f(-\mathrm{ad}\, X)^{-1} \| \leq (2 - f(\|\mathrm{ad}\, X\|))^{-1} \, .$$

This is used in $(*)$ with $\|\mathrm{ad}\| \leq \frac{1}{9}$ and $|X| := \rho \leq 4$ to end the proof

$$|\langle X, \ddot{X} \rangle_\mu| \leq \rho \cdot (2 - f(\rho/9))^{-1} \cdot f'(\rho/9) \cdot \frac{1}{9} |\dot{X}|_\mu^2$$

$$\leq 0.93 |\dot{X}|_\mu^2 \quad (\text{if } \rho/9 \leq \frac{1}{2}) \, .$$

In many applications (e.g. 8.2.7, 5.4) one knows a point x which can be expected to be close to the center; with the following result one can exploit such a guess by estimating $d(x,\mathcal{C}_f)$ in terms of $|v_f(x)|$.

8.1.10 Proposition. (Length of $v_f(x)$ and distance to the center)
Let $f : A \to B$ be a normalized mass distribution (8.1.1).

(i) Under the assumption 8.1.3 with $B = B_\rho$ (and recalling 6.3.4) holds

$$|\operatorname{grad} P_f(x)| \geq d(x,\mathcal{C}_f) \cdot 2\rho \frac{\mathbf{s}'_\Delta}{\mathbf{s}_\Delta}(2\rho) ;$$

if in addition $|K| \leq \Lambda^2$ and $\rho \leq \frac{1}{2}\pi\Lambda^{-1}$ then also

$$|\operatorname{grad} P_f(x) + \exp_x^{-1} \mathcal{C}_f| \leq d(x,\mathcal{C}_f) \cdot 2(\Lambda\rho)^2 \quad \text{(see 8.1.2 (ii))}.$$

(ii) In the case 8.1.6 of a compact Lie group with biinvariant Finsler metric holds (recall 7.5.1, 7.5.3)

$$||v_f(x) - \exp_x^{-1} \mathcal{C}_f|| \leq d_{||\ \ ||}(x,\mathcal{C}_f) \cdot (1 - \frac{\rho}{\tan \rho}) .$$

(iii) In the case 8.1.8 of a nilpotent Lie group N we have with any left invariant Finsler metric: If $C \subset \exp D_r$ where $D_r = \{X ; ||X|| \leq r\}$, put $2\tilde{r} := r(1 + e^{r\,||\mathrm{ad}||})$, then

$$||v_f(x) - \exp_x^{-1}\mathcal{C}_f|| \leq ||\exp_x^{-1}\mathcal{C}_f|| \cdot (1 - \frac{\mathrm{id}}{\tan}(\tilde{r} \cdot ||\mathrm{ad}||)) \cdot e^{\tilde{r}\,||\mathrm{ad}||} .$$

Proof. In all cases we join the center $\mathcal{C}_f$ by a geodesic to the point x. For the first inequality in (i) we have from 6.4.6

$$|\operatorname{grad} P_f(x)| \cdot |\dot\gamma(1)| \geq \int_0^1 \frac{d^2}{dt^2}(P_f \circ \gamma)dt \geq |\dot\gamma|^2 \cdot 2\rho \frac{\mathbf{s}'_\Delta}{\mathbf{s}_\Delta}(2\rho) .$$

For the other inequalities we have

$$|| \frac{D}{dt}(v_f(\gamma(t)) - t\,\dot\gamma(t)) || \leq$$

$$\begin{cases} |\dot{\gamma}| \cdot 2(\Lambda\rho)^2 & \text{in case (i) from 6.4.6,} \\ ||\dot{\gamma}|| \cdot (1 - \frac{\rho}{\tan \rho}) & \text{in case (ii) from 7.4.5,} \\ ||\dot{\gamma}|| \cdot (1 - \frac{\mathrm{id}}{\tan} (\tilde{r}\, ||\mathrm{ad}||)) & \text{in case (iii) from 7.4.5;} \end{cases}$$

for (iii) observe that any geodesic which joins two points of $\exp D_r$ has by 7.8.2 (ii) a length $\leq r(1 + e^{r\,||\mathrm{ad}||}) = 2\tilde{r}$.

