J. DIFFERENTIAL GEOMETRY
16 (1981) 333–345

TOTALLY CONVEX SETS IN COMPLETE RIEMANNIAN MANIFOLDS

VICTOR BANGERT

1. Introduction

While convex sets in Riemannian manifolds have similar local properties as convex sets in Euclidean space, their behavior in the large can be very different. If one chooses an appropriate notion of convexity, however, global similarities exist and two of them will be discussed in this paper. First, we generalise the following theorem of Euclidean geometry to the Riemannian case:

Monotony Theorem. *Let $D \subseteq \mathbf{R}^{m+1}$ be a compact set containing a convex set C. Then the m-dimensional Hausdorff measures of the topological boundaries satisfy*

$$\mathcal{H}^m(\partial C) \leqslant \mathcal{H}^m(\partial D).$$

For a proof note that the metric projection P: $\mathbf{R}^{m+1} \to C$ onto C is distance-nonincreasing and maps ∂D onto ∂C.

The Riemannian version of this monotony theorem is Theorem 1 in §3. The appropriate assumption is that C is totally convex in D. For 2-dimensional Riemannian manifolds a related problem has been treated in [11, 4.14]. Theorem 1 improves this result even for dimension 2. In §4 the case of equality in Theorem 1 is investigated. Under additional curvature assumptions we obtain a splitting theorem for the Riemannian structure of $D - \mathring{C}$.

Subsequently we treat the topological implications of total convexity. In Theorem 4 we prove that the inclusion i: $C \to D$ is a homotopy equivalence provided C is totally convex in the locally convex set D. This result may be viewed as a counterpart to the fact that all convex sets in Euclidean space are contractible, hence homotopy equivalent. A corollary to Theorem 4 states that the only totally convex set in a compact connected Riemannian manifold M is M itself. This result is used in §4.

Received December 6, 1979, and, in revised form, April 10, 1981. Work partially supported by the program Sonderforschungsbereich 40 "Theoretische Mathematik" at the University of Bonn.

Finally we note that some of our results have simpler proofs if one considers sublevels of convex functions instead of totally convex sets. However, contrary to the Euclidean case, there exist totally convex sets which cannot be sublevels of convex functions; cf. [1, p. 94]. For sublevels of convex functions a slightly weaker form of Theorem 4 has recently been proved in [10].

2. Notation and definitions

Throughout this paper M will denote an $(m + 1)$-dimensional, smooth, connected manifold with complete Riemannian metric $\langle\, , \rangle$ and Levi-Cività connection ∇. We denote by π: $TM \to M$ the tangent bundle, and by σ: $T^1M \to M$ the unit tangent bundle of M. Let d: $M \times M \to \mathbf{R}$ be the metric induced by $\langle\, , \rangle$, and let $B(p, \varepsilon)$ be the closed metric ball about p of radius ε.

A function f: $M \to \mathbf{R}$ is convex if for every geodesic c in M the function $f \circ c$ is convex.

For the convexity of sets we shall use the following notions.

Let C be a nonvoid closed connected subset of M.

(i) C is strongly convex if for $p, q \in C$ the minimal geodesic segment $\overline{pq}$ is unique within M and $\overline{pq} \subseteq C$.

(ii) C is locally convex if for all $p \in C$ there exists $\varepsilon > 0$ such that $C \cap B(p, \varepsilon)$ is strongly convex.

(iii) Suppose the interior $\mathring{D}$ of $D \subseteq M$ contains C. Then C is totally convex in D if any geodesic c: $[0, 1] \to D$ which joins two points of C is contained in C. In the case $D = M$, C is simply said to be totally convex.

Obviously both (i) and (iii) imply (ii). Details concerning these definitions can be found in [2] and [7].

We are now going to describe the integration theory by means of which we calculate the boundary volumes of sets.

A subset L of M is a k-dimensional strong Lipschitz submanifold of M if locally L is the graph of a Lipschitz function defined on $\mathbf{R}^k$, i.e., for every $p \in L$ there exist a chart ϕ: $U^\phi \to \mathbf{R}^{m+1}$ of M at p and a Lipschitz function f: $U \to \mathbf{R}^{m+1-k}$ defined on an open subset U of $\mathbf{R}^k$ such that $\phi(L \cap U^\phi) = \{(x, f(x)) | x \in U\}$; cf. [16]. Then α: $U \to L$, $\alpha(x) = \phi^{-1}(x, f(x))$ is called a coordinate system of L at p. The notions "measurable" and "set of measure zero" make sense on L since they are invariant under locally bilipschitz homeomorphisms. L is said to be differentiable at $p \in L$ if one (and hence every one) coordinate system α of L at p is differentiable in $\alpha^{-1}(p)$. Let L' denote the set of points at which L is differentiable. By Rademacher's theorem (cf. [8]) $L - L'$ is a set of measure zero. For $p \in L'$ the tangent space T_pL of L at p is defined as the linear span of the vectors $\{\partial\alpha/\partial x^1|\alpha^{-1}(p), \cdots, \partial\alpha/\partial x^k|\alpha^{-1}(p)\}$. The k-dimensional volume element of M is the symmetric form $\mathrm{dvol}_k^M =: \mathrm{dvol}_k$ defined by

