RIEMANNIAN COMPARISON CONSTRUCTIONS

Hermann Karcher

When S. Kobayashi wrote his chapter for the first edition of this book he reported on amazing results which had just been obtained. Those results triggered a rapid development of comparison theory: in 1975 the book by Cheeger and Ebin [CE] dealing mostly with curvature > 0 appeared. In 1978 Gromov's theorem classifying almost flat manifolds appeared, see [**BK**]. In 1985 the case of curvature < 0 was treated in a book by Ballman, Gromov, and Schroeder [**BGS**]. In survey articles, by Burago and Zalgaller [**BZ**] and M. Berger [**B**], the field was explained to a wider audience. The article, "Comparison and Finiteness Theorems in Riemannian Geometry," by T. Sakai [S] contains a very complete bibliography.

I have been asked to present, "with complete proofs," part of this development. I selected material which the reader will, hopefully, see as a direct continuation of the research portrayed by Kobayashi. It therefore seemed natural to assume his chapter as background; I refer to its sections as (K.1), (K.2), etc. In particular I will use the exponential map, minimizing geodesics, conjugate and cut points, Jacobi fields, and the second variation formula without further motivation and almost without separate introduction. The only other prerequisite I am aware of is: not to be scared by the covariant derivative. I am using notation which resembles as closely as possible the use of Euclidean directional derivatives. The following three examples should explain the notation.

0.1 The fact that the Riemannian scalar product $g(\ ,\ )$ is *parallel* for the covariant derivative D but usually not constant for the local derivative ∂ of a chart is expressed by the following formula (U and V are vector fields):

$$\partial_X(g(U,V)) = \begin{cases} g(D_X U, V) + g(U, D_X V) \\ (\partial_X g)(U,V) + g(\partial_X U, V) + g(U, \partial_X V). \end{cases}$$

0.2 The so-called *symmetry* of the covariant derivative is expressed as

$$D_X Y - D_Y X = [X,Y] \quad (= \partial_X Y - \partial_Y X),$$

or with the help of a map $c\colon I \times I \to M$ as $(D/\partial s)(\partial/\partial t)c = (D/\partial t)(\partial/\partial s)c$. Both identities follow from the local expression of D in terms of the chart derivative and the symmetric Christoffel symbols:

$$D_X Y = \partial_X Y + \Gamma(X,Y), \qquad \Gamma(X,Y) = \Gamma(Y,X).$$

Usually this symmetry is (axiomatically) assumed; then one finds with (0.1):

$$2g(\Gamma(X,Y),Z) = -(\partial_Z g)(X,Y) + (\partial_X g)(Y,Z) + (\partial_Y g)(Z,X).$$

0.3 The gradient of a function f is the vector field defined by

$$\partial_X f = g(\operatorname{grad} f, X).$$

Further covariant differentiation gives the Hessian

$$D^2_{X,Y}f := \partial_X(\partial_Y f) - \partial_{D_X Y} f = g(D_X \operatorname{grad} f, Y)$$

$$= \partial^2_{X,Y} f - \partial_{\Gamma(X,Y)} f \qquad \text{(local chart expression)}.$$

The local formula shows the symmetry of $D^2 f$; $D \operatorname{grad} f$ is a symmetric endomorphism field. A function f with a positive definite Hessian is convex along any geodesic $c(D_{c'}c' = 0)$: $(f \circ c)'' = \partial_{c'}(g(\operatorname{grad} f, c')) = g(D_{c'}\operatorname{grad} f, c') > 0$.

CONTENTS

Guideline: observe how global properties are concluded from infinitesimal (= curvature) assumptions.

1. The setup to get curvature control started.

Natural functions, Jacobi equation, Riccati equation, constant curvature case, reduction of the n-dimensional case to a 1-dimensional discussion, the Riccati inequality, principal curvature—and Hessian bounds, generalized Rauch bounds, Bishop-Gromov volume bounds.

2. Immediate applications.

The Hadamard-Cartan theorem, fixed points of isometries, growth of the fundamental group, the Ricci-diameter bound with equality discussion.

3. Busemann functions.

$K \geqslant 0$ and compact totally convex exhaustion, Ricci $\geqslant 0$ and the splitting of a line as factor.

4. Triangle comparison theorems.

Angle and secant comparisons with upper or lower curvature bounds. A new proof of Toponogov's theorem.

5. Applications of the triangle theorems.

Bound for the number of generators of the fundamental group; critical points of the distance function, cut locus estimates (Klingenberg, Cheeger, Toponogov), sphere theorems (Rauch, Berger, Klingenberg, Shikata, Grove-Shiohama).

6. Complex projective space and its distance spheres.
Description from scratch: metric, embedding, equivariant isometries, curvature tensor of $\mathbf{C}P^n$, curvature tensor and short closed geodesics of the distance spheres.

1. THE SETUP TO GET CURVATURE CONTROL STARTED

1.1 *Natural maps and functions.* If one tries to generalize the arclength parametrization of curves to get good coordinates for a Riemannian manifold M, then a natural map *from* a Euclidean space $\mathbf{R}^n$ (e.g. T_pM) *into* M is the exponential map (K.1). It is defined via an initial value problem for geodesics from a point p:

$$\exp_p\colon T_pM \to M$$

$$\exp_p(v) = c(1), \quad \text{where } c \text{ is the geodesic with} \qquad (1.1.1)$$

$$c(0) = p, \qquad c'(p) = v.$$

Particularly natural functions from (parts of) a Riemannian manifold M into $\mathbf{R}$ are distance functions (from a point or a submanifold), i.e. functions which satisfy

$$|\operatorname{grad} f| = 1 \qquad \text{(distance function)}. \qquad (1.1.2)$$

Integral curves of distance functions are geodesics; namely, let u be an arbitrary tangent vector then

$$0 = \partial_u g(\operatorname{grad} f, \operatorname{grad} f) \underset{(0.1)}{=} 2g(D_u \operatorname{grad} f, \operatorname{grad} f)$$

$$\underset{(0.3)}{=} 2g(D_{\operatorname{grad} f} \operatorname{grad} f, u) \qquad \text{(symmetry of } D \operatorname{grad} f\text{)},$$

so indeed

$$D_{\operatorname{grad} f} \operatorname{grad} f = 0. \qquad (1.1.3)$$

The level surfaces are therefore called a family of "parallel" surfaces.

1.2 *Connections with Jacobi fields* (K.2). The differential equation for Jacobi fields J along a geodesic c,

$$J'' - R(J, c')c' = 0 \qquad (J'' = D_{c'}(D_{c'}J)), \qquad (1.2.1)$$

is the linearization of the geodesic equation along c. It is thus by definition that estimates on Jacobi fields as in Rauch's theorem (K.3.2) contain estimates of the differential of the exponential map or the Hessian of distance functions. We only have to relate the *initial conditions* for the Jacobi equation to the relevant families of geodesics.

First, for the exponential map we get

$$\partial \exp_p\big|_{s\cdot v}(sw) := \frac{\partial}{\partial t}\exp_p(s\cdot(v+tw))\Big|_{t=0} =: J(s), \quad (1.2.2)$$

$$J(0) = 0, \qquad J'(0) = w$$

since

$$J'(0) = \frac{D}{\partial s}\frac{\partial}{\partial t}\exp_p(s(v+tw))\Big|_{0,0}$$

$$\underset{(0.2)}{=} \frac{D}{\partial t}\left(\frac{\partial}{\partial s}\exp_p(s(v+tw))\Big|_0\right)\Big|_0$$

$$= \frac{D}{\partial t}(v+tw) = w.$$

Rauch type estimates $|J'(0)|\cdot h(s) \leqslant |J(s)| \leqslant |J'(0)|\cdot H(s)$ therefore translate into $|w|\cdot h(s) \leqslant |\partial \exp_p|_{sv}(sw)| \leqslant |w|\cdot H(s)$.

Second, for distance functions one has a natural unit normal field N along the level surfaces, $N := \operatorname{grad} f$, and the shape operator S with respect to that normal is the Hessian of f:

$$S\cdot u := D_u N = D_u \operatorname{grad} f \qquad (u \text{ tangential to a level surface}).$$

(1.2.3)

The eigenvalues of $S = D\operatorname{grad} f$ are called principal curvatures (of the level surface).

The sign convention in (1.2.3) is such that the principal curvatures of the sphere are positive for the outer normal. This is the better choice when dealing with level surfaces.

The family of normal geodesics defines natural diffeomorphisms from one level surface $f = \text{const.}$ to nearby ones:

$$E_s(p) := \exp_p s \cdot N(p), \qquad p \in f^{-1}(\text{const.}). \tag{1.2.4}$$

As in (1.2.2) one can describe the differentials of the E_s by Jacobi fields and the specifics of the construction determine the initial conditions: let $p(t)$ be a curve in a level surface then

$$\partial E_s\Big|_p \cdot \dot p(0) := \frac{\partial}{\partial t} E_s(p(t))\Big|_{t=0} =: J(s), \tag{1.2.5}$$

$$J(0) = \dot p(0), \qquad J'(0) = S \cdot \dot p(0)$$

since

$$J'(0) \underset{(0.2)}{=} \frac{D}{\partial t}\frac{\partial}{\partial s}\exp_p(s \cdot N(p(t)))\Big|_{s=0} = \frac{D}{\partial t} N(p(t)) \underset{(1.2.3)}{=} S \cdot \dot p(0).$$

Since $E_s(p(t))$ is a curve in another level surface $f^{-1}(s + \text{const.})$, (1.2.5) in fact describes all the shape operators S_s of the level surfaces (along each normal geodesic):

$$S_s \cdot J(s) = J'(s). \tag{1.2.6}$$

Following definition (1.2.3), interpretations of the shape operator usually emphasize the turning speed of the normal along the level surface. But the shape operator also controls the change of length of the geodesic projections E_s between level surfaces:

$$\begin{aligned} &\frac{d}{ds} g\left(\frac{\partial}{\partial t} E_s(p(t)), \ \frac{\partial}{\partial t} E_s(p(t))\right) \\ &\quad \underset{(1.2.5)}{=} 2g\left(\frac{\partial}{\partial t} E_s(p(t)), S_s \cdot \frac{\partial}{\partial t} E_s(p(t))\right). \end{aligned} \tag{1.2.7}$$

1.3 *The Riccati equation for the shape operators of the level surfaces of a distance function.* The claim of this heading, namely that the shape operators of the level surfaces of a distance function satisfy the *first*-order equation (1.3.1) along the normal geodesics is a key observation. Together with (1.2.7) it allows us to split Rauch's estimates for a second-order equation into two first-order steps—each with a geometric interpretation. Because of its importance I give three derivatives of (1.3.1) which connect three different points of view of the basic comparison construction.

First, differentiate (1.2.6) and use the Jacobi equation:

$$J''(s) = \frac{D}{\partial s}\left(S_s \cdot J(s)\right) = S_s' \cdot J(s) + S_s \cdot J'(s)$$

$$= -R(J, N)N(s) = \left(S_s' + S_s^2\right) \cdot J(s).$$

We abbreviate $R(J, N)N =: R_N \cdot J$; also, for fixed s we can consider $J(s)$ as an arbitrary tangent vector to a level surface. Then the last equation is a *Riccati equation* for S_s:

$$S_s' = -R_N - S_s^2. \tag{1.3.1}$$

Second, insert a parallel vector field $U(s) \perp N$ into (1.2.3) and differentiate in the direction $\partial/\partial s = N = \operatorname{grad} f$:

$$\frac{D}{\partial s}\left(S_s \cdot U(s)\right) = S_s' \cdot U(s)$$

$$= \frac{D}{\partial s}\left(D_{U(s)} \operatorname{grad} f\right) = D^2_{N,U} \operatorname{grad} f$$

$$\left(\text{note } \frac{D}{\partial s} U = 0\right).$$

Insert in this equality the definition of the curvature tensor $D^2_{N,U} \operatorname{grad} f = D^2_{U,N} \operatorname{grad} f - R(U, N) \operatorname{grad} f$ and the definition of the second differential of a vector field $D^2_{U,N} \operatorname{grad} f \underset{(\text{Def.})}{=} D_U(D_N \operatorname{grad} f) - D_{D_U N} \operatorname{grad} f \underset{(1.1.3,\, 1.2.3)}{=} -S^2 \cdot U$ to obtain (1.3.1) again.

Third, consider a 2-dimensional Riemannian metric in Gaussian form:

$$ds^2 = du^2 + G^2(u, v)\, dv^2, \qquad \text{with curvature } K = -\frac{G_{uu}}{G}.$$

The u-lines are geodesics and their orthogonal trajectories are parallel curves. The rate of change of their lengths gives their geodesic curvature:

$$\frac{d}{du}\int G\, dv = \int \frac{G_u}{G} G\, dv = \int \kappa_g\, ds \quad \text{or} \quad \kappa_g = \frac{G_u}{G}.$$

One further differentiation gives the Riccati equation,

$$\frac{d}{du}\kappa_g = -K - \kappa_g^2. \tag{1.3.2}$$

The 2-dimensional formula has been known a long time; we also shall see that the n-dimensional case is close to (1.3.2).

1.3.3. It turns out to be rather easy to reduce the *comparison* discussion of (1.3.1) both for upper or lower curvature bounds to 1-dimensional Riccati inequalities—as if we had upper or lower curvature bounds in (1.3.2) (see (1.5) below). So the original Rauch line is replaced by:

Step 1. Prove inequalities for the principal curvatures of level surfaces of distance functions via 1-dimensional Riccati inequalities.

Step 2. Use Step 1 to integrate (1.2.6) or (1.2.7).

1.4 *The constant curvature case.* The Rauch comparison theorem (K.3.2) is formulated in such a way that upper or lower bounds for the curvature seem to play a completely symmetric role. Most applications so far have gone via constant curvature models. Integration of the Rauch estimates to distance or volume control always requires in the case of *upper* curvature bounds some size restriction, e.g. stay away from some cut locus. That such restrictions are *not* needed in comparisons with *smaller* curvature models

(Toponogov, Bishop-Gromov) made these results prime tools in the development of comparison theory.

I summarize the relevant constant curvature data.

1.4.1. We denote the hyperbolic space of curvature $\kappa < 0$, Euclidean space ($\kappa = 0$), or the sphere of curvature $\kappa > 0$ either jointly by M_κ, or, if the sign of κ is specified, by H^n_κ, $\mathbf{R}^n$, S^n_κ.