Finally, in cases (i) , (ii) use

$$|\, ||X(\gamma(1))|| - ||X(\gamma(o))||\, | \leq \int_o^1 ||\frac{D}{dt} X \circ \gamma||dt$$

for any differentiable vectorfield and any metric connection; in case (iii) this inequality holds for the components relative to parallel basis fields, so that 7.2.5 gives another factor $e^{\tilde{r}\,||\mathrm{ad}||}$.

8.2 Almost homomorphisms of compact groups

In this section we prove that almost homomorphisms can be improved to homomorphisms [13]; on this result depends the identification (3.6.4) of the deckgroup of the finite covering in Gromov's theorem with a subgroup of $O(n)$.

8.2.1 Proposition. (Almost homomorphisms are near homomorphisms)
Let H and G be compact Lie groups, H with volume normalized to 1 and G with biinvariant Finsler metric normalized to $||[X,Y]|| \leq ||X||\cdot||Y||$.
Let $w_o : H \to G$ be a (continuous) q-almost homomorphism, i.e. assume for all $h,k \in H$

$$d(w_o(k \cdot h) \cdot w_o(h)^{-1}, w_o(k)) \leq q \leq \frac{\pi}{6} .$$

Then there exists a (continuous) homomorphism

$$w : H \to G \text{ near } w_o \text{, i.e. for all } h \in H$$

$$d(w_o(h), w(h)) \leq q + \frac{1}{2} q^2 + q^4 .$$

Proof. Assume that a q_i-almost homomorphism $w_i : H \to G$ has already been defined. Then define for each $k \in H$ the map

$$\eta_k : H \to G \quad , \quad \eta_k(h) := w_i(k \cdot h) \cdot w_i(h)^{-1} \in B_{q_i}(w_i(k)) \ ,$$

and define

8.2.2 $$w_{i+1} : H \to G \quad \text{by} \quad w_{i+1}(k) := \mathcal{C}_{\eta_k} \ .$$

We plan to prove for all $h,k \in H$

8.2.3 $$d(w_{i+1}(k \cdot h) \cdot w_{i+1}(h)^{-1}, w_{i+1}(k)) \le \tfrac{1}{2} q_i^2 \, / \cos 1.25\, q_i =: q_{i+1} \ .$$

This shows that the sequence $\{w_i\}$ of almost homomorphisms converges rapidly to a limit, the homomorphism w .

8.2.4 Lemma. w_{i+1} $(i \ge 0)$ commutes with inversion (an isometry of G).

Proof. $$w_{i+1}(k)^{-1} = (\mathcal{C}_{\eta_k})^{-1} = \mathcal{C}_{(\eta_k)^{-1}} =$$

$$\mathcal{C}_{h \to w_i(h) \cdot w_i(k \cdot h)^{-1}} = \mathcal{C}_{h \to w_i(k^{-1}h) \cdot w_i(h)^{-1}}$$

(since left multiplication bei k^{-1} is volume preserving in H)

$$= w_{i+1}(k^{-1}) \ .$$

8.2.5 Lemma. For all $k,k',h \in H$ holds

$$\eta_{k'}(h) \cdot \eta_k(h)^{-1} \in B_{q_i}(w_i(k' \cdot k^{-1})) \ ,$$

$$\mathcal{C}_{\eta_{k'} \cdot \eta_k^{-1}} = w_{i+1}(k' \cdot k^{-1}) \ .$$

<u>Proof.</u> $\eta_{k'}(h) \cdot \eta_k(h)^{-1} = w_i(k'k^{-1} \cdot kh) \cdot w_i(kh)^{-1}$.