$$\mathrm{dvol}_k(p): (T_pM)^k \to \mathbf{R},$$

$$\mathrm{dvol}_k(p)(a_1, \cdots, a_k) = (\det\langle a_i, a_j\rangle)^{1/2}.$$

For a bounded measurable function $h: L \to \mathbf{R}$ with compact support in the range $\alpha(U)$ of a coordinate system α of L we define

$$\text{(2.1)} \qquad \int_L h \,\mathrm{dvol}_k := \int_U h \circ \alpha \,\mathrm{dvol}_k\left(\frac{\partial\alpha}{\partial x^1}, \cdots, \frac{\partial\alpha}{\partial x^k}\right) dx^1 \cdots dx^k.$$

Since the formula for changes of variables is valid for locally bilipschitz homeomorphisms, (2.1) is independent of the choice of α. Thus using an appropriate subdivision of L into measurable subsets we can define a measure vol_k on L in such a way that locally the integral with respect to vol_k is given by (2.1). On the set of k-dimensional strong Lipschitz submanifolds, vol_k coincides with the k-dimensional Hausdorff measure $\mathcal{H}^k$ induced by d; cf. [8, 3.2.5 and 3.2.46]. However, we will not use this fact, but work with our analytical definition instead.

Finally, let $H \subseteq M$, $L \subseteq N$ be strong Lipschitz submanifolds of Riemannian manifolds M, N, respectively and suppose $\dim H = \dim L = k$. If f: $H \to L$ is locally Lipschitz and $K \subseteq H$ is measurable, we have the following formula:

$$\text{(2.2)} \qquad \int_K |\det f_*| \mathrm{dvol}_k^M \geqslant \mathrm{vol}_k^N(f(K)),$$

where the equality holds if $f|K$ is injective. This follows from [8, 3.2.5] after everything is reduced to locally Lipschitz functions from R^k to $\mathbf{R}^k$, since for differentiable f at p, $|\det f_*|(p)$ is computed with respect to the scalar products induced on T_pH and $T_{f(p)}L$.

The admissible sets in our monotony theorem will have the property defined below. A closed nonvoid subset D of M is said to have strong Lipschitz boundary, if $D = \overline{\mathring{D}}$ and $\partial D = D - \mathring{D}$ is a strong Lipschitz submanifold of M. If $\partial D \neq \varnothing$, then ∂D separates M; hence $\dim \partial D = m$. By [16, Theorem 6.1], every locally convex set with nonvoid interior has a strong Lipschitz boundary. Actually the boundaries of locally convex sets satisfy even stronger regularity conditions; cf. [2].

The nonstandard integration theory described above is used since it enables us to calculate the boundary volumes of convex sets by analytical means.

Adding smoothness assumptions to our theorems would not simplify the proofs essentially except that one could use the standard integration techniques in this case. Note however that the problem of approximating an arbitrary convex set by convex sets with smooth boundary is still open.

3. The monotony theorem

In this section we prove Theorem 1, the monotony theorem for Riemannian manifolds. Formulas (3.1)–(3.3) below play a crucial role in our proof; they are closely related to Santalo's formula (cf. [12, p. 488]). Since (3.1)–(3.3) are known in the smooth case (cf. [5, §VIII, 8]), we only give the arguments which are necessary to generalize (3.1)–(3.3) to the Lipschitz case.

We first note some generalities on the Riemannian structure of T^1M. On TM we will use the canonical metric defined by requiring that for all $v \in TM$

$$\pi_* \oplus K: T_v TM \to T_{\pi(v)}M \oplus T_{\pi(v)}M$$

be isometric, where $T_{\pi(v)}M \oplus T_{\pi(v)}M$ denotes the orthogonal sum, and K is the connection map of the Levi-Civitá connection. T^1M will be endowed with the metric induced by the inclusion $T^1M \to TM$. Then $\sigma\colon T^1M \to M$ is a Riemannian submersion. Considering M as the 0-section of TM we can use the volume elements $\mathrm{dvol}_k := \mathrm{dvol}_k^{TM}$ on both T^1M and M. The volume of the standard k-sphere is denoted by $\mathrm{vol}(S^k)$. According to Liouville's theorem the geodesic flow

$$\Phi\colon T^1M \times \mathbf{R} \to T^1M$$

preserves dvol_{2m+1}.