1.4.2. Similarly, we use a common notation for the functions which control the trigonometry of those spaces. Denote the solutions of the differential equation

$$f'' + \kappa \cdot f = 0$$

which have the same initial conditions as sin, respectively, cos by

$$s_\kappa \text{ resp. } c_\kappa; \qquad s'_\kappa = c_\kappa, \qquad c'_\kappa = -\kappa s_\kappa.$$

1.4.3. Distance spheres of radius r have

$$\text{principal curvatures} = \frac{s'_\kappa}{s_\kappa}(r) =: ct_\kappa(r); \quad ct'_\kappa = -\kappa - ct^2_\kappa;$$

$$\text{the length of their great circles} = 2\pi s_\kappa(r).$$

Parallel surfaces at distance r from totally geodesic hyperplanes have

$$\text{principal curvatures} = \frac{c'_\kappa}{c_\kappa}(r).$$

The Hessian of the distance function has in the radial direction the eigenvalue 0. For the proof of Toponogov's theorem I need to rescale the distance function so that all eigenvalues of the Hessian become equal. This "modified distance" function is:

$$md_\kappa(r) := \int_0^r s_\kappa = \begin{cases} \frac{1}{2}r^2 & \text{if } \kappa = 0 \\ 1 - \cos r & \text{if } \kappa = 1 \\ \cosh r - 1 & \text{if } \kappa = -1 \end{cases}.$$

This function also avoids case distinctions in the *cosine formula* for triangles (of edge lengths a, b, c, in M_κ):

$$md_\kappa(c) = md_\kappa(a-b) + s_\kappa(a)\cdot s_\kappa(b)\cdot(1-\cos\gamma).$$

1.4.4. If $\kappa \leqslant \Delta$ then we have in appropriate intervals

$$s_\Delta(r) \leqslant s_\kappa(r), \qquad c_\Delta(r) \leqslant c_\kappa(r), \qquad ct_\Delta(r) \leqslant ct_\kappa(r),$$

which expresses comparisons of lengths, areas and principal curvatures by explicit formulas (compare 1.4.3).

1.4.5. Rauch's estimates as well as the ones we shall prove are formulated for Jacobi fields perpendicular (or "normal") to their geodesics. This is all one needs since the tangential part of a Jacobi field is always a *linear* field and hence explicitly known from the initial data (independently from the curvature tensor):

$$J_{\tan} := g(J, c')\cdot c' \qquad (\text{normalization } |c'| = 1).$$

Indeed, the skew symmetries of R imply $J''_{\tan} = 0$ and $R(J_{\tan}, c')c' = 0$. This and the Pythagorean theorem extend estimates for normal Jacobi fields to arbitrary ones.

1.5 *Reduction of the discussion of (*1.3.1*) to a* 1*-dimensional inequality.* All estimates in this section are pointwise. One can therefore allow that the lower and upper curvature bounds vary from point to point. This generalization will not be pursued.

1.5.1. Assume a lower bound $\delta \leqslant K$.
To discuss $S' = -R_N - S^2$ let u be a parallel unit field along a geodesic normal to the level surfaces. Then we obtain a first-order Riccati inequality as follows:

$$g(Su,u)' = -g(R(u,N)N,u) - g(S^2u,u),$$

$$g(Su,u)' \leqslant -K(u\wedge N) - g(Su,u)^2 \leqslant -\delta - g(Su,u)^2.$$

1.5.2. Assume an upper curvature bound $K \leqslant \Delta$.
To discuss $S' = -R_N - S^2$ let $p(t)$ be a curve in one level surface of the distance function. Consider the 2-dimensional "ruled" surface

$$F(s,t) := \exp_{p(t)} s \cdot N(p(t)).$$

Since the s-lines are geodesics we have a Gaussian parametrization and the geodesic curvatures of the parallel t-curves satisfy (by 1.3.2)

$$\kappa_g' = -K^F - \kappa_g^2,$$

$$\kappa_g' \geqslant -\Delta - \kappa_g^2 \qquad \text{from } (1.5.3).$$

Bounds derived from $K \leqslant \Delta$ therefore have a 2-dimensional geometric interpretation!

1.5.3. For the ruled surface of (1.5.2),

$$K^F \leqslant K^M\left(\frac{\partial}{\partial s}F \wedge \frac{\partial}{\partial t}F\right).$$

Proof. The s-lines are geodesics in M and hence in F. The t-derivative of this family, therefore, gives Jacobi fields $J(s)$ for both spaces. The covariant derivative D^F in F is the *orthogonal* projection of the covariant derivative D^M in M. The computation

$$\frac{\partial^2}{(\partial s)^2} g(J(s), J(s))$$

$$= \begin{cases} 2g\left(\dfrac{D^M}{ds}J, \dfrac{D^M}{ds}J\right) - 2g(R^M(J,F')F',J) \\ 2g\left(\dfrac{D^F}{ds}J, \dfrac{D^F}{ds}J\right) - 2g(R^F(J,F')F',J) \end{cases}$$

therefore proves

$$g(R^F(J,F')F',J) \leqslant g(R^M(J,F')F',J).$$

Volume estimates also fit in this discussion. For any family of invertible linear maps, $L(s)$, between Euclidean spaces we have (Flanders 7.10)

$$\frac{d}{ds}\det(L(s)) = \operatorname{trace}(L' \cdot L^{-1}) \cdot \det(L(s)).$$

We apply this to the differentials ∂E_s of (1.2.4) and with the abbreviation

$$a(s) := \det(\partial E_s)$$

we get with (1.2.5, 1.2.6)

$$\frac{d}{ds} a(s) = (\operatorname{trace} S_s) \cdot a(s). \tag{1.5.4}$$

This says that trace S_s is the growth rate of the hypersurface volume $a(s)$—tangent space-wise.

On the other hand, taking the trace of the Riccati equation (1.3.1) for S gives

$$\frac{d}{ds} \operatorname{trace} S = -\operatorname{trace} R_N - \operatorname{trace} S^2. \tag{1.5.5}$$

1.5.6. Recall: $H := \frac{1}{n-1} \cdot \operatorname{trace} S$ is the mean curvature of the level hypersurface.

ricci$(N, N) :=$ trace$(Y \to R(Y, N)N)$ is the Ricci curvature of M^n in the direction N.

Note that (1.5.4) and (1.5.5) are again controls in two first-order steps: (1.5.4) controls the volume growth of the level surfaces in terms of their mean curvature, and (1.5.5) controls the change of the mean curvature in terms of the Ricci curvature, except that trace S^2 and trace S are only related through

Schwarz' inequality for endomorphisms

$$(n-1) \cdot \operatorname{trace} S^2 \geqslant (\operatorname{trace} S)^2, \qquad \text{“ = ” iff } S = \lambda \cdot \mathrm{id}. \tag{1.5.7}$$

This and (1.5.5) give:

1.5.8. Assume a lower bound $(n-1)\cdot\rho\leqslant \mathrm{ric}(N,N)$. Then we have, for the mean curvature $H=1/(n-1)\mathrm{trace}\,S$ of the level surfaces of a distance function (1.1.2) along the geodesics normal to the level surfaces, the Riccati inequality

$$H'\leqslant -\frac{1}{n-1}\mathrm{ric}(N,N)-H^2\leqslant -\rho-H^2.$$

Now (1.5.4) and (1.5.8) are a perfect control. Because of the use of (1.5.7) one does not have a corresponding result assuming upper bounds on the Ricci curvature—in fact almost no consequences of such upper bounds are known.

For equality discussions see (2.4.2, 3.6).

1.6. *The Riccati comparison argument.* Note the simplicity of the following arguments. The assumptions for (1.6.1) came from (1.5.1, 1.5.2, 1.5.8).

Consider two functions f, F which satisfy on some interval

$$f'\leqslant -\rho-f^2,\qquad F'\geqslant -\rho-F^2. \tag{1.6.1}$$

Then

$$\left((f-F)\cdot e^{\int(f+F)}\right)'\leqslant 0.$$

1.6.2. COROLLARIES. *If* a) $f(r_0)\geqslant F(r_0)$, *respectively,* b) $f(r_0)\leqslant F(r_0)$ *then*

a) $f(r)\geqslant F(r)$ *for* $r\leqslant r_0$ *(as long as* $f<\infty$, $F>-\infty$*),*

b) $f(r)\leqslant F(r)$ *for* $r\geqslant r_0$ *(as long as* $f>-\infty$, $F<\infty$*).*

1.6.3. *Assume in addition to* (1.6.1)

a) $\lim_{r\to 0} f(r)=+\infty$, *respectively*

b) $\lim_{r\to 0} F(r)=+\infty$, *then*

a) $f(r)\leqslant ct_\rho(r)$ *as long as* $f=\infty$,

b) $F(r)\geqslant ct_\rho(r)$ *as long as* $ct_\rho=\infty$. *(Definition of* ct_ρ *in* (1.4.3).*)*

Proof. a) Let f be defined and finite on $(0,R)$ and assume for some $r_0\in(0,R)$ that $f(r_0)>ct_\rho(r_0)$, i.e. $f(r_0)\geqslant ct_\rho(r_0-\epsilon)$ for some $\epsilon > 0$. Because of (1.6.2) a) we have on $(0, r_0)$ $f(r) \geqslant ct_\rho(r-\epsilon)$, a contradiction since $\lim_{r \downarrow \epsilon} ct_\rho(r-\epsilon) = +\infty$. The proof of b) is essentially the same.

REMARK. We will apply these estimates to (1.5.1, 1.5.2, 1.5.8) to get estimates for solutions of the Riccati equation (1.3.1). There are other ways to deal with (1.3.1). I hope the reader finds it helpful for an intuitive understanding that the estimates are explained essentially in a 2-dimensional picture.

1.7–1.9 *Basic geometric comparison results.* Depending on the choice of the distance function, we obtain from the just proved estimates various geometric comparison statements. In particular, distance functions from a point, from a closed geodesic (5.3.1), and from a hypersurface have been used. For the purpose of this exposition it will be sufficient to treat here only the *distance function from a point*. Executing step 1 of (1.3.3) I first derive principal curvature estimates; in step 2 these are integrated to length, respectively, volume comparison results. All the explicit bounds are *sharp* for the constant curvature models. Conditions involving $\Delta^{-1/2}$ etc. are to be ignored if $\Delta \leqslant 0$.

1.7. *Bounds for the principal curvatures of distance spheres.* Let $c(r)$, $0 \leqslant r \leqslant R$, be a geodesic arc which does not meet a conjugate point. Within a sufficiently narrow neighborhood U the arc c is length minimizing and we can define on U a "local" distance function f from $p = c(0)$. Sections 1.1 to 1.6 apply to such local distance functions. Estimates for the principal curvatures $\kappa_i(r)$ of the level surfaces or for the *Hessian* $D\,\mathrm{grad}\, f$ are reformulations of (1.6.3). I find the bounds for the Hessian more useful when they are rewritten for the modified distance function $md_\kappa \circ f$ (1.4.3). This can be done using

$$\mathrm{grad}(h \circ f) = (h' \circ f) \cdot \mathrm{grad}\, f$$

$$D\,\mathrm{grad}(h \circ f) = (h' \circ f) \cdot D\,\mathrm{grad}\, f + (h'' \circ f) \cdot df \otimes \mathrm{grad}\, f.$$

1.7.1. If $\delta \leqslant K$, then (1.5.1) and (1.6.3) imply up to the first conjugate point

$$\kappa_i(r) \leqslant ct_\delta(r),$$

or equivalently

$$D\,\mathrm{grad}(md_\delta \circ f) \leqslant (c_\delta \circ f) \cdot id.$$

If $\delta > 0$ then conjugate points are not farther than $\pi\delta^{-1/2}$ away.

1.7.2. If $(n-1) \cdot \rho \leqslant \mathrm{ric}$, then (1.5.8) and (1.6.3) imply up to the first conjugate point the mean curvature estimate

$$h(r) := \frac{1}{n-1} \sum \kappa_i(r) \leqslant ct_\rho(r) =: h_\rho(r), \qquad \text{or}$$

$$-\Delta f := \mathrm{trace}(D\,\mathrm{grad}\, f) \leqslant (n-1) ct_\rho(r).$$

If $\rho > 0$ then conjugate points are not farther than $\pi\rho^{-1/2}$ away.

1.7.3. If $K \leqslant \Delta$, then (1.5.2) and (1.6.3) imply for $0 < r < \pi \cdot \Delta^{-1/2}$

$$\kappa_i(r) \geqslant ct_\Delta(r),$$

or equivalently

$$D\,\mathrm{grad}(md_\Delta \circ f) \geqslant (c_\Delta \circ f) \cdot id.$$

There are no conjugate points in $(0, \pi \cdot \Delta^{-1/2})$.

REMARK. I repeat that the estimates are sharp in the constant curvature models. Also note, that (0.3) and (1.7.3) give convexity statements for the distance functions from a point, if $r < \pi/2 \cdot \Delta^{-1/2}$.

1.8. *Generalized Rauch estimates.* We execute step 2 of (1.3.3), namely integrate (1.2.6) using the principal curvature estimates of (1.7). If a Jacobi field J is $\neq 0$ on some interval, then $|J|' =$

$\langle J, J' \rangle / |J|$, hence

$$\left(\frac{|J|}{s_\kappa} \right)' = \frac{|J|}{s_\kappa} \cdot \left(\frac{\langle J, J' \rangle}{\langle J, J \rangle} - \frac{s_\kappa'}{s_\kappa} \right).$$

1.8.1. Assume $\delta \leqslant K$, $J(0) = 0$ and let r_{conj} be the distance to the first conjugate point along c. Then (1.7.1) and (1.2.6) imply

$$\frac{|J(r)|}{s_\delta(r)} \quad \text{is nonincreasing in} \quad 0 < r < r_{\mathrm{conj}},$$

in particular $|J(r)| \leqslant |J'(0)| \cdot s_\delta(r)$ (Rauch).

1.8.2. Assume $K \leqslant \Delta$, $J(0) = 0$ and $r < \pi \cdot \Delta^{-1/2}$. Then (1.7.3) and (1.2.6) imply

$$\frac{|J(r)|}{s_\Delta(r)} \quad \text{is nondecreasing in} \quad 0 < r < \pi \cdot \Delta^{-1/2},$$

in particular $|J(r) \geqslant |J'(0)| \cdot s_\Delta(r)$ (Rauch).

Remarks. (i) We have already seen in (1.2.2) how these estimates control the change of arclength under the exponential map (still sharp in the constant curvature models). This arclength control will be improved to distance control in Section 4.

(ii) In a situation where the curvatures approach 0 as the distance r from some distinguished point $*$ grows, one wants the curvature bounds $\delta(r)$, $\Delta(r)$ to depend on r. The generalizations of (1.8.1, 1.8.2) are easy. The tricky part (which rarely works) is the improvement to distance control.

1.9. *Bounds for the volume of distance spheres and balls.* We just saw that Rauch's original comparison results were later improved to monotonicity statements. The same is true for volumes: first one had Bishop's comparison results, later Gromov pointed out that the corresponding global monotonicity statements are true even beyond conjugate points; they are more powerful and easier to use (see 2.4.2, 3.6 for applications).

We saw in (1.5.4) that the trace of the shape operator is the growth rate of the tangent space-wise hypersurface volume $a(s)$. This growth rate is controlled in (1.7.2, 1.7.3). One further radial integration,

$$v(r) := \int_0^r a(s)\, ds, \tag{1.9.1}$$

gives the ray-wise contribution of the geodesic in question to the volume of the ball. Finally, an integration over all directions at the center of the ball gives the volumes of spheres and balls. I first state the ray-wise inequalities.