The fact that w_i is q_i-almost homomorphic gives the first statement; therefore the center of $\eta_{k'} \cdot \eta_k^{-1}$ exists in $B_{q_i}(w_i(k' \cdot k^{-1}))$. Now the second statement follows from

$$w_i(k'k^{-1} \cdot kh) \cdot w_i(kh)^{-1} = \eta_{k' \cdot k^{-1}}(k \cdot h)$$

(left multiplication by k is volume preserving in H).

Next we translate $\eta_{k'}$ and η_k so that their center goes to the identity $e \in G$ and shorten our notation:

$$\eta' := w_{i+1}(k')^{-1} \cdot \eta_{k'} \quad , \quad \eta := w_{i+1}(k^{-1}) \cdot \eta_k \quad .$$

These definitions imply (for all $h \in H$)

$$\eta'(h) \in B_{q_i}(e) \quad , \quad \eta(h) \in B_{q_i}(e) \ ,$$

$$\mathcal{C}_{\eta'} = e = \mathcal{C}_{\eta} = \mathcal{C}_{\eta^{-1}} \qquad (8.1.5,\ 8.2.4)$$

$$\mathcal{C}_{\eta' \cdot \eta^{-1}} = w_{i+1}(k')^{-1} \cdot \mathcal{C}(\eta_{k'} \cdot \eta_k^{-1}) \cdot w_{i+1}(k^{-1})^{-1} \qquad (8.1.5)$$

$$= w_{i+1}(k')^{-1} \cdot w_{i+1}(k' \cdot k^{-1}) \cdot w_{i+1}(k) \qquad (8.2.5)$$

This shows the

8.2.6 <u>Lemma.</u> w_{i+1} is q_{i+1}-almost homomorphic for

$$q_{i+1} = \max_{k',k \in H} d(e, \mathcal{C}_{\eta' \cdot \eta^{-1}}) \ .$$

8.2.7 <u>Lemma.</u> $d(e, \mathcal{C}_{\eta' \cdot \eta^{-1}}) \leq \left| \int_H \exp_e^{-1}(\eta'(h) \cdot \eta^{-1}(h)) dh \right| \cdot \dfrac{\tan 2q_i}{2q_i}$.

Proof. Since for all $h \in H$ $\eta'(h) \cdot \eta^{-1}(h) \in B_{2q_i}(e)$, the lemma follows directly from 8.1.10 (ii).

8.2.8 Lemma. $\left| \int_H \exp_e^{-1}(\eta'(h) \cdot \eta^{-1}(h))dh \right| \leq \frac{1}{2} q_i^2 \cdot \frac{q_i}{\sin q_i}$.

Proof. Define maps $X, Y : H \to T_e G$ by

$$X(h) := \exp_e^{-1} \eta'(h) \ , \ Y(h) := \exp_e^{-1}(\eta(h)^{-1}) \ .$$

Then $\int_H X(h)dh = 0 = \int_H Y(h)dh$ because of 8.1.4 and since e is the center of η' and of η^{-1} .

Next let $H(\ , \ )$ be the Campbell-Hausdorff map (7.2.7, 7.5), then

$$H(X(h), Y(h)) = \exp_e^{-1} (\eta'(h) \cdot \eta^{-1}(h)) \ .$$

Finally we have from 7.5.4 (iii) together with

$$\|H(X,Y)\| \leq \|X\| + \|Y\| \leq 2q_i$$

$$\|H(X(h),Y(h)) - (X(h)+Y(h))\| \leq \frac{1}{2} \frac{q_i}{\sin q_i} \cdot \|[X(h),Y(h)]\|$$

$$\leq \frac{1}{2} \frac{q_i}{\sin q_i} \|X(h)\| \cdot \|Y(h)\| \leq \frac{1}{2} \frac{q_i^3}{\sin q_i} \ ,$$

which proves 8.2.8.