Throughout this section, A denotes a closed nonvoid subset of M with strong Lipschitz boundary B, and $\mathcal{A} := \sigma^{-1}(A)$, $\mathfrak{B} := \sigma^{-1}(B)$. The inner unit normal (vector field) $N\colon B' \to \mathfrak{B}$ is measurable and defined at points of differentiability of B. We set

$$\mathfrak{B}^+ := \{v \mid v \in \sigma^{-1}(B') \text{ and } \langle v, N \circ \sigma(v)\rangle > 0\}.$$

Now suppose $G \subseteq B$ is measurable, and $\mathcal{G}^+ := \sigma^{-1}(G) \cap \mathfrak{B}^+$. Obviously $\mathfrak{B} \times \mathbf{R}$ is a $(2m+1)$-dimensional strong Lipschitz submanifold of $T^1M \times \mathbf{R}$. Endowing $T^1M \times \mathbf{R}$ with the Riemannian product structure we can apply (2.2) with $K := \mathcal{G}^+ \times [0, s]$ and $F := \Phi|\mathfrak{B} \times \mathbf{R}$, $F\colon \mathfrak{B} \times \mathbf{R} \to T^1M$. Then

$$\mathrm{vol}_{2m+1}\big(\Phi(\mathcal{G}^+ \times [0, s])\big) \leqslant \int_{\mathcal{G}^+ \times [0, s]} |\det F_*| \mathrm{dvol}_{2m+1}^{\mathfrak{B} \times \mathbf{R}}, \tag{3.1}$$

where the equality holds if $\Phi|\mathcal{G}^+ \times [0, s]$ is one-to-one. For $v \in \mathfrak{B}'$ and $t \in [0, s]$, we can compute $|\det F_*|(v, t)$ as in the smooth case. Using the fact that Φ preserves dvol_{2m+1}, we get

$$|\det F_*|(v, t) = \langle N \circ \sigma(v), v\rangle.$$

Since the right-hand side is independent of t we easily conclude from (2.1)

$$\int_{\mathcal{G}^+ \times [0, s]} |\det F_*| \mathrm{dvol}_{2m+1}^{\mathfrak{B}\times\mathbf{R}} = s \int_{\mathcal{G}^+} \langle N \circ \sigma(v), v\rangle \mathrm{dvol}_{2m}. \tag{3.2}$$

At all points of differentiability of $\mathfrak{B}$, the map $\sigma|\mathfrak{B}: \mathfrak{B} \to B$ has the defining property of a Riemannian submersion. Using (2.2) and the same arguments as in the smooth case this implies that for every nonnegative measurable function f: $\mathfrak{B} \to \mathbf{R}$

$$\int_{\mathfrak{B}} f \,\mathrm{dvol}_{2m} = \int_B \left(\int_{\sigma^{-1}(p)} f \,\mathrm{dvol}_m \right) \mathrm{dvol}_m.$$

Since $\sigma^{-1}(p) \subseteq T^1M$ is isometric to the standard sphere, we can compute

$$\int_{\sigma^{-1}(p) \cap \mathfrak{B}^+} \langle N \circ \sigma(v), v\rangle \mathrm{dvol}_m = \frac{1}{m} \mathrm{vol}(S^{m-1})$$

Hence

$$\int_{\mathcal{G}^+} \langle N \circ \sigma(v), v\rangle \mathrm{dvol}_{2m} = \frac{1}{m} \mathrm{vol}(S^{m-1}) \mathrm{vol}_m(G). \tag{3.3}$$

We note that combining (3.1)–(3.3) proves Santalo's formula in the Lipschitz case.

Santalo's formula.

$$\mathrm{vol}_{2m+1}(\Phi(\mathcal{G}^+ \times [0, s])) \leqslant \frac{1}{m} \mathrm{vol}(S^{m-1}) \mathrm{vol}_m(G) s,$$

where the equality holds if $\Phi|\mathcal{G}^+ \times [0, s]$ is one-to-one.

A geodesic c: $[a, b] \to A$ with $c(a) \in \partial A$, $c(b) \in \partial A$ and $c((a, b)) \subseteq \mathring{A}$ will be called a chord of A (of length $L(c)$). Theorem 1 will be obtained as a corollary to the following more quantitative result.

Proposition 1. *Let $A \subseteq M$ have strong Lipschitz boundary $B = \partial A$ and finite volume $\mathrm{vol}_{m+1}(A)$. Suppose G is a measurable subset of B with $\mathrm{vol}_m(G) > \mathrm{vol}_m(B - G)$. Then the infimum t_0 of the lengths of chords of A with both end-points on G satisfies*

$$t_0 \leqslant k_m \frac{\mathrm{vol}_{m+1}(A)}{\mathrm{vol}_m(G) - \mathrm{vol}_m(B - G)} \quad \textit{with } k_m = m \frac{\mathrm{vol}(S^m)}{\mathrm{vol}(S^{m-1})}.$$

Remark. In particular we prove the existence of such chords with the inclusion of the cases $\mathrm{vol}_m(B) = \infty$ or $G = B$.