1.9.2. *Assume $(n-1)\rho \leqslant \mathrm{ric}$. Then (1.5.4) and (1.7.2) imply up to the first conjugate point*

$$\frac{a}{a_\rho}(r) \quad \textit{is nonincreasing},$$

where $a_\rho(r) = s_\rho(r)^{n-1}$ is the integral of (1.5.4) in the constant curvature model M_ρ. Applying this to (1.9.1) gives

$$\frac{v}{v_\rho}(r) \quad \textit{is nonincreasing}.$$

Proof.

$$\left(\frac{a}{a_\rho}\right)'(r) = \frac{a}{a_\rho}\cdot(h - h_\rho) \leqslant 0.$$

$$\frac{v}{v_\rho}(r) = \frac{\int_0^r \frac{a}{a_\rho}(s)\cdot a_\rho(s)\, ds}{\int_0^r 1\cdot a_\rho(s)\, ds}$$

is monotone since $(a/a_\rho)(s)$ is monotone.

1.9.3. *Assume $K \leqslant \Delta$ and $r < \pi\cdot\Delta^{1/2}$. Then (1.5.4), (1.7.3) and (1.9.1) imply (with the same argument as in 1.9.2)*

$$\frac{a}{a_\Delta}(r) \quad \textit{and} \quad \frac{V}{V_\Delta} \quad \textit{are nondecreasing}.$$

At the final step (the integration over all directions at the center of the ball) the cut locus interferes: Beyond the cut point (of the center) a geodesic ray does not contribute to the volume of distance spheres and balls. This cannot be rescued without further assumptions in the case of lower volume bounds; on the other hand, this actually helps in integrating (1.9.2): If a geodesic hits a cut point in M we can set $a(r) = 0$ for larger r and can integrate up to the cut point distance in M_ρ even if (1.9.2) is not true that far! Therefore we have Gromov's global monotonicity result for distance balls B_r.

1.9.4. *Assume* $(n-1)\cdot\rho \leqslant \text{ric}$. *Then*

$$\frac{\text{vol}_{n-1}(\partial B_r \subset M)}{\text{vol}_{n-1}(\partial B_r \subset M_\rho)}$$

and

$$\frac{\text{vol}_n(B_r \subset M)}{\text{vol}_n(B_r \subset M_\rho)}$$

are nonincreasing functions of r with $\lim_{r\to 0}$ *vol-ratio(r)* $= 1$.

1.9.5. *Assume* $K \leqslant \Delta$ *and that* B_r *does not meet a cut point of its center. Then*

$$\frac{\text{vol}_{n-1}(\partial B_r \subset M)}{\text{vol}_{n-1}(\partial B_r \subset M_\Delta)}$$

and

$$\frac{\text{vol}_n(B_r \subset M)}{\text{vol}_n(B_r \subset M_\Delta)}$$

are nondecreasing functions of r with $\lim_{r\to 0}$ *vol-ratio(r)* $= 1$.

2. IMMEDIATE APPLICATIONS OF THE CURVATURE CONTROLLED BOUNDS

2.1. NONPOSITIVELY CURVED MANIFOLDS (Hadamard, Cartan). *Let* M^n *be complete and assume curvature bounds* $K \leqslant \Delta \leqslant 0$.

2.1.1. *If M is simply connected, then (for each p):* $\exp_p$*:* $T_pM \to M$ *is an expanding* ($\geqslant 1$) *diffeomorphism.*

2.1.2. *If M is not simply connected, then in each homotopy class of paths from p to q there is exactly one geodesic.*

2.1.3. *If M is simply connected and we take* $f(x) := d(p,x)$ *then* $D\,\mathrm{grad}(\frac{1}{2}f^2) \geqslant id$*; thus* f^2 *is a strictly convex function and all distance balls are strictly convex.*

Proof. From (1.2.2, 1.8.2) and $(1/r)s_\Delta(r) \geqslant 1$ we have $|\partial \exp_p \cdot w| \geqslant |w|$ everywhere. Therefore $\exp_p$ is an expanding local diffeomorphism. If $\exp_p$ were not injective we would have two geodesic arcs γ_1, γ_2 from p to some $q \in M$. By simple connectivity γ_1 and γ_2 are homotopic. The homotopy can be lifted to T_pM since $\exp_p$ is an expanding local diffeomorphism; note that local inverse images of Cauchy sequences in M are Cauchy in T_pM. But the lifts of different geodesics give different radial segments and can therefore not have a common endpoint. The contradiction proves injectivity, hence (2.1.1).

It is useful to consider T_pM not only as an Euclidean space, but also with the Riemannian metric $\tilde{g}$ pulled back by the (local) diffeomorphism $\exp_p$. This makes $\exp_p$: $(T_pM, \tilde{g}) \to M$ a locally isometric map and suggests taking $(T_pM, \tilde{g})$ as a metric realization of the universal covering $\tilde{M}$. Different geodesic arcs in the same homotopy class would then give different geodesic connections in $(T_pM, \tilde{g})$ for which (2.1.1) holds, proving (2.1.2).

(2.1.3) combines (0.3, 1.7.3) in the case $\Delta = 0$, i.e. $md_\Delta \circ f = (1/2)f^2$ (which is differentiable because of (2.1.1)).

2.2 Fixed Points of Isometries (Cartan 1928). *Let M be simply connected and assume* $K \leqslant 0$. *Then every bounded set is contained in a unique smallest convex ball. In particular, every isometry group of M which has a bounded orbit has a fixed point.*

Proof (Eberlein). The intersection of two closed balls B_1, B_2 of radius r and midpoints m_1, m_2 is contained in a smaller ball. Let $p \in B_1 \cap B_2$ have maximal distance from the midpoint m between m_1, m_2; the strict convexity (2.1.3) of $x \to d(p,x)^2$ together with $d(p, m_1) \leqslant r$, $d(p, m_2) \leqslant r$ implies $d(p,m) < r$. A ball of smallest radius which contains a given bounded set is therefore unique. If the bounded set is the orbit of an isometry group, then all isometries of the group map a smallest containing ball to a smallest containing ball; the center of the unique such ball is a fixed point. Note that the arguments are, verbatim, the same as for the Euclidean case.

2.3. *Growth of the fundamental group* (Svarc 1955, Milnor 1968).

2.3.1 Let $\{\gamma_1, \ldots, \gamma_N\}$ be a finite set of generators for a group π. Define the growth function

growth(k) := Number of elements in π which can be written as a product of at most k factors in the generators.

This is justified since if the growth function for one set of generators grows exponentially (respectively, polynomially of degree d) then it does so for all other finite sets of generators (not difficult).

2.3.2. *Let M be compact and assume $K \leqslant \Delta < 0$. Then the fundamental group $\pi_1(M,p)$ grows exponentially.*

2.3.3. *Let M^n be complete and assume $0 \leqslant \mathrm{ric}$. Then every subgroup $G \subset \pi_1(M,p)$ which has a finite set Γ of generators has at most polynomial growth of degree n.*

REMARK. For bounds on the number of generators see (5.1).

Proof. The fundamental group can be considered as a group π_1 which acts isometrically on the universal cover $\tilde{M}$. This is easy for (2.3.2) where we have a nice metric model for $\tilde{M}$, and from a sufficiently abstract point of view (2.3.3) is the same. π_1 is called the group of deck transformations.

For (2.3.2) we have to define a suitable set of generators. Pick some $\tilde{p} \in \tilde{M}$ and define the Dirichlet fundamental domain

$$F := \{ q \in \tilde{M};\ d(q, \tilde{p}) \leqslant d(q, \gamma\tilde{p}) \text{ for all } id \neq \gamma \in \pi_1 \}.$$

Let D be the diameter of the compact manifold M; clearly F is contained in the ball of radius D around $\tilde{p}$. Now define a finite subset $\Gamma \subset \pi_1$ (which will be shown to generate π_1)

$$\Gamma := \{\gamma \in \pi_1;\ d(F, \gamma F) \leqslant 1\},$$

with $d(A, B) := \min\{d(p,q);\ p \in A,\ q \in B\}$. The open sets $\gamma \cdot \mathring{F}$ are disjoint and by the triangle inequality $\Gamma \cdot F \subset B_{3D+1}(\tilde{p})$. Therefore Γ is finite, namely

$$|\Gamma| \leqslant \operatorname{vol}(B_{3D+1}(\tilde{p})) \cdot \operatorname{vol}(F)^{-1}.$$

Also $\Gamma \cdot F \supset \{q \in \tilde{M};\ d(F,q) \leqslant 1\}$, hence $\Gamma^k \cdot F \supset B_k(\tilde{p})$; therefore $\bigcup_k(\Gamma^k \cdot F) = \tilde{M}$ and Γ generates. Moreover, growth$(k) = |\Gamma^k| \geqslant \operatorname{vol}(B_k(\tilde{p})) \cdot \operatorname{vol}(F)^{-1}$. Finally, (1.9.5) says that $\operatorname{vol}(B_k,(\tilde{p}))$ is at least as large as the volume of a ball of radius k in the hyperbolic space H^n_Δ. Hence

$$\operatorname{vol}(B_k(\tilde{p})) \geqslant \operatorname{vol}(B_k^\Delta) = \operatorname{vol}(S^{n-1}) \cdot \int_0^k s_\Delta^{n-1}(r)\,dr,$$

which grows exponentially, proving (2.3.2).

The proof of (2.3.3) needs only minor modifications. Since the Dirichlet fundamental domain for G may not have finite diameter put $F_1 = \mathring{F} \cap B_1(\tilde{p})$.

From the generating set Γ we need its maximal displacement $L := \max\{d(\tilde{p}, \gamma\tilde{p});\ \gamma \in \Gamma\}$. The sets $\gamma \cdot F_1$ are disjoint (for $\gamma \in G$) and $\Gamma^k \cdot F_1 \subset B_{k \cdot L+1}(\tilde{p})$ (triangle inequality). Finally we get from the upper volume bound (1.9.4)

$$|\Gamma^k| \cdot \operatorname{vol}(F_1) \leqslant \operatorname{vol}(B_{k\cdot L+1}(\tilde{p})) \leqslant \operatorname{vol}(B_{k\cdot L+1} \subset \mathbf{R}^n) \leqslant \text{const} \cdot k^n.$$

2.4 Ricci Diameter Bound. *Let M^n be complete and assume a positive lower bound $0 < (n-1)\rho \leqslant$ ric. Then:*

2.4.1. (Myers 1935) $\operatorname{diam}(M) \leqslant \pi \cdot \rho^{-1/2}$ *and M is compact.*

2.4.2. (Cheng 1975) *If* $\operatorname{diam}(M^n) = \pi \cdot \rho^{-1/2}$ *then M^n is isometric to S^n_ρ.*

Proof. (2.4.1) follows from (1.7.2) and (K.4.1) (no geodesic minimizes beyond its first conjugate point.)

Proof of (2.4.2) (Itokawa, Shiohama 1983). The argument shows the power of Gromov's global monotonicity extension (1.9.4) of the volume bound. We may normalize to $\rho = 1$ (to have $\pi \cdot \rho^{-1/2} = \pi$). From (1.9.4) we have: The volume ratio $f(r) := \text{vol}(B_r) \cdot \text{vol}(B_r^\rho)^{-1}$ is nonincreasing, in particular $1 = f(0) \geqslant f(r) \geqslant f(\pi)$ for $0 \leqslant r \leqslant \pi$. In the sphere S_ρ^n we have for $\mu(r) := \text{vol}(B_r^\rho) \cdot \text{vol}(S_\rho^n)^{-1}$ obviously $\mu(r) + \mu(\pi - r) = 1$. Let $p, q \in M$ be such that $d(p,q) = \text{diam}(M) = \pi$. The open balls $B_r(p), B_{\pi-r}(q)$ are then *disjoint*, hence

$$\begin{aligned}\text{vol}(M) &\geqslant \text{vol}(B_r(p)) + \text{vol}(B_{\pi-r}(q)) \\ &= (f(r)\cdot\mu(r) + f(\pi - r)\cdot\mu(\pi - r))\cdot\text{vol}(S_\rho^n) \\ &\geqslant f(\pi)\cdot\text{vol}(S_\rho^n) = \text{vol}(B_\pi(p)) = \text{vol}(M).\end{aligned}$$

Thus we have equality in all estimates involved! First this gives $f(\pi) = 1$, hence $f(r) = 1$ $(0 \leqslant r \leqslant \pi)$. And next $\overline{B}_r(p) \cup \overline{B}_{\pi-r}(q) = M^n$, which says: For each $x \in M^n$ (put $r := d(x,p)$) we have $d(p,x) + d(x,q) = d(p,q)$, so that together the shortest geodesics from p to x and from x to q are segments from p to q. All geodesics starting at p therefore reach q precisely at distance π.

2.4.3. Along all of these segments we must have equality in the estimates leading to (1.9.4). First, the mean curvature of the distance spheres along each segment is $h(r) = h^\rho(r) = ct_\rho(r)$ (1.7.2, 1.9.2). Equality in (1.5.8) requires equality in (1.5.7), i.e. $S_r = h(r) \cdot id$. This gives (1.3.1) $R(Y,N)N = \rho \cdot Y$ (i.e. all sectional curvatures of 2-planes containing tangents to segments from p to q are $= \rho$). For such R_N the Jacobi equation (1.2.1) can be solved explicitly $J(r) = J'(0) \cdot s_\rho(r)$ (up to parallel translation), and (1.2.2) implies that M^n can be mapped isometrically to S_ρ^n by sending segments from p to q in M^n isometrically to meridians from pole to pole in S_ρ^n.

3. BUSEMANN FUNCTIONS

3.1. DEFINITION. The Busemann function b of a ray c is defined as an increasing limit of (shifted) distance functions:

$$b(x) := \lim_{t\to\infty} (t - d(x, c(t))) \leqslant d(x, c(0))$$

Note $d(c(t_2), c(t_1)) = t_2 - t_1$ if $0 \leqslant t_1 \leqslant t_2$, hence $t_2 - d(x, c(t_2)) \geqslant (t_1 - d(x, c(t_1)))$ by the triangle inequality. We have the Lipschitz bound $|b(x) - b(y)| \leqslant d(x, y)$. The sets $\{x \in M;\ b(x) > a\}$ are called horoballs.

3.2. LOCAL SUPPORT FUNCTIONS.

CLAIM: *For every $y \in M$ there is a unit vector $Y \in T_yM$ such that $c_Y(r) := \exp_y r \cdot Y$ is a ray and*

$$b(x) \geqslant b(y) + r - d(x, c_Y(r)) =: b_{Y,r}(x), \qquad (3.2.1)$$

in particular

$$b(c_Y(r)) = b(y) + r.$$

NOTE. By construction $c_Y(r)$ is not in the cut locus $C(y)$ and hence vice versa (K.4.2); the distance function $x \to d(x, c_Y(r))$ is therefore *differentiable* at y and—because of (3.2.1)—its level sphere through y stays inside the horoball. The ray c_Y need not be unique, so b is not differentiable in general, but if so then $\operatorname{grad} b(y) = Y$.