This sequence of lemmas proves

$$q_{i+1} \leq \frac{1}{2} q_i^2 \frac{\tan 2q_i}{2 \sin q_i} \leq \frac{1}{2} q_i \ , \quad \text{if } q_i \leq \frac{\pi}{6} \ .$$

Since therefore $d(e, \mathcal{C}_{\eta' \cdot \eta^{-1}}) \leq \frac{1}{2} q_i$ we can in the proof of 8.2.7, which was based on 8.1.10, apply the estimate 7.4.5 to Jacobi fields along geodesics of length $2.5q_i$ instead of $4q_i$, which gives the improvement

$$q_{i+1} \leq \frac{1}{2} q_i^2 \frac{\tan 1.25q_i}{1.25 \sin q_i} \leq \frac{1}{2} q_i^2 / \cos 1.25q_i \leq \frac{1}{2} q_i^2 \, (1+q_i^2) \, (\text{if } q_i \leq \frac{\pi}{6})$$

and completes the inductive step 8.2.3.

One also checks that $\sum_{i=0}^{\infty} q_i \leq q + \frac{1}{2} q^2 + q^4$ (if $q \leq \frac{\pi}{6}$), which proves the estimate for $d(w_o(h), w(h))$ in 8.2.1.

8.3 Averages of differentiable maps

In this section we prove the result on which the maximal rank proof 5.4 for the map $F : \tilde{M} \to N$ depends.

8.3.1 Assumptions. Let M be a complete Riemannian manifold with injectivity radius $\geq r > 0$ and curvature bounds $|K| \leq \Lambda^2$. Let $\{m_i \in M ; i \in \mathbb{N}\}$ be a discrete set of points, $r/10$-dense in M (2.1.4). Let

$$F_i : B_r(m_i) \to N$$

be local differentiable maps into another manifold N (Riemannian or at least with affine connection D). Assume for each $x \in M$

$$F_i(x) \in B_x \subset N \quad , \quad \text{if} \quad d(x,m_i) \leq r ,$$

where B_x satisfies the assumptions of 8.1.3 or 8.1.4 so that a center can be defined for mass distributions in B_x.

Let $\psi : \mathbb{R}_+ \to [0,1]$ be a C^∞-cut off function such that

$$\psi\Big|_{[0,8]} = 1 \ , \ \psi\Big|_{[10,\infty)} = 0 \ , \ -1 \leq \psi' \leq 0 \ ,$$

and define weights $\phi_i(x)$ for the points $F_i(x) \in B_x$ by

$$\phi_i(x) := \psi\left(\frac{10\, d(x,m_i)}{r}\right) \cdot \left(\sum_j \psi\left(\frac{10\, d(x,m_j)}{r}\right)\right)^{-1} .$$

In this situation we can interpolate the local maps to a map $F : M \to N$ via the

Definition. $F(x) := \mathcal{C}\{F_i(x), \phi_i(x)\}$ (8.1.3 or 8.1.4).

The following equivariance result will be used in the case where a group Γ acts isometrically on M resp. connection preserving on N and where the local maps are compatible with these actions, i.e. $\gamma \cdot F_i(x) = F_{\gamma_i}(\gamma x)$ with γ_i defined by $\gamma \cdot m_i = m_{\gamma_i}$.

8.3.2 <u>Proposition</u>. (Equivariance of F)
Let $U : M \to M$ be an isometry which permutes the points m_i , $U(m_i) = m_{Ui}$; let $V : N \to N$ be an affine diffeomorphism; assume for the local maps

$$V \circ F_i(x) = F_{Ui}(Ux) \qquad \text{("local equivariance")}$$

then

$$V \circ F(x) = F(Ux) \qquad \text{("equivariance").}$$

<u>Proof</u>.