Proof. For $v \in \mathfrak{B}^+$ we define $s(v) := \sup\{s | \Phi_t v \in \mathring{\mathfrak{A}}$ for $t \in (0, s)\}$. Then $s(v) > 0$ since B is differentiable in $\sigma(v)$. If c: $[a, b] \to A$ is a chord of

A, and $c(a) \in G$, $c(b) \in B - G$, then $\dot{c}(b) \in \sigma^{-1}(B - G) - \mathcal{B}^+ =: (\mathcal{B} - \mathcal{G})^-$. Obviously Santalo's formula holds for $(\mathcal{B} - \mathcal{G})^-$ as well. Hence

$$\text{(3.4)} \quad \mathrm{vol}_{2m+1}\big(\Phi((\mathcal{B} - \mathcal{G})^- \times [0, s])\big) \leqslant \frac{1}{m}\mathrm{vol}(S^{m-1})\mathrm{vol}_m(B - G)s.$$

Now let us assume that for some fixed $t > 0$ we cannot find a chord of A with both end-points on G and length $\leqslant t$. This implies

$$\text{(3.5)} \qquad \text{If } v \in \mathcal{G}^+ \text{ and } s(v) \leqslant t, \text{ then } \Phi_{s(v)}v \in (\mathcal{B} - \mathcal{G})^-.$$

Our proof consists in finding a lower bound on the volume of

$$\mathcal{G}_t^+ := \{\Phi_s v | v \in \mathcal{G}^+, 0 \leqslant s \leqslant \min\{t, s(v)\}\}.$$

which in turn yields an upper bound on t. (3.5) implies that for all $\varepsilon \in (0, t)$

$$\Phi_\varepsilon(\mathcal{G}_{t-\varepsilon}^+) - \Phi\big((\mathcal{B} - \mathcal{G})^- \times [0, \varepsilon]\big) \subseteq \mathcal{G}_t^+ - \mathcal{G}_\varepsilon^+.$$

Using (3.4) we conclude

$$\text{(3.6)} \quad \begin{aligned} &\mathrm{vol}_{2m+1}(\mathcal{G}_t^+) - \mathrm{vol}_{2m+1}(\mathcal{G}_{t-\varepsilon}^+) \\ &\quad \geqslant \mathrm{vol}_{2m+1}(\mathcal{G}_\varepsilon^+) - \frac{1}{m}\mathrm{vol}(S^{m-1})\mathrm{vol}(B - G)\varepsilon. \end{aligned}$$

Setting $\mathcal{D}_\varepsilon := \{v \in \mathcal{G}^+ | s(v) > \varepsilon\}$ we have $\bigcup_{\varepsilon>0} \mathcal{D}_\varepsilon = \mathcal{G}^+$, $\Phi(\mathcal{D}_\varepsilon \times [0, \varepsilon]) \subseteq \mathcal{G}_\varepsilon^+$, and $\Phi|\mathcal{D}_\varepsilon \times [0, \varepsilon]$ is one-to-one. From (3.1) and (3.2) it follows

$$\Phi(\mathcal{D}_\varepsilon \times [0, \varepsilon]) = \varepsilon \int_{\mathcal{D}_\varepsilon} \langle N \circ \sigma(v), v \rangle d\mathrm{vol}_{2m},$$

where

$$\lim_{\varepsilon \to 0} \int_{\mathcal{D}_\varepsilon} \langle N \circ \sigma(v), v \rangle d\mathrm{vol}_{2m} = \int_{\mathcal{G}^+} \langle N \circ \sigma(v), v \rangle d\mathrm{vol}_{2m}.$$

This implies, in consequence of (3.3) and (3.6), that for the function $f(s) := \mathrm{vol}_{2m+1}(\mathcal{G}_s^+)$

$$\liminf_{\varepsilon \downarrow 0} \frac{1}{\varepsilon}(f(t) - f(t - \varepsilon)) \geqslant \frac{1}{m}\mathrm{vol}(S^{m-1})(\mathrm{vol}_m(G) - \mathrm{vol}_m(B - G)).$$

Since our assumption on t is a fortiori satisfied for all $s \in (0, t)$, the preceding formula holds with t replaced by s for all $s \in (0, t)$. Hence $f(0) = 0$ implies

$$f(t) \geqslant \frac{1}{m}\mathrm{vol}(S^{m-1})(\mathrm{vol}_m(G) - \mathrm{vol}_m(B - G))t.$$

By definition we have $\mathcal{G}_t^+ \subseteq \mathcal{Q}$, hence

$$f(t) \leqslant \mathrm