Proof of 3.2. To construct the ray $c_Y(r)$ let $\gamma_t(r)$ be a minimizing geodesic from y to $c(t)$ (c the given ray). For a subsequence $(t_n \to \infty)$ we have convergence of the initial unit directions $\gamma'_{t_n}(0)$ to some $Y \in T_yM$, and $c_Y(r) := \exp_y r \cdot Y$ is a ray (since each subarc is a limit of segments). To prove the inequality, we have by definition

$$r = d(y, \gamma_t(r)) = d(y, c(t)) - d(\gamma_t(r), c(t))$$
$$\text{if} \quad 0 \leqslant r \leqslant d(y, c(t))$$

and

$$b(x) - b(y) = \lim_{t \to \infty} \left(t - d(x, c(t)) - t + d(y, c(t)) \right).$$

The triangle estimate $d(x, c(t)) \leqslant d(x, c_Y(r)) + d(c_Y(r), c(t))$ and the first relation are inserted in the second:

$$\begin{aligned} b(x) - b(y) \geqslant \lim_{t_n \to \infty} \Big(-d(x, c_Y(r)) - d(c_Y(r), c(t_n)) \\ + r + d(\gamma_{t_n}(r), c(t_n)) \Big) \\ \geqslant -d(x, c_Y(r)) + r + \lim_{t_n \to \infty} -d(\gamma_{t_n}(r), c_Y(r)) \\ = -d(x, c_Y(r)) + r \qquad \text{(by definition of } c_Y(r)). \end{aligned}$$

3.3. *Hessian estimates.* For the local support functions $b_{Y,r}$ we translate (1.7.1, 1.7.2) into:

$$\text{If ric} \geqslant 0 \text{ on } M \text{ then trace} \, D \, \text{grad} \, b_{Y,r}(y) \geqslant -\frac{n-1}{r}, \quad (3.3.1)$$

$$\text{if } K \geqslant 0 \text{ on } M \text{ then } D \, \text{grad} \, b_{Y,r}(y) \geqslant -\frac{1}{r} \cdot id. \quad (3.3.2)$$

Inequality (3.3.1) allows us to use the Calabi-Hopf maximum principle, see (3.7) below. With (3.3.2) we are only one more argument away from the next result (3.5).

3.4. DEFINITION. A subset $A \subset M$ is called totally convex if for arbitrary $p, q \in M$ *all* geodesic connections (not just the minimizing ones) are in A.

3.5. COMPACT TOTALLY CONVEX EXHAUSTION (Cheeger-Gromoll-Meyer 1969, 1972). Let M be complete, noncompact, $K \geqslant 0$. Let c_α denote the set of all rays from some point $p \in M$, b_α

the corresponding Busemann functions and $bm := \max_\alpha b_\alpha$. Then:

a) The sublevels of Busemann functions $\{x \in M;\ b(x) \leqslant a\}$ are totally convex.

b) The sublevels of bm are a continuous exhaustion of M by compact totally convex sets.

3.5.1. *Classification* (Gromoll-Meyer 1969). Assume in addition to (3.5) $K > 0$ then:

a) No level surface of a Busemann function or of bm contains a geodesic arc, i.e., the minimal level surface of bm is a point.

b) M^n is diffeomorphic to $\mathbf{R}^n$.

3.5.2 *Classification* (Cheeger-Gromoll 1972). Let M^n be complete, noncompact, and $K \geqslant 0$. Then M^n contains a compact, totally geodesic, totally convex submanifold S ("soul") and M^n is diffeomorphic to the normal bundle of S in M^n.

Proof of (3.5). Let γ: $[0, L] \to M$ be a not necessarily minimizing geodesic arc of length L. Assuming $b(\gamma(0)) = a$, $b(\gamma(L)) = a_1 \leqslant a$ we have to show $b \circ \gamma(t) \leqslant a$. By possibly shortening γ we may assume $a_1 = a$. If the continuous function $\tilde{h}(t) := b(\gamma(t)) - a$ has a positive maximum 2μ then

$$h(t) := b(\gamma(t)) - a - \mu \cdot L^{-2} \cdot t(L-t)$$

also has an (interior!) maximum $\geqslant \mu$ at some t_0. Consider the local support functions $b_{Y,r}$ of b at $y = \gamma(t_0)$. The local, smooth functions

$$h_{Y,r}(t) := b_{Y,r}(\gamma(t)) - a - \mu L^{-2} \cdot t(L-t)$$

also have a local maximum at t_0, but from (3.3.2)

$$h''_{Y,r}(t_0) \geqslant -\frac{1}{r} + 2\mu L^{-2} > 0 \qquad \text{for large } r.$$

The contradiction proves (3.5a).

For b) it is clear that the intersection of totally convex sets is totally convex. We have to show that the sublevels of bm are compact. If not we would find a divergent sequence q_n with $bm(q_n) \leqslant a$ (assume $a \geqslant 0 = bm(p)$, p from (3.5)) and segments γ_n from p to q_n; by convexity $bm \circ \gamma_n \leqslant a$. Any limit of $\{\gamma_n\}$ is a *ray* in the sublevel $\{bm \leqslant a\}$, contradicting the definition of bm.

Proof of (3.5.1a). Let b be a Busemann function. We improve (3.3.2) to

3.5.3. *For every $y \in M$ there exists $\epsilon > 0$ and $R > 0$ such that the local support functions $b_{Y,r}$ (3.2) satisfy*

$$D \operatorname{grad} b_{Y,r}(y) \geqslant \epsilon \cdot id, \qquad \textit{if} \quad r \geqslant R.$$

Namely, put $3k := \min\{\text{sectional curvatures of } M \text{ in the unit ball around } y\}$, $\epsilon = k$, $R = 2 + (1/k)$. The Riccati equation and inequality (1.5.1) control the principal curvatures of the level surfaces of the local support functions $b_{Y,r}$ (of 3.2). Any solution $\kappa(r)$ with $\kappa(0) \leqslant \epsilon$ of $\kappa' \leqslant -K(u \wedge N) - \kappa^2 \leqslant -3k$ (1.5.1) drops on $[0,1]$ at least to $\epsilon - 3k < -\epsilon$ and, because of (1.6), then stays below $F(r) = (r - 1 - (1/\epsilon))^{-1}$, which solves $F' = -F^2$, $F(1) = -\epsilon$. For $r \geqslant R$ the $b_{Y,r}$ are defined beyond the pole of F so that, by (1.6.3), their level surfaces cannot have principal curvatures $\leqslant \epsilon$ at y. This proves (3.5.3), which in turn implies that no Busemann function can be constant on any geodesic arc (no weak interior maximum) —proving (3.5.1a). (To deal with bm extended 3.2.1 to $\sup b_\alpha(x) \geqslant \sup(b_\alpha(y) + r - b_{Y_\alpha,r}(x))$.

b) The proof of the diffeomorphism statement is very similar to (5.4.3) and I omit it.

The proof of (3.5.2) still requires a lot of work since the minimal level of bm is not yet the soul. One needs that lower dimensional totally convex sets are top dimensional in some totally geodesic submanifold, see **[CE]**.

3.6. Ricci Splitting Theorem (Cheeger-Gromoll 1971). *Let M^n be complete and assume* $\mathrm{ric} \geqslant 0$. *If M^n contains a line then this line splits off as a Riemannian factor, $M^n = N^{n-1} \times \mathbf{R}$.*

3.7. Calabi-Hopf Maximum Principle (1957). *Let (M, g) be a connected Riemannian manifold and f a continuous function on M.*

Assume, that for any $x \in M$ *and any* $\epsilon > 0$ *there exists a* C^∞*-support function* $f_{x,\epsilon} \leqslant f$, $f_{x,\epsilon}(x) = f(x)$ *(see* 3.2*) with*

$$-\Delta f_{x,\epsilon} = \operatorname{trace} D \operatorname{grad} f_{x,\epsilon} \geqslant -\epsilon.$$

Then f *attains no maximum unless it is constant.*

Proof of 3.6 (Eschenburg-Heintze 1984). Let γ be a line in M. Consider the two Busemann functions $b_\pm$ for the rays $c_\pm(t) := \gamma(\pm t)$. They satisfy

$$b_+(x) + b_-(x) = \lim_{t \to \infty} \left(t - d(x, \gamma(t)) + t - d(x, \gamma(-t))\right) \leqslant 0$$

and

$$(b_+ + b_-)(\gamma(t)) = 0 \qquad \text{(maximum!)}.$$

The sums of local support functions for b_+, b_- (3.2) satisfy the assumptions of the maximum principle (3.7) because of (3.3.1), which implies $b_+ + b_- = 0$! The level surfaces of $b_+ = -b_-$ can now be touched by large spheres from both sides:

$$b_{Y,r}(x) \leqslant b_+(x) = -b_-(x) \leqslant -b_{-Y,r}(x), \qquad \text{equality at} \quad x = y.$$

In particular $b_+ = -b_-$ is differentiable, $|\operatorname{grad} b_+| = 1$, the radial rays c_Y (3.2) are the integral curves of $\operatorname{grad} b_+$, and all the rays extend to lines. The proof can be finished with elementary arguments, but another application of the maximum principle works more elegantly. Every Busemann function b is subharmonic ($\operatorname{ric} \geqslant 0$!) (indeed, if b agrees with some harmonic function h on the boundary of some geodesic ball, then (3.7) implies that $b - h$ cannot have a positive maximum, i.e. $b \leqslant h$). Therefore $b_+ = -b_-$ is sub- and superharmonic, hence harmonic, hence C^∞ and $\Delta b_+ = 0$. Finally we have another equality discussion as in (2.4.3): b_+ is a distance function (1.1.2) whose level surfaces have constant mean curvature 0. (1.5.8) implies $0 \leqslant -\operatorname{ric} \circ \gamma$ ($\leqslant 0$). Therefore we must have equality in Schwarz' inequality (1.5.7): $S = 0 \cdot id$. From this and (1.3.1) we have $R(\ , \operatorname{grad} b_+) \operatorname{grad} b_+ = 0$. The Jacobi equation

(1.2.1) along the rays has explicit (parallel!) solutions and F: $b_+^{-1}(0) \times \mathbf{R} \to M^n$, $f(x, t) := \exp_x t \cdot \operatorname{grad} b_+(x)$ is an isometry.

4. THE ALEKSANDROW-TOPONOGOV ANGLE COMPARISON THEOREMS

Results closely related to those in this chapter were used by E. Cartan in the twenties and by Preissmann (1943) under curvature assumptions $K \leqslant 0$. Aleksandrow used triangle comparison theorems in the forties and fifties as a substantial tool, in particular in his theory of convex surfaces. Some years after Rauch, Toponogov proved in the Riemannian context the n-dimensional angle comparison for lower curvature bounds $\delta \leqslant K$ (1959). Remarkably, this theorem is true without any size restrictions. The proof, originally long and technical, has been considerably simplified.

4.0. DEFINITION. A triangle T in a Riemannian manifold is given by its three geodesic edges (which I assume minimizing although generalizations can be handled with the same proof). Assume lower curvature bounds $\delta \leqslant K$ or upper bounds $K \leqslant \Delta$. A triangle with the *same edgelengths* as T in the plane of constant curvature M_δ, respectively, M_Δ is called an "Aleksandrow triangle" T_δ, respectively, T_Δ. The two segments and the angle between them is called a hinge; a "Rauch hinge" in M_δ, respectively, M_Δ has the same edgelength, angle, edgelength as occur at one vertex of T.

4.1. TRIANGLE COMPARISON THEOREMS ASSUMING $K \leqslant \Delta$. *Size restrictions on T are necessary, namely, T does not meet the cut locus of its vertices, and, the circumference satisfies* $l(T) < 2\pi\Delta^{-1/2}$ *(ignore this, if* $\Delta \leqslant 0$*). Then an Aleksandrow triangle* T_Δ *exists and the angles of T and* T_Δ *satisfy*

$$\alpha \leqslant \alpha_\Delta, \qquad \beta \leqslant \beta_\Delta, \qquad \gamma \leqslant \gamma_\Delta. \tag{4.1.1}$$

The third edge c_Δ^* *closing a Rauch hinge in* M_Δ *satisfies*

$$|c| \geqslant |c_\Delta^*|. \tag{4.1.2}$$

With the obvious definitions of corresponding points on the edges of T and T_Δ and secants σ, σ_Δ between them one has

$$|\sigma| \leqslant |\sigma_\Delta|. \tag{4.1.3}$$

4.2. TRIANGLE COMPARISON THEOREMS ASSUMING $\delta \leqslant K$ *(Toponogov).*

4.2.0. *An Aleksandrow triangle T_δ always exists, more precisely: a circumference $l(T) > 2\pi\delta^{-1/2}$ (if $\delta > 0$) does not occur and $l(T) = 2\pi\delta^{-1/2}$ occurs only on S_δ^n; if $l(T) < 2\pi\delta^{-1/2}$ then the three triangle inequalities in T are sufficient for the existence of T_δ.*

The angles of T and T_δ satisfy (take $l(T) < 2\pi\delta^{-1/2}$, 4.2.0)

$$\alpha_\delta \leqslant \alpha, \qquad \beta_\delta \leqslant \beta, \qquad \gamma_\delta \leqslant \gamma. \tag{4.2.1}$$

The third edge c_δ^* closing a Rauch hinge in M_δ satisfies

$$|c| \leqslant |c_\delta^*|. \tag{4.2.2}$$

Secants σ, σ_δ between corresponding points on T, respectively, T_δ satisfy

$$|\sigma| \geqslant |\sigma_\delta|. \tag{4.2.3}$$

4.3. REMARKS. Because of the cosine law in M_δ, respectively, M_Δ (e.g. on S^n: $\cos c = \cos a \cos b + \sin a \sin b \cos \gamma$) it is trivial that the opposite edgelength of a hinge varies monotonely increasing with the hinge angle. (4.1.1) and (4.1.2) are therefore immediately equivalent, and so are (4.2.1) and (4.2.2). If one considers the limit of short secants across a vertex, then (4.1.3) implies (4.1.1) and (4.2.3) implies (4.2.1), again immediately. The converse (4.1.1) $\Rightarrow$ (4.1.3) and (4.2.1) $\Rightarrow$ (4.2.3) is also true but needs more trigonometry into which I do not want to go. (4.1.3) and (4.2.3) extend immediately to infinite triangles if $\Delta < 0$ or $\delta < 0$. They also extend to Gromov's limits of Riemannian spaces in which minimizing curves exist but angles cannot be defined. I shall prove (4.1) using estimates on $\partial \exp$ (1.2.2, 1.8.2) and Toponogov's theorem (4.2) with the Hessian bounds (1.7.1). It is enough to prove (4.2.3) for the secants to the opposite vertex, see (4.5.4).