$$\begin{aligned} V \circ F(x) &= V(\mathcal{C}\{F_i(x),\phi_i(x)\}) && (8.3.1) \\ &= \mathcal{C}\{V \circ F_i(x),\phi_i(x)\} && (8.1.5) \\ &= \mathcal{C}\{F_{Ui}(Ux),\phi_{Ui}(Ux)\} && \text{(local equivariance)} \\ &= \mathcal{C}\{F_j(Ux),\phi_j(Ux)\} && (8.1.5) \\ &= F(Ux) && (8.3.1) \end{aligned}$$

Of course one wants 8.3.1 to define a differentiable map:

8.3.3 <u>Proposition</u>. (Differentiability of F)
The interpolation (8.3.1) $F : M \to N$ is differentiable. More precisely: Define in a neighborhood $\mathcal{N}$ of graph F the differentiable vectorfield (see 8.1.4) $v : \mathcal{N} \to TN$, $v(x,n) := -\sum_i \exp_n^{-1} F_i(x) \cdot \phi_i(x)$, then

(i) $v(x,F(x)) = 0$;

(ii) $\partial_1 v_{(x,F(x))} + D_2 v_{(x,F(x))} \cdot dF_x = 0$;

(iii) $D_2 v(x,F(x))$ is an invertible linear map;

where the above partial derivatives are defined by

(iv) $\partial_1 v_{(x,n)} : T_x M \to T_n N \ , \ \partial_1 v_{(x,n)} \cdot \dot{x}(0) := \frac{d}{dt} v(x(t),n)$

$(t \to v(x(t),n)$ is a curve in the vector space $T_n N$),

(v) $D_2 v_{(x,n)} : T_n N \to T_n N \ , \ D_2 v_{(x,n)} \cdot \dot{n}(0) := \frac{D}{dt} v(x,n(t))$

$(t \to v(x,n(t))$ is a vectorfield along the curve

$t \to n(t))$.

Proof. (i) is the definition of F , since for each x the center $\mathcal{C}\{F_i(x),\phi_i(x)\}$ is defined as the unique zero of the vectorfield $n \to v(x,n)$ on the set B_x (8.3.1). (iv) and (v) are the definitions such that (ii) follows by differentiation from (i) if the differentiability of F were already known. But (iii) - which is only a restatement of 8.1.3 (ii), resp. 8.1.6 (ii), resp. 8.1.8 (*) - is the crucial assumption in the implicit function theorem:

If one computes in local coordinates for TN the directional derivative

$$\frac{d}{dt} v(x,n(t)) = \partial_2 v_{(x,n)} \cdot \dot{n}(0) \in T_{v(x,n)} TN ,$$

then we have the so called horizontal and vertical components of this derivative:

$$\left(\frac{d}{dt} v(x,n(t))\right)^{hor} := \frac{d}{dt} (\pi_N \circ v(x,n(t))) = \frac{d}{dt} n(t)$$

$$\left(\frac{d}{dt} v(x,n(t))\right)^{ver} := \frac{D}{dt} v(x,n(t)) .$$

Now, if $v(x,n(0)) = 0$, then the horizontal components span the tangent space of the zero section in TN while the vertical components - if $\neq 0$ - are transversal to the zero section. Therefore (iii) says that the tangent space of the zero section and the image of $\partial_2 v_{(x_o,F(x_o))}$ span the tangent space $T_{v(x_o,F(x_o))} TN$ (for each x_o), so that by the implicit function theorem $v(x,G(x)) = 0$, $G(x_o) = F(x_o)$ locally defines a differentiable map G , but locally $G = F$ since $n \to v(x,n)$ has only the zero $n = F(x)$. This proves the differentiability of F .

8.3.4 Remark. (Maximal rank of dF)
We see from 8.3.3 (ii) and (iii):

dF_x has maximal rank if and only if the euclidean derivative $\partial_1 v_{(x,F(x))}$ (8.3.3 (iii)) has maximal rank.

In 5.4 we have enough information about the local maps F_i to prove from

$$\partial_1 v_{(x,n)} = - \sum_i d(\exp_n^{-1} \circ F_i)_x \cdot \phi_i(x) + \exp_n^{-1} F_i(x) \cdot (d\phi_i)_x$$

that $\partial_1 v$ has maximal rank.