4.4. *Proof* of (4.1.2). Because of the size restrictions we can parametrize the edge c of T in exponential coordinates from the opposite vertex, i.e.

$$c(t) = \exp_p X(t), \qquad X\colon [0,1] \to T_pM \qquad \left(|X| < \pi\Delta^{-1/2}\right).$$

We identify T_pM and $T_{p_\Delta}M_\Delta$ isometrically (e.g. with $\mathbf{R}^n$) and define

$$\tilde{c}(t) := \exp_{p_\Delta} X(t), \qquad X\colon [0,1] \to T_{p_\Delta}M_\Delta \approx T_pM.$$

Now follow the procedure outlined in (1.2.2):

$$\dot{c}(t) = \partial \exp_p\Big|_{X(t)} \cdot \dot{X}(t) = \frac{\partial}{\partial t}\left(\exp_p s \cdot X(t)\right)\Big|_{s=1} = J_t(1),$$

$$\dot{\tilde{c}}(t) = \partial \exp_{p_\Delta}\Big|_{X(t)} \cdot \dot{X}(t) = \tilde{J}_t(1);$$

and the Jacobi fields have initial conditions

$$J_t(0) = 0 = \tilde{J}_t(0), \qquad \frac{D}{\partial s} J_t(0) = \dot{X}(t) = \frac{D}{ds}\tilde{J}_t(0)$$

so that (1.8.2), combined with (1.4.5), proves

$$|\dot{c}(t)| \geqslant |\dot{\tilde{c}}(t)|, \qquad \text{hence length}\,(c) \geqslant \text{length}\,(\tilde{c}).$$

Finally length $(\tilde{c}) \geqslant |c_\Delta^*|$ since $\tilde{c}$ is some connection of its endpoints and c_Δ^* is the shortest one.

Remark. The corresponding proof of (4.2) gets considerably more complicated, mainly because one does not want the size restrictions. Also, the triangle inequality at the very end has to be used in M rather than in M_δ, i.e. the proof has to start by parametrizing c_δ^*. This causes the corresponding definition of $\tilde{c}$ in M to run into problems with conjugate points.

zur Gleichheitsdiskussion.
Falls $\delta \geq 0$, $\lambda \geq 0$, λ''_{loc} ... ⇒ ...
Der Beweis zeigt auch: Es gibt keine Lösung von $f''+\delta f = c$ die auf einem Teilintervall oberhalb von λ, am Rande $=\lambda$ ist.
Daraus resultiert die Gleichheitsdiskussion.

4.5. *Proof* of (4.2.3).

4.5.1. Because of (2.4.2) (and its simple proof) we can start with the alternative: either $\operatorname{diam}(M) = \pi \cdot \delta^{-1/2}$, then M^n is isometric to S^n_δ or else $\operatorname{diam}(M) \leqslant \pi \cdot \delta^{-1/2} - 2\epsilon$. Of course we know the sphere, therefore we have during the following proof all segments $\leqslant \pi\delta^{-1/2} - 2\epsilon$ which eliminates many separate considerations. It helps the exposition without lengthening the full argument very much if I treat the simpler cases separately.

4.5.2. First assume that no edge of T meets the cut locus of a vertex. Let r, respectively, r_δ denote the (differentiable) distance function from the vertex $p \in T$ to get the modified distance functions $md_\delta \circ r$, respectively, $md_\delta \circ r_\delta$. Their restrictions to the opposite edge c, respectively, c_δ are—because we stay away from the cut locus—differentiable functions

$$h := md_\delta \circ r \circ c, \qquad H := md_\delta \circ r_\delta \circ c_\delta. \tag{4.5.3}$$

They satisfy

$$\begin{aligned} &h'' + \delta h \leqslant 1, \qquad &&H'' + \delta H = 1 \quad \text{with} \\ &h(0) = H(0), \qquad &&h(|c|) = H(|c|) \end{aligned}$$

because of the Hessian estimate (1.7.1).

4.5.4. *Claim*. The difference $\lambda := h - H$ satisfies

$$\lambda'' + \delta \cdot \lambda \leqslant 0, \qquad \lambda(0) = 0, \qquad \lambda(|c|) = 0,$$

which implies

$$\lambda \geqslant 0 \qquad \text{on} \quad [0, |c|].$$

(In fact, $\lambda(0) \geqslant 0$, $\lambda(|c|) \geqslant 0$ is enough, and handles the case of general secants, (4.3).)

Proof. a) If $\delta < 0$ then $\lambda'' \leqslant (-\delta) \cdot \lambda$ shows that λ cannot have a negative minimum, i.e. $h \geqslant H$ (4.2.3).

b) If $\delta = 0$ then λ cannot have a negative minimum $= -2\mu$ since

$$\tilde{\lambda} := \lambda + \mu \cdot |c|^{-2} \cdot t(|c| - t)$$

would also have an (interior) minimum $\leqslant -\mu$, contradicting $\tilde{\lambda}'' \leqslant -2\mu|c|^{-2} < 0$.

c) If $\delta > 0$ we have from (4.5.1) $|c| \leqslant \pi\delta^{-1/2} - 2\epsilon$, hence

$$\sigma_\epsilon(t) := s_\delta(t+\epsilon) - s_\delta\left(\frac{\epsilon}{2}\right) > 0 \qquad \text{for } 0 \leqslant t \leqslant |c|.$$

Again, suppose λ has some negative minimum. Then $\tilde{\lambda} := \lambda/\sigma_\epsilon$ also has a negative minimum $-\mu$ at, say, $t_0 \in (0, |c|)$. But this contradicts

$$\tilde{\lambda}'(t_0) = (\lambda'\sigma_\epsilon - \lambda\sigma_\epsilon') \cdot \sigma_\epsilon^{-2}(t_0) = 0,$$

$$\begin{aligned} \tilde{\lambda}''(t_0) &= (\lambda''\sigma_\epsilon - \lambda\sigma_\epsilon'') \cdot \sigma_\epsilon^{-2}(t_0) \\ &= \left((\lambda'' + \delta\lambda) \cdot \sigma_\epsilon + \delta \cdot \lambda \cdot s_\delta\left(\frac{\epsilon}{2}\right)\right) \cdot \sigma_\epsilon^{-2}(t_0) \\ &\leqslant -\delta \cdot \mu \cdot s_\delta\left(\frac{\epsilon}{2}\right) \cdot \sigma_\epsilon^{-1}(t_0) < 0. \end{aligned}$$

4.5.5. In a second step we allow the edge c to meet the cut locus of the opposite vertex p, but *not* in a conjugate point. Then each segment from p to $c(t_0)$ is *locally* minimizing beyond $c(t_0)$. So we get differentiable local distance functions

$$r_{\text{loc}} \geqslant r, \qquad \text{defined in particular near } c(t_0).$$

The principal curvature estimates (1.7.1) were stated for these local distance functions. In addition to (4.5.3) we now also have for each $t_0 \in (0, |c|)$

$$\begin{gathered} h_{\text{loc}} := md_\delta \circ r_{\text{loc}} \circ c|_{[t_0-\tau,\, t_0+\tau]}, \\ h_{\text{loc}}'' + \delta \cdot h_{\text{loc}} \leqslant 1, \qquad h_{\text{loc}} \geqslant h, \qquad h_{\text{loc}}(t_0) = h(t_0). \end{gathered} \tag{4.5.6}$$

This is enough to make the proof of (4.5.4) work. At that t_0 at which the negative minimum of $\tilde{\lambda}$ occurs we replace h by h_{loc}, and get also $\lambda_{\text{loc}} \geqslant \lambda$ and $\tilde{\lambda}_{\text{loc}} \geqslant \tilde{\lambda}$, $\tilde{\lambda}_{\text{loc}}(t_0) = \tilde{\lambda}(t_0)$. The smooth function $\tilde{\lambda}_{\text{loc}}$ therefore also has a local interior minimum at t_0, but $\tilde{\lambda}''_{\text{loc}}(t_0) < 0$ as before.

4.5.7. In the last step we deal with conjugate endpoints of segments from p to $c(t_0)$. We cannot get local superfunctions which satisfy the same differential inequality as before, but they almost do.

Let $\gamma(s)$ be a segment from p to $c(t_0)$ and define for small $\eta > 0$ local superdistance functions, differentiable near $c(t_0)$, by

$$r_{\text{loc},\eta}(x) := d(x, \gamma(\eta)) + \eta \geqslant r(x) = d(x, \gamma(0)).$$

The Hessian of $md_\delta \circ r_{\text{loc},\eta}$ has (as in the computation for (1.7.1)) the radial eigenvalue $c_\delta(r_{\text{loc},\eta})$ and spherical eigenvalues

$$\kappa_{sp} \leqslant ct_\delta(r_{\text{loc},\eta} - \eta) \cdot s_\delta(r_{\text{loc},\eta}) = c_\delta(r_{\text{loc},\eta}) + \frac{s_\delta(\eta)}{s_\delta(r_{\text{loc}} - \eta)}. \tag{4.5.8}$$

We can assume $r_{\text{loc}}(c(t_0)) > 0$ (since we do not need to prove anything if the vertex p lies on the opposite edge, i.e. $|a| + |b| = |c|$) and replace the denominator in (4.5.8) by a constant; this also uses (4.5.1) once more! With this we get instead of (4.5.6)

$$\begin{aligned} &h_{\text{loc},\eta} := md_\delta \circ r_{\text{loc},\eta} \circ c|_{[t_0-\tau,\, t_0+\tau]}, \\ &h_{\text{loc},\eta} \geqslant h, \qquad h_{\text{loc},\eta}(t_0) = h(t_0), \end{aligned} \tag{4.5.9}$$

$$h''_{\text{loc},\eta} + \delta \cdot h_{\text{loc},\eta} \leqslant 1 + \text{const} \cdot s_\delta(\eta), \qquad \text{const. independent of } \eta.$$

This is *still* good enough to make the proof of (4.5.4) work. At t_0 where the continuous function $\tilde{\lambda}$ has its negative minimum $\leqslant -\mu$, the smooth functions $\tilde{\lambda}_{\text{loc},\eta}$ *all* have a minimum. As before

$$\text{if} \quad \delta \leqslant 0 \quad \text{then } \tilde{\lambda}''_{\text{loc},\eta}(t_0) \leqslant -2\mu|c|^{-2} + \text{const} \cdot s_\delta(\eta),$$

$$\text{if} \quad \delta > 0 \quad \text{then } \tilde{\lambda}''_{\text{loc},\eta}(t_o) \leqslant -\mu \cdot \delta \cdot s_\delta\left(\frac{\epsilon}{2}\right) \cdot \sigma_\epsilon(t_0)^{-1} + \text{const} \cdot s_\delta(\eta),$$

which gives the same contradiction ($\tilde{\lambda}''_{\rm loc}(t_0) < 0$) for sufficiently small η.

4.6. *Triangles T with circumference* $= 2\pi\delta^{-1/2}$ *occur only on* S^n_δ *and triangles with circumference* $l(T) > 2\pi\delta^{-1/2}$ *do not occur.*

Proof. (i) Again with (4.5.1) we may assume $\mathrm{diam}(M) \leqslant 2\pi\delta^{-1/2} - 2\epsilon$. The proof given in (4.5) formally includes the case $l(T) = 2\pi\delta^{-1/2}$, but this does not occur: T_δ would have to be a great circle so that each edge meets the antipodal point of the opposite vertex. Now (4.2.3) requires a segment of length $\pi \cdot \delta^{-1/2}$ in M. (ii) A triangle with $l(T) > 2\pi\delta^{-1/2}$ by continuity gives a triangle with $l(T') = 2\pi\delta^{-1/2}$, so $M = S^n_\delta$—contradicting $l > 2\pi\delta^{-1/2}$.

5. APPLICATIONS OF THE TRIANGLE THEOREMS

5.1. *Number of generators for* $\pi_1(M,p)$ (Gromov 1978).

5.1.1. *Let* M^n *be complete and assume* $K \geqslant 0$. *Then the fundamental group* $\pi_1(M,p)$ *can be generated by* $N \leqslant 2 \cdot 5^{(1/2)n}$ *elements (compare* 2.3.3*).*

5.1.2. *Let* M^n *be compact. Assume curvature and diameter bounds* $-\Lambda^2 \leqslant K$, $\mathrm{diam}(M) < \frac{1}{2}D$. *Then the fundamental group* $\pi_1(M,p)$ *can be generated by* $N \leqslant 2 \cdot (3 + 2\cosh \Lambda D)^{(1/2)n}$ *elements. (If* $n = 2$ *then this bound is* $\geqslant 5 \cdot \mathit{genus}^4$*, hence never sharp.)*

Proof. Define for each $\alpha \in \pi_1(M, p)$ the "length" $|\alpha|$ as the length of a shortest geodesic loop in the homotopy class α. Now define a "short" set of generators $\{\alpha_1, \ldots, \alpha_N, \ldots\}$:

(i) α_1 is a shortest element in $\pi_1(M, p) \setminus \{id\}$.

(ii) If $\alpha_1, \ldots, \alpha_k$ are already chosen, denote by $\langle \alpha_1, \ldots, \alpha_k \rangle$ the subgroup of $\pi_1(M, p)$ which they generate. Then α_{k+1} is a shortest element in $\pi_1(M, p) \setminus \langle \alpha_1, \ldots, \alpha_k \rangle$.

5.1.3. CLAIM. *No short set of generators contains more elements than the bound in (*5.1.1*) respectively, (*5.1.2*).*

Proof of the claim. By definition $|\alpha_1| \leqslant |\alpha_2| \leqslant \cdots$ and also

$$\left|\alpha_i \cdot \alpha_j^{-1}\right| \geqslant \max\{|\alpha_i|, |\alpha_j|\} \tag{5.1.4}$$

otherwise α_i or α_j was not chosen minimally.

Now apply Toponogov's theorem to the triangle T obtained in the universal cover $\tilde{M}$ by lifting the shortest geodesic loops which represent the classes $\alpha_i, \alpha_j, \alpha_i\alpha_j^{-1}$; the edgelengths of T are then $|\alpha_i|, |\alpha_j|, |\alpha_i\alpha_j^{-1}|$. (4.2.1) gives a lower bound for the angle Φ opposite the longest edge:

$$(K \geqslant 0) \quad \left|\alpha_i \cdot \alpha_j^{-1}\right| \leqslant |\alpha_i|^2 + |\alpha_j|^2 - 2|\alpha_i||\alpha_j|\cos\Phi, \tag{5.1.5}$$

i.e. $\Phi \geqslant 60°$;

$$(K \geqslant -\Lambda^2) \quad \cos\Phi \leqslant \frac{\cosh\Lambda|\alpha_i| \cdot \cosh\Lambda|\alpha_j| - \cosh\Lambda|\alpha_i\alpha_j^{-1}|}{\sinh\Lambda|\alpha_i| \cdot \sinh\Lambda|\alpha_j|}, \tag{5.1.6}$$

i.e.

$$\cos\Phi \leqslant \frac{\cosh\Lambda D}{1 + \cosh\Lambda D} \quad \text{or} \quad \sin^{-2}\frac{1}{2}\Phi \leqslant 2 + 2\cosh\Lambda D.$$

(The last inequality uses $|\alpha_j| \leqslant D$, i.e. generators can always be chosen $\leqslant 2 \cdot \operatorname{diam}(M) + \epsilon$, by dividing any loop from p into segments shorter than ϵ and joining the dividing points with p, back and forth.)

Finally, consider the initial unit vectors of the short loops representing $\alpha_1, \alpha_2, \ldots$. They are points on the unit sphere in T_pM with mutual distance $\geqslant \Phi$. The open balls of radius $\sin\frac{1}{2}\Phi$ are therefore disjoint and the inner halfs of these balls are all contained in the ball of radius $(1 + \sin^2\frac{1}{2}\Phi)^{1/2}$. Therefore their number N is bounded by the volume ratio, i.e.

$$N \leqslant 2 \cdot \left(1 + \sin^2\frac{1}{2}\Phi\right)^{n/2} \cdot \left(\sin\frac{1}{2}\Phi\right)^{-n} = 2\left(1 + \sin^{-2}\frac{1}{2}\Phi\right)^{+n/2}. \tag{5.1.7}$$

Insertion of (5.1.5) respectively, (5.1.6) into (5.1.7) proves the theorem.

5.2. *Critical points of the distance function.* The following definition has turned out to be very useful and suggestive. The arguments in this section have been developed in connection with the sphere theorems.

5.2.1. DEFINITION. A point $q \in M$, $q \neq p$, is called *not critical* for the distance function from p, if the initial tangents of *all* segments qp lie in an open halfspace in T_qM. Otherwise q is called critical.

The following fact justifies the name.

5.2.2. *Existence of a gradient like vector field.* Let A be a compact set in M with no critical points for the distance function from p. Then one has a vector field X on an open neighborhood of A such that the distance from p is strictly decreasing along the integral curves of X.

Proof. For every $x \in A$ we have by definition a vector $Y_x \in T_xM$ and $\epsilon_x > 0$ such that the angle between Y_x and any segment xp is $\leqslant (\pi/2) - 2\epsilon_x$. By radial parallel translation we extend Y_x to a smooth local vector field X_x on some ball $B(x)$. We can choose $B(x)$ so small that for all $y \in B(x)$ the angle between $X_x(y)$ and any segment yp is $\leqslant (\pi/2) - \epsilon_x$. Denote by $B'(x)$ the concentric balls of half the radius as $B(x)$ and choose (by compactness of A) finitely many $B'(x_i)$ which cover A. With nonnegative C^∞-functions φ_i which are 1 on $B'(x_i)$ and 0 outside $B(x_i)$ we define a smooth vector field

$$X := \frac{1}{\Sigma\varphi_i} \cdot \Sigma\varphi_i X_{x_i}.$$

The following now holds for each $x \in \cup B'(x_i)$: Each segment xp has an angle $< \pi/2$ with all $\varphi_i X_{x_i}(x)$ which are $\neq 0$; therefore $X(x) \neq 0$ and the angle between xp and $X(x)$ is $< \pi/2$. So X is the desired vector field on $\cup B'(x_i) \supset A$.

Of course, (5.2.2) alone cannot be of too much help. The discovery was that the existence of "large" distances in M can be combined with *Toponogov's theorem* (4.2.2) to draw rather strong conclusions about these critical points.

5.2.3. ASSUMPTION. Let M be complete and $0 < \delta \leqslant K$. We shall consider critical points for the distance function from p. Let q be a *farthest* point from p and assume $D := d(p, q) > (\pi/2)\delta^{-1/2}$. (Note: In general p is not farthest from q.)

CONCLUSIONS.
5.2.4. q is a unique farthest point and critical for the distance function from p (compare 6.1.2).
5.2.5. A point $x \neq p, q$ is *not* critical for the distance from p if

a) D is maximal, i.e. $D = \operatorname{diam}(M)$

or

b) if x is far from p, i.e. $d(x, p) \geqslant (\pi/2)\,\delta^{-1/2}$.

Any fixed segment xq can be used to define the open halfspace (in 5.2.1) since all segments xp have angles $\varphi > \pi/2$ (see 5.2.7) with the chosen segment xq.
5.2.6. Every $x \in M$ which is far from q is close to p

$$d(x, q) > \frac{\pi}{2}\delta^{-1/2} \Rightarrow d(p, x) < \frac{\pi}{2}\delta^{-1/2}.$$

In the middle the angle estimate is more precise than in (5.2.5),

$$d(p, x) = d(x, q) \Rightarrow \measuredangle(p, x, q) \geqslant D \cdot \delta^{1/2}.$$

Proofs. For uniqueness in (5.2.4) we show that the midpoint m on a segment between two (assumed) farthest points $q, \bar{q}$ (from p) has to be at a greater distance. Let $p_\delta q_\delta \bar{q}_\delta$ be the Aleksandrow triangle of $pq\bar{q}$ (4.0) and m_δ the midpoint of $q_\delta\bar{q}_\delta$. (4.2.3) implies $d(p, m) \geqslant d(p_\delta, m_\delta)$ and from $d(p_\delta, q_\delta) = d(p_\delta, \bar{q}_\delta) > (\pi/2)\,\delta^{-1/2}$ we have $d(m_\delta, p_\delta) > d(p_\delta, q_\delta) = d(p, q)$, as claimed.

Proof of (5.2.5). For each of the triangles pxq with $\varphi := \measuredangle(pxq)$ we have from (4.2.1),

$$\cos\varphi \leqslant \frac{\cos\delta^{1/2}D - \cos(\delta^{1/2}|qx|)\cdot\cos(\delta^{1/2}|px|)}{\sin(\delta^{1/2}|qx|)\cdot\sin(\delta^{1/2}|px|)} < 0, \quad (5.2.7)$$

where the last strict inequality follows since uniqueness in (5.2.4) supplies the following strict inequalities:

a) If $D = \operatorname{diam}(M)$ then $|px|, |qx| < D$.

b) If $d(x, p) \geqslant (\pi/2)\cdot\delta^{-1/2}$ then also $|px| < D$.

Now (5.2.5b) forces q to be critical in (5.2.4): For any $Y \in T_qM$ and all $0 < \epsilon < D - (\pi/2)\,\delta^{-1/2}$ we can take $x := \exp_q \epsilon\cdot Y$ in (5.2.5b). Any limit of segments xp (as $\epsilon \to 0$) is a segment qp. For this limit segment we conclude from $\measuredangle(xq, xp) > (\pi/2)$ (see 5.2.7) that $\measuredangle(-Y, qp) \geqslant (\pi/2)$—in other words: In every closed half-space of T_qM some segment to p starts.

Proof of (5.2.6). Fix a segment qx and choose with (5.2.4) a segment qp such that $\alpha := \measuredangle(qp, qx) \leqslant (\pi/2)$. For the Rauch hinge in M_δ corresponding to xqp in M we have (4.2.2)

$$\begin{aligned}
&\cos\delta^{1/2}|px| \\
&\underset{(4.2.2)}{\geqslant} \cos(\delta^{1/2}D)\cdot\cos(\delta^{1/2}|qx|) + \sin(\delta^{1/2}D)\cdot\sin(\delta^{1/2}|qx|)\cos\alpha \\
&> 0.
\end{aligned}$$

Finally, if $d(p, x) = d(q, x)$ then (5.2.7) improves to

$$\cos\varphi \leqslant \frac{\cos(\delta^{1/2}D) - \cos^2(\delta^{1/2}|qx|)}{1 - \cos^2(\delta^{1/2}|qx|)} \leqslant \cos(\delta^{1/2}D). \quad (5.2.8)$$

5.3. *Cut locus estimates* (see K.5).

5.3.1. (Klingenberg 1960). *Let M^n be compact, orientable, even dimensional. Assume curvature bounds* $0 < \mathrm{K} \leqslant 1$. *Then*

The cut locus distance satisfies $d(p, C(p)) \geqslant \pi$.

This estimate is sharp *not* only for round spheres, see (6.1.2).

5.3.2. (Klingenberg 1961). *Let M^n be compact, simply connected (and odd dimensional). Assume curvature bounds $\frac{1}{4} \leqslant K \leqslant 1$. Then*

$$d(p, C(p)) \geqslant \pi.$$

The (current) best counterexamples to an extension of (5.3.2) are given in (6.1.4).

5.3.3. (Easy version of 5.3.2). *Let M^n be compact. Assume curvature bounds $\frac{1}{4} < \delta \leqslant K \leqslant 1$ and* $\operatorname{diam}(M) > (\pi/2)\delta^{1/2}$. *Then*

$$d(p, C(p)) \geqslant \pi.$$

5.3.4. (Cheeger 1970). *Let M^n be compact. Assume curvature bounds $\delta \leqslant K \leqslant 1$. Then*

a) $\delta > 0$: $d(p, C(p)) \geqslant \min(\pi, \pi \cdot \delta^{-1/2} \cdot \operatorname{vol}(M) \cdot \operatorname{vol}(S^n_\delta)^{-1})$, *this is sharp for round spheres.*

b) $\delta \leqslant 0$: $d(p, C(p)) \geqslant$

$$\min\Big(\pi, \operatorname{vol}(M) \cdot \operatorname{vol}(S^{n-2})^{-1} \cdot (n-1) \cdot s_\delta^{1-n}(\operatorname{diam}(M))\Big),$$

this is never sharp.

5.3.5. (Toponogov). *Let M^n be complete and noncompact. Assume curvature bounds $0 < K \leqslant 1$ (i.e.* $\inf K = 0$ *by* 3.2.1*). Then*

$$d(p, C(p)) \geqslant \pi.$$

5.3.6. (Compact version of (5.3.5)). *Let M be compact. Assume curvature bounds $0 < \delta \leqslant K \leqslant 1$, and that for each $p \in M$ there is a $q \in M$ with $d(p, q) > (\pi/2)\delta^{-1/2}$. Then*

$$d(p, C(p)) \geqslant \pi.$$

REMARK. (5.3.3) is included since its proof is a nice application of critical points and the proof of (5.3.2) is too long to be given here. (5.3.6) is included since it explains why in the noncompact case (5.3.5) the parity of the dimension plays no role. (5.3.6) also says that the counterexamples to an extension of (5.3.2) are "small"—compared to what the curvature bounds would allow.

Proofs. 5.3.7. All proofs start with Klingenberg's result that a closest nonconjugate cut point q of p is the midpoint of a geodesic loop (K.4.4). If the cut locus distance happens to have a minimum $m < \pi$ at p, then (K.4.5) implies (by reversing the role of p and q) that we have a closed geodesic γ through p and q of length $2m = 2\pi - 2\eta$, $\eta > 0$. All proofs then show that such a closed geodesic cannot exist.

The proof of (5.3.1) uses the famous

5.3.8. SYNGE LEMMA (1926). *Let M^n be compact, orientable, even dimensional. Assume $K > 0$. Then any closed geodesic γ has shorter parallel curves.*

REMARK. Synge used his lemma to conclude that if M were *not* simply connected there would exist a *shortest closed* geodesic in a nontrivial homotopy class—contradicting (5.3.8).

Proof of (5.3.8). Parallel translation around γ is an orientation preserving isometry of the odd dimensional subspace $(\gamma'(0))^{\perp} \subset T_{\gamma(0)}M$ and therefore has a fixed vector—in other words, we have a closed parallel unit vector field $v(t) \perp \gamma'(t)$ along γ. The strip $c(\epsilon, t) := \exp_{\gamma(t)}\epsilon \cdot v(t)$ is similar to a band around the equator of S^2, i.e. the closed geodesic γ has *shorter* parallel curves $t \to c(\epsilon, t)$, $\epsilon > 0$ small. This follows from the second variation formula (K.4 Lemmata 1, 2):

$$\frac{d^2}{d\epsilon^2}L(\epsilon)\bigg|_{\epsilon=0} = -\int_0^{2m} g(R(v, \gamma')\gamma', v)\,dt < 0,$$

since we assumed $K > 0$.

We finish the proof of (5.3.1) with (4.1.1): Choose three equidistant points $\gamma(t_i)$ on γ (from (5.3.7) and consider for small $\epsilon > 0$ the triangle T with vertices $c(\epsilon, t_i)$ $(i = 1, 2, 3)$ and unique edges of length $< \frac{2}{3}m$ (using the second variation again). The circumference of T is short enough so that no edge meets the cut locus of the opposite vertex and (4.1.1) applies. As $\epsilon \to 0$ the edges of T converge to the arcs of γ between the $\gamma(t_i)$ (there are no other segments between the $\gamma(t_i)$) and thus all angles of T converge to π. By (4.1.1) the same is true for the angles of the Aleksandrow triangle T_Δ; its edgelengths converge (as $\epsilon \to 0$) to $\frac{2}{3}m = \frac{2}{3}(\pi - \eta)$. This contradicts spherical trigonometry: $\lim a(\epsilon) = \lim b(\epsilon) = \lim c(\epsilon) = \frac{2}{3}(\pi - \eta)$ and the spherical cosine law imply

$$\lim_{\epsilon \to 0} \cos \gamma(\epsilon) = \frac{\cos \frac{2}{3}(\pi - \eta)}{1 + \cos \frac{2}{3}(\pi - \eta)} > -1.$$

Proof of (5.3.4) (Heintze-Karcher (1978). Either $d(p, C(p)) \geqslant \pi$ and we are done, or there is the closed geodesic γ of (5.3.7) and we proceed by estimating the volume of M using the level surfaces of the distance function from γ (compare (1.9)). For every $p \in M$ there is a segment from p to γ which (necessarily) meets γ perpendicularly. The set Z of points p for which these segments are not unique has volume zero and $M \setminus Z$ is covered by unique minimizing geodesics to γ. Along these we exploit (1.5.1, 1.6). Choose orthonormal parallel vector fields $u_1, \ldots, u_{n-1}$, $u_1(0) = \gamma'$ along each segment $c \perp \gamma$. As in (1.7.1) we get for the shape operator S of the distance tubes around γ

$$g(Su_1, u_1)(r) \leqslant -\delta \frac{s_\delta}{c_\delta}, \quad g(Su_j, u_j) \leqslant \frac{c_\delta}{s_\delta}(r) \qquad (j = 2, \ldots, n-1)$$

$$\operatorname{trace} S_r \leqslant -\delta \frac{s_\delta}{c_\delta}(r) + (n-2) \frac{c_\delta}{s_\delta}(r). \tag{5.3.9}$$