REFERENCES

[1] Gromov, M.: Almost flat manifolds, J. Diff. Geom. 13 (1978), 231-242.

[2] Aleksandrow, A. D.: Metrische Räume mit einer Krümmung nicht größer als K . In: Der Begriff des Raumes in der Geometrie, Bericht von der Rieman Tagung 1957.

[3] Berger, M.: Les variêtés Riemanniennes $\frac{1}{4}$ -pincées, Ann. Scuola Norm. Sup. Pisa (III) 14 (1960), 161-170.

[4] Bieberbach, L.: Über die Bewegungsgruppen der Euklidischen Räume, I : Math. Ann. 70 (1911), 297-336, II: Math. Ann. 72 (1912), 400-412.

[5] Bourguignon, J. P., Karcher, H.: Curvature operators: Pinching estimates and geometric examples, Ann. scient. Éc. Norm. Sup., 4e série, t. 11 (1978), 71-92.

[6] Cheeger, J., Gromoll, D.: On the structure of complete manifolds of nonnegative curvature, Ann. Math. 96 (1972), 413-443.

[7] Cheeger, J., Ebin, D. G.: Comparison theorems in Riemannian geometry, North-Holland, Amsterdam 1975.

[8] Eberlein, P.: Lattices in spaces of non positive curvature, Ann. of Math. 111 (1980), 435-476.

[9] Gromoll, D., Meyer, W. T.: On complete open manifolds of positive curvature, Ann. of Math. 90 (1969), 75-90.

[10] Gromov, M.: Manifolds of negative curvature, J. Diff. Geom. 13 (1978), 223 - 230.

[11] Gromov, M.: Synthetic Riemannian geometry, preprint.

[12] Grove, K., Karcher, H.: How to conjugate C^1-close group actions, Math. Z. 132 (1973), 11-20.

[13] Grove, K., Karcher, H., Ruh, E. A.: Group actions and curvature, Inv. math. 23 (1974), 31-48.

[14] Hausner, M., Schwartz, J. T.: Lie groups, Lie algebras, Gordon and Breach, New York, 1968.

[15] Heintze, E., Im Hof, H. C.: Geometry of horospheres, J. Diff. Geom. 12 (1977), 481-491.

[16] Heintze, E., Karcher, H.: A general comparison theorem with applications to volume estimates for submanifolds, Ann. scient. Éc. Norm. Sup., 4e série, t. 11 (1978), 451-470.

[17] Im Hof, H. C., Ruh, E. A.: An equivariant pinching theorem, Comment. Math. Helv. 50 (1975), 389-401.

[18] Karcher, H.: Riemannian center of mass and mollifier smoothing, Comm. Pure and Appl. Math. 30 (1977), 509-541.

[19] Karcher, H.: Report on Gromov's almost flat manifolds, Sêm. Bourbaki, (1978/79) Exp. 526.

[20] Klingenberg, W.: Über Riemannsche Mannigfaltigkeiten mit positiver Krümmung, Comment. Math. Helv. 35 (1961), 47-54.

[21] Kneser, H.: Funktionentheorie, Vandenhoeck und Ruprecht, Göttingen 1958.

[22] Malcev, A. I.: On a class of homogeneous spaces, Amer. Math. Soc. Transl. No. 39 (1951), 276-307.

[23] Min Oo, Ruh, E. A.: Comparison theorems for compact symmetric spaces, Ann. scient. Éc. Norm. Sup., 4e série, t. 12 (1979), 335-353.

[24] Min Oo, Ruh, E. A.: Vanishing theorems and almost symmetric spaces of non-compact type, preprint Bonn 1980.

[25] Raghunathan, M. S.: Discrete subgroups of Lie groups, Ergebnisse Band 68, Springer Verlag Berlin etc. 1972.