As in (1.9.2) this implies (along each normal segment c)

$$a(r) \leqslant c_\delta \cdot s_\delta^{n-2}(r), \qquad 0 \leqslant r \leqslant l(c),$$

where $l(c) \leqslant \operatorname{diam}(M)$ denotes the distance up to which c is a minimizing segment to γ. If $\delta > 0$ we do not need separate diameter bounds (because of (1.7.1) no segment c of length $> (\pi/2)\cdot \delta^{-1/2}$ can be minimizing to γ). Since nonminimizing geodesics no longer contribute to the volume of the (outer) tubes we only worsen the volume estimate if we integrate the bound for $a(r)$ for $r \leqslant (\pi/2)\cdot\delta^{-1/2}$ respectively, $r \leqslant \operatorname{diam}(M)$ first at each $\gamma(t)$ over all normal directions and then along γ:

$$\operatorname{vol}_{n-1}(\text{level surface at distance } r \text{ from } \gamma)$$

$$\leqslant l(c)\cdot \operatorname{vol}(S^{n-2})\cdot c_\delta(r)\cdot s_\delta(r)^{n-2},$$

and a final radial integration proves (5.3.4):

b) $\operatorname{vol}(M) \leqslant l(c)\cdot\operatorname{vol}(S^{n-2})\cdot(n-1)^{-1}s_\delta(\operatorname{diam}(M))^{n-1}$,

a) $\operatorname{vol}(M) \leqslant l(c)\cdot\operatorname{vol}(S^{n-2})\cdot(n-1)^{-1}s_\delta((\pi/2)\cdot\delta^{-1/2})^{n-1}$

$$\leqslant \frac{1}{2\pi}\cdot\delta^{1/2}\cdot l(c)\cdot\operatorname{vol}(S^n_\delta).$$

The work to prove (5.3.5) was already done in (3.5): If for some $p \in M$ we had $d(p, C(p)) < \pi$ then we could find a compact totally convex set B with $p \in B$ (3.5b). Changing names we assume that the cut locus distance assumes its nonconjugate minimum on B at p. Again we have the loop γ from p through the closest cut point q. By total convexity this loop is in B, hence $d(q, C(q)) \geqslant d(p, C(p))$ and γ cannot have an angle $< \pi$ at q either (we repeated 5.3.7). By total convexity of the sublevels of the function *bm* (3.5) γ has to lie on a level of *bm* which contradicts (3.5.1).

To prove (5.3.6) we have the short closed geodesic γ from (5.3.7). Choose some $p \in \gamma$ and let $\bar q$ be a point farthest from p, i.e. $d(p, \bar q) > (\pi/2)\cdot\delta^{-1/2}$. Since the case $\delta > \frac14$ is covered by (5.3.3) we assume $\delta \leqslant \frac14$, i.e. $d(p,\bar q) > (\pi/2)\cdot\delta^{-1/2} \geqslant \pi > \frac12$length (γ). Next, let $q \in \gamma$ be a *closest* point to $\bar q$; note $\gamma \perp q\bar q$. We apply (4.2.2) to the 90°-hinge $\bar q q p$:

$$0 \underset{(\text{Choice of } \bar q)}{>} \cos\delta^{1/2}\cdot d(p,\bar q) \underset{(4.2.2)}{\geqslant} \cos\delta^{1/2} d(p,q)\cdot\cos\delta^{1/2} d(q,\bar q)$$

to conclude $d(q,\bar q) > \frac12\pi\cdot\delta^{-1/2}$ (note $d(p,q) \leqslant \frac12$ length $(\gamma) < \frac12\pi\cdot\delta^{-1/2}$). Now the contradiction arises for the same reason as in the noncompact case (3.5.3): $\bar{q}q$ is too long to even locally minimize the distance from $\bar{q}$ to the closed geodesic γ since the principal curvatures of the distance sphere around $\bar{q}$ through q are negative (1.7.1), i.e. convex to the outside.

To prove (5.3.3) consider the set $G := \{x \in M, d(x, C(x)) \geq \pi\}$ which is closed because of the continuity of the cut locus distance (K.4.3 corollary). We show $G \neq \emptyset$ and open, hence (by the assumed connectedness of M) $G = M$. To prove $G \neq \emptyset$ choose $y, \bar{q} \in M$ such that $d(y, \bar{q}) = \mathrm{diam}(M)$; to show G open choose $y \in \{x \in M; d(x, C(x)) > \frac{1}{2}\pi \cdot \delta^{-1/2}\}$, i.e. from an open neighborhood of G. In each case assume by contradiction $d(y, C(y)) < \pi$, so that we have the geodesic loop γ (5.3.7) from y through the nearest cut point q. Of course q is critical for the distance function from y (two segments in opposite directions). But q is so far from y that because of (5.2.5, 5.2.4) there is only one critical point around: the farthest point $\bar{q}$ from y, i.e. $q = \bar{q}$. By definition of q every geodesic from y is minimizing at least up to distance $d(y, q) = d(y, \bar{q})$, so that *all* endpoints are farthest points and by (5.2.4) equal to $\bar{q}$. This makes $q = \bar{q}$ conjugate to y, contradicting $d(y, q) < \pi$.

5.4. *Sphere Theorems.*

5.4.1. (Rauch, Berger, Klingenberg 1951–1961). *Let M^n be complete, simply connected and assume*

$$\tfrac{1}{4} < \delta \leq K \leq 1 \qquad (\text{“}\tfrac{1}{4}\text{-pinching”}).$$

Then M^n is homeomorphic to S^n. The constant $\frac{1}{4}$ cannot be improved, see (6.1).

5.4.2. (Shikata 1967). *A bi-Lipschitz bound for the homeomorphism goes to* 1 *as* $\delta \to 1$. *The homeomorphism can therefore be smoothed to a diffeomorphism if δ is close enough to* 1.

5.4.3. (Grove-Shiohama 1977). *Let M^n be complete. Assume* $0 < \delta \leq K$ *and* $\mathrm{diam}(M) > \dfrac{\pi}{2} \cdot \delta^{-1/2}$. *Then M^n is homeomorphic to S^n.*

5.4.4. REMARKS. The $\frac{1}{4}$-pinching theorem (5.4.1) is probably the most widely known comparison result. Rauch's version did not get the sharp constant $\frac{1}{4}$, but he did not use the cut locus estimates (5.3) either—instead he got such estimates as corollaries. Simple proofs of (5.4.1) depend on cut locus estimates. Given those, (5.4.3) is a nice generalization of (5.4.1): upper curvature bounds are replaced by an (implied) lower diameter bound. Note, that the diameter assumption also implies cut locus estimates (5.3.3). Exotic spheres had been discovered a few years earlier, so it was natural to try and replace "homeomorphic" by "diffeomorphic". Gromoll and Calabi proved the first diffeomorphism theorems in 1966; but, given what I have developed in this chapter, Shikata's result is most easily explained. It also points to a difference in the conclusions in (5.4.1) and (5.4.3) which the formulation "homeomorphic" deemphasizes: no metric control on the homeomorphism is obtained in (5.4.3), while (5.4.1) naturally leads to (5.4.2). Some of the arguments which were developed for the proofs of (5.4) I have put, as more general tools, into (5.2) and used them in (5.3).

Proof of (5.4.3). Choose $p, q \in M$ such that $d(p, q) = \operatorname{diam}(M)$. Because of (5.2.4, 5.2.5) we have p and q as the *only critical points* for the distance function from p. We modify the construction (5.2.2) to obtain a vector field X on M with only one radial source at q and one radial sink at p. First choose two balls $B(p), B(q)$ which do not meet the cut locus $C(p)$ respectively, $C(q)$ and define on these the local vector fields not quite as in (5.2.2): for each $y \in B(p)$ respectively, $y \in B(q)$ let $X_p(y)$ respectively, $X_q(y)$ be the unit tangent vector of the unique segment yp respectively, qy. Now cover M with the concentric balls $B'(p), B'(q)$ and the $B'(x_i)$ of (5.2.2) (for $A = M \setminus (B'(p) \cup B'(q))$. The vector field $X = (1/\Sigma\varphi_i) \cdot \Sigma\varphi_i X_i$ then agrees with X_p near p, X_q near q and has no singularities except p, q. All integral curves run from q to p and their finite arc length depends differentiably on the initial direction. Now identify T_qM isometrically with a tangent space T_NS^n and map the integral curves of X proportional to arc length onto the corresponding meridians of S^n from N. This defines a continuous bijective map $M^n \to S^n$, hence a homeomorphism. The map is of maximal rank differentiable in $M^n \setminus \{q, p\}$;

the problem at q is harmless, but at p no information can be obtained as to how the angle between integral curves in M is related to the angle between the corresponding meridians in S^n.

Proof of (5.4.1). Choose $p, q \in M$ such that $d(p, q) = \operatorname{diam}(M)$. Assuming the cut locus estimate $d(y, C(y)) \geqslant \pi$ (5.3) we conclude from (5.2.6) that *every* geodesic which starts from p respectively, q (since it is minimizing up to a distance $\geqslant \pi > (\pi/2) \cdot \delta^{-1/2}$) *reaches* the "equator" set

$$E := \left\{ x \in M; d(p, x) = d(x, q) \quad \left(\leqslant \frac{\pi}{2} \cdot \delta^{-1/2}, 5.2.6 \right) \right\}$$

before it reaches the cut locus $C(p)$ respectively, $C(q)$. (5.2.6) also shows that the gradients of $d(q, x)$ and $-d(p, x)$ make along E an angle $\alpha \leqslant \pi - \delta^{1/2} \cdot \operatorname{diam}(M) \leqslant \pi - \delta^{1/2}\pi < (\pi/2)$ (and $\alpha \to 0$ as $\delta \to 1$). The function $f(x) := d(q, x) - d(p, x)$ is therefore differentiable near E and $\operatorname{grad} f \neq 0$, i.e. E is a differentiable submanifold. In particular, the distance from p to E (or q to E) depends differentiably on the initial direction of the segment from p (or q) to E, i.e. the radial map from the unit sphere in T_pM (or T_qM) to E is a diffeomorphism. So we call E the equator of M and the segments from p or q to E half meridians. The homeomorphism $M^n \to S^n$ is now clear: Identify T_qM with a tangent space T_NS^n and map the "meridians" of M proportional to arc length onto the corresponding meridians of S^n.

About the proof of (5.4.2): The homeomorphism from M^n to S^n which we just obtained is of maximal rank differentiable on $M \setminus \{E \cup \{p, q\})$. The nonsmoothness along E is not serious; more simply than in the proof of (5.4.3) one can combine $\operatorname{grad} d_q$ and $-\operatorname{grad} d_p$ to a smooth vector field on $M \setminus \{p, q\}$ such that the integral curves are essentially the broken meridians above, only slightly changed near E to smooth the corners. Crucial for the application of Shikata's smoothing result is the control of the "antipodal map" $T_q^1(M) \to T_p^1M$ which maps the initial directions ($\in T_qM$) of the broken meridians to the final directions ($\in T_pM$). This map is defined by geodesics, its derivative can therefore be described by Jacobi fields along the broken meridians; the Jacobi fields vanish at p, q, they match at E and describe the derivative as

$J'|_q \to J'|_p$. Since Rauch's estimates (1.8) control the maximum and minimum growth of Jacobi fields with $J(0)=0$, we get the desired bi-Lipschitz bound from the ratio of the upper and lower bound in (1.8)

$$L(\delta) \leqslant \frac{s_\delta}{s_\Delta}\left(\frac{\pi}{2}\cdot\delta^{-1/2}\right) = \sin^{-1}\left(\frac{\pi}{2}\cdot\delta^{-1}\right) \xrightarrow[\delta\to 1]{} 1.$$

Shikata's smoothing construction has been simplified since. It also depends on Hessian bounds for the distance function. But it is high time to get to the examples, so I omit the smoothing arguments.

6. COMPLEX PROJECTIVE SPACE $\mathbf{C}P^n$ AND ITS DISTANCE SPHERES

The following properties make these examples relevant to the preceding comparison theorems.

6.1.1. *Complex projective space* $\mathbf{C}P^n$ *has a natural Riemannian metric (Fubini-Study) which has curvature bounds* $1 \leqslant K \leqslant 4$.

6.1.2. *Diameter = cut locus distance* $= \pi\cdot\Delta^{-1/2}$ *(compare* 2.4.2, 5.2.4, 5.3.1*).*

6.1.3. $\mathbf{C}P^n$ *is not homeomorphic to* S^{2n} *(compare* 5.3*).*

6.1.4. *The odd dimensional distance spheres of radius r have curvature bounds*

$$0 < \delta(r) := \frac{\cos^2 r}{\sin^2 r} \leqslant K \leqslant 4 + \frac{\cos^2 r}{\sin^2 r} =: \Delta(r) \xrightarrow[r\to(\pi/2)]{} 4$$

and have closed geodesics of length

$$l(r) = 2\pi\cos r\sin r \xrightarrow[r\to(\pi/2)]{} 0.$$

6.1.5. *If* $\delta(r)$: $\Delta(r) < \frac{1}{9}$ *then* $l(r) < 2\pi\cdot\Delta^{-1/2}$ *(compare* 5.3.2*) and diam(r)* $\leqslant \frac{1}{2}\pi\cdot\delta(r)^{-1/2}$ *(compare* 5.3.6*).*

6.2.1. *Definition as a metric space.* $\mathbf{C}P^n$ is defined as the set of complex lines in $\mathbf{C}^{n+1}$, or, what is the same, as the set of Hopf circles $\{e^{i\varphi}\cdot p;\ \varphi\in[0,2\pi]\}_{p\in S^{2n+1}\subset\mathbf{C}^{n+1}}$. Any two circles have, as disjoint compact sets in S^{2n+1}, a natural distance, namely the length of a shortest great circle arc which joins them; this distance is $\leq \pi/2$ and $= \pi/2$ if the two Hopf circles lie in totally orthogonal subspaces of $\mathbf{R}^{2(n+1)} = \mathbf{C}^{n+1}$.

6.2.2. *Natural embedding* into a Euclidean sphere. Denote by Sym the vector space of complex linear hermitian symmetric endomorphisms of the Euclidean vector space $\mathbf{C}^{n+1}$ with the scalar product

$$\langle\langle A, B\rangle\rangle = \operatorname{Re}\operatorname{trace}(\bar{A}^t \cdot B) = \operatorname{Re}\Big(\sum_{i,j} \bar{a}_{ij} b_{ij}\Big), \qquad A, B \in \mathrm{Sym}.$$

Define the quadratic map

$$V\colon S^{2n+1} \to \mathrm{Sym}, \qquad V(p)(z) := \frac{1}{2}(\langle z, p\rangle \cdot p + \langle z, ip\rangle \cdot ip),$$

that is, $V(p)$ is (up to the factor $\frac{1}{2}$ which is more convenient later) the orthogonal projection onto the complex line $\mathbf{C} \cdot p$, in particular $V(e^{i\varphi} \cdot p) = V(p)$. V gives, therefore, an injective map from the set of Hopf circles into Sym. Since $V(p)$ has two eigenvalues $\frac{1}{2}$ and all others 0 we have $\langle\langle V(p), V(p)\rangle\rangle = \frac{1}{2}$, so the image lies in a sphere.

6.2.3. *Submanifold metric.* The map V has constant rank $2n$ on S^{2n+1}, therefore the image is a $2n$-dimensional submanifold in Sym. We show more. Curves $p(t)$ on S^{2n+1} which are perpendicular to Hopf circles are mapped by V onto curves of the same length in Sym. With its induced submanifold metric the image is therefore isometric to $\mathbf{C}P^n$ (as defined in 6.2.1).

Proof. The assumptions are $\dot{p} \perp p, ip$, hence also $i\dot{p} \perp p, ip$. The tangent vector of $V(p(t))$ is given by

$$\frac{d}{dt} V(p(t))(z) = \frac{1}{2}(\langle z, \dot{p}\rangle p + \langle z, p\rangle \dot{p} + \langle z, i\dot{p}\rangle ip + \langle z, ip\rangle i\dot{p}),$$

hence—with $\{e_i\}$ an orthonormal basis— $\langle\langle \partial V \cdot \dot{p}, \partial V \cdot \dot{p}\rangle\rangle =$

$$\sum_{i=1}^{2n} \Big\langle \frac{d}{dt} V(p(t))(e_i), \frac{d}{dt} V(p(t))(e_i)\Big\rangle = \frac{1}{4} \cdot 4\langle \dot{p}, \dot{p}\rangle.$$

6.3.1. *Isometries.* Unitary maps $U: \mathbf{C}^{n+1} \to \mathbf{C}^{n+1}$ are isometries of S^{2n+1} which map Hopf circles to Hopf circles, therefore they give isometries of $\mathbf{C}P^n$. These extend to isometries of Sym (conjugations)

$$V(U \cdot p) = U \cdot V(p) \cdot U^{-1},$$

and $\langle\langle U \cdot A, U \cdot B \rangle\rangle = \langle\langle A, B \rangle\rangle = \langle\langle A \cdot U, B \cdot U \rangle\rangle$.

For any fixed orthonormal complex basis $\{v_0, \dots, v_n\}$ of $\mathbf{C}^{n+1}$ we have a complex conjugation $\overline{\Sigma z_i \cdot v_i} := \Sigma \bar{z}_i \cdot v_i$. These are also isometries which preserve Hopf circles: $\overline{e^{i\varphi} \cdot p} = e^{-i\varphi} \cdot \bar{p}$, hence give isometries of $\mathbf{C}P^n$, and as before extend to isometries of Sym (conjugation of the matrices).

6.3.2. *Totally geodesic submanifolds.* Connected fixed point sets of isometries of Riemannian manifolds are totally geodesic.

a) The fixed point sets in S^{2n+1} of conjugations are equatorial spheres cut out by at most $(n+1)$-$\dim_{\mathbf{R}}$ subspaces. These n-spheres are orthogonal to the Hopf circles, so V maps them, preserving length, into the image—but the map is not injective: antipodal points are on the same Hopf circle. The totally geodesic fixed point sets in $\mathbf{C}P^n$ obtained from conjugations in $\mathbf{C}^{n+1}$ are therefore real projective spaces $\mathbf{R}P^n$ of constant curvature 1.

b) To each subspace $\mathbf{C}^{k+1} \subset \mathbf{C}^{n+1}$ we define a complex reflection

$$U: \mathbf{C}^{n+1} \to \mathbf{C}^{n+1}, \qquad U|_{\mathbf{C}^{k+1}} := id, \quad U|_{(\mathbf{C}^{k+1})^{\perp}} := -id.$$

This map leaves the Hopf circles in $\mathbf{C}^{k+1}$ and $(\mathbf{C}^{k+1})^{\perp}$ fixed, and no others. As an isometry of $\mathbf{C}P^n$ (6.3.1) it has lower dimensional complex projective subspaces $\mathbf{C}P^k$, $\mathbf{C}P^{n-k-1}$ as totally geodesic fixed point sets.

6.3.3. *The Riemann spheres* $\mathbf{C}P^1$ *have curvature* 4 *in* $\mathbf{C}P^n$. Since the metric on any $\mathbf{C}P^1 \subset \mathbf{C}P^n$ can be computed by intersecting S^{2n+1} with the appropriate $\mathbf{C}^2$, it is enough to consider the case $n = 1$ (real dimension 2). The unitary maps of $\mathbf{C}^2$ are transitive on (the Hopf circles of) S^3; this and (6.2.1) give: the metric on $\mathbf{C}P^1$

has constant curvature and diameter $\pi/2$. This leaves two possibilities: S^2 with curvature 4 and $\mathbf{R}P^2$ with curvature 1; but multiplication by i gives $\mathbf{C}P^1$ an orientation—excluding $\mathbf{R}P^2$.

6.3.4. REMARK. The parametrization for $S^3 \subset \mathbf{R}^4 = \mathbf{C}^2$

$$F(\alpha, t, t_0) = \begin{pmatrix} \cos\alpha \cdot \begin{pmatrix} \cos(t+t_0) \\ \sin(t+t_0) \end{pmatrix} \\ \sin\alpha \cdot \begin{pmatrix} \cos(t-t_0) \\ \sin(t-t_0) \end{pmatrix} \end{pmatrix}, \qquad i = \left(\begin{array}{cc|cc} 0 & -1 & \multicolumn{2}{c}{0} \\ 1 & 0 & & \\ \hline \multicolumn{2}{c|}{0} & 0 & -1 \\ & & 1 & 0 \end{array}\right)$$

has the t-lines as Hopf circles since $(\partial/\partial t)F = iF$. The distance between any two Hopf circles can be obtained from the metric which measures the length of curves perpendicular to Hopf circles,

$$ds^2 = d\alpha^2 + \left(\frac{1}{2}\sin 2\alpha\right)^2 dt_0^2.$$

It is the metric of a sphere of curvature 4 in polar coordinates.

6.3.5. *Symmetric space structure.* For every $p \in S^{2n+1}$ consider

$$U\colon \mathbf{C}^{n+1} \to \mathbf{C}^{n+1}, \qquad U\Big|_{\mathrm{span}(p,ip)} = id, \qquad U\Big|_{\{p,ip\}^\perp} = -id.$$

The differential of this isometry (use the submanifold picture) on the tangent space of the fixed point $V(p)$ is $(-id)$. Riemannian manifolds M which have to every $p \in M$ an isometry σ_p: $M \to M$ such that $\sigma_p(p) = p$, $\partial\sigma_p|_p = -id$ are called *symmetric spaces.*

An immediate consequence is that the curvature tensor is parallel. Namely, the differential of an isometry at a fixed point preserves any isometry-invariant tensor; in particular, using $\partial\sigma_p|_p = -id$ we get

$$\begin{aligned} -(D_X R)(U,V)W &= (D_{-X}R)(-U,-V)-W \\ &= (D_X R)(U,V)W. \end{aligned}$$

Similarly, multiplication by i on each tangent space defines a tensor field J ($J(X) := i \cdot X$)) which is compatible with all the unitary isometries, hence it is parallel:

$$-(D_X J)(Y) = (D_{-X} J) \cdot (-Y) = (D_X J)(Y).$$

6.3.6. *$\mathbf{C}P^n$ is very different from S^{2n}.* Define the ("Kähler"-) 2-form

$$\omega(X, Y) := g(J \cdot X, Y).$$

Because of (6.3.5) it is parallel: $D\omega(X, Y) = g(DJ \cdot X, Y) = 0$, in particular, the exterior derivative vanishes: $d\omega = 0$. Also, ω is obviously the area form when restricted to any $\mathbf{C}P^1$ (if $X \in T_p\mathbf{C}P^1$, $|X| = 1$, then X and $J(X)$ are an orthonormal basis of $T_p\mathbf{C}P^1$ and $\omega(X, J(X)) = 1$), hence $\int_{\mathbf{C}P^1} \omega = (1/4) \cdot 4\pi$.

Observe that in S^{2n} ($n > 1$) we can differentiably contract any S^2 to a point and then evaluate $\int_{S^2} \omega = \int_{\text{Cone}(S^2)} d\omega$ (Stokes) for any 2-form ω. We showed that this cannot be done in $\mathbf{C}P^n$. In the appropriate language this is expressed as: The second cohomology of $\mathbf{C}P^n$ is not zero.

6.4. *The curvature tensor of* $\mathbf{C}P^n$. If X, Y, Z are vector fields tangential to a totally geodesic submanifold then $D_Y Z$ is also tangential and hence $R(X, Y)Z$ is tangential. Let Y be any tangent vector of $\mathbf{C}P^n$, then Y, iY span the tangent space of a totally geodesic $\mathbf{C}P^1$ of curvature 4, hence

$$R^{\mathbf{C}P^n}(iY, Y)Y = 4 \cdot iY.$$

If $X \perp Y, iY$, then X and Y span the tangent space of a totally geodesic $\mathbf{R}P^2$ of curvature 1, hence

$$R^{\mathbf{C}P^n}(X, Y)Y = 1 \cdot X \quad \text{if } X \perp Y, iY.$$

Combining,

$$R^{\mathbf{C}P^n}(X, Y)Y = \langle Y, Y \rangle \cdot X - \langle X, Y \rangle \cdot Y + 3\langle X, iY \rangle \cdot iY. \quad (6.4.1)$$

Again for any curvature tensor, the first Bianchi identity gives

$$6R((X,Y)Z) = R(X,Y+Z)Y+Z - R(X,Y-Z)Y-Z$$
$$+R(X+Z,Y)X+Z - R(X-Z,Y)X-Z.$$

This and (6.4.1) give the full curvature tensor.

$$R^{\mathbf{C}P^n}(X,Y)Z = \langle Y,Z\rangle\cdot X - \langle X,Z\rangle\cdot Y$$
$$2\cdot\langle X,iY\rangle\cdot iZ + \langle iY,Z\rangle\cdot iX - \langle iX,Z\rangle\cdot iY, \qquad (6.4.2)$$

but already (6.4.1) shows that the curvature range is [1, 4].

6.5. *Metric and principal curvatures of the distance spheres in* $\mathbf{C}P^n$. Every family of concentric distance spheres $S(r)$ cuts each totally geodesic $\mathbf{C}P^1$ through the midpoint of the $S(r)$ into concentric distance circles of geodesic curvature $2\operatorname{ctg} 2r$. We call those circles Hopf circles on the distance spheres since their tangent field is obtained by multiplying the radial vector field N by i. Hence

6.5.1. The Hopf circles on a distance sphere of radius r in $\mathbf{C}P^n$ have length $2\pi\cdot\cos r\cdot\sin r$. They are geodesics and principal curvature lines with principal curvature $\kappa_1(r) = 2\operatorname{ctg} 2r$.

The radial direction N and a tangent vector $\perp N, iN$ span a tangent space of a totally geodesic $\mathbf{R}P^2$ of curvature 1. Hence

6.5.2. The $\mathbf{R}P^2$'s through the midpoints of distance spheres in $\mathbf{C}P^n$ intersect those spheres perpendicular to their Hopf circles in closed geodesics of length $2\pi\cdot\sin r$ (namely distance circles in $\mathbf{R}P^2$). These are also principal curvature lines of curvature $\kappa_2(r) = \operatorname{ctg} r$.

Remark. The Hopf circles on the distance spheres in $\mathbf{C}P^n$ shrink by a factor $\cos r$ faster than great circles of distance spheres in S^{2n}. As $r\to\pi/2$ the length goes to zero and the distance spheres in $\mathbf{C}P^n$ get collapsed to $\mathbf{C}P^{n-1}$.

6.6. *The curvature tensor of the distance spheres.* Let S_r denote the shape operator and R_r the curvature tensor of a distance sphere of radius r. The Gauss equations are

$$\langle R^{\mathbf{C}P^n}(X,Y)Y,X\rangle = \langle R_r(X,Y)Y,X\rangle - \langle S_rX,X\rangle\langle S_rY,Y\rangle + \langle S_rX,Y\rangle^2.$$

This and (6.5.1, 6.5.2) give

$$K_r(iN\wedge Y) = 1 + \frac{2\cos 2r}{\sin 2r}\cdot\frac{\cos r}{\sin r} = \left(\frac{\cos r}{\sin r}\right)^2, \quad (6.6.1)$$

$$K_r(X\wedge iX) = 4 + \left(\frac{\cos r}{\sin r}\right)^2 \quad (X\perp iN), \quad (6.6.2)$$

$$K_r(X\wedge Y) = 1 + \left(\frac{\cos r}{\sin r}\right)^2 \quad (X,Y\perp iN). \quad (6.6.3)$$

So the curvatures have at least the range claimed in (6.1.4). But since we saw that the eigenspaces of S_r are compatible with the eigenspaces of $R^{\mathbf{C}P^n}(\ \ ,Y)Y$ one checks easily that indeed (6.6.1) gives min K and (6.6.2) max K.

Finally, we get

$$\text{length of Hopf circles} < 2\pi(\max K)^{-1/2}$$

$$\text{iff } \cos^2 r < (1+3\sin^2 r)^{-1} \ (\text{or } 2 < 3\sin^2 r)$$

$$\text{iff } \min K : \max K < \tfrac{1}{9}.$$

Remark. Berger (1960) discovered the curvature and cut locus properties of these metrics on S^3. Weinstein (1973) observed that they occur as distance spheres in $\mathbf{C}P^n$. They also appear on the quotient $\mathbf{R}P^3 = SO(3)$ as the kinematic metric of a rotating solid body with two equal moments of inertia and one smaller moment.

Acknowledgements. I wish to thank many colleagues, in particular Karsten Grove and Wolfgang Ziller, for many helpful discussions and suggestions. Most of the manuscript was written while I enjoyed the hospitality and ideal working conditions of the I.H.E.S. in Bures.

REFERENCES

[B] Berger, M.: La Géométrie Métrique des Variétés Riemanniennes, Soc. Math. France, Astérisque, hors série (1985), 9–66.

[BGS] Ballmann, W., Gromov, M., Schroeder V.: Manifolds of Nonpositive Curvature, Birkhäuser 1985.

[BK] Buser P., Karcher, H.: Gromov's Almost Flat Manifolds, Soc. Math. France, Astérisque 81 (1981).

[BZ] Burago, Y., Zalgaller, D.: Convex Sets in Riemannian Spaces of Nonnegative Curvature, Russian Math. Surveys, t. 32-3 (1977), 1–57.

[CE] Cheeger, J., Ebin, D. G.: Comparison Theorems in Riemannian Geometry, American Elsevier, New York 1975.

[G] Gromov, M.: Almost Flat Manifolds, J. Diff. Geom. 13 (1978), 231–241.

[S] Sakai, T.: Comparison and Finiteness Theorems in Riemannian Geometry, Advanced Studies in Pure Mathematics 3 (1984), 125–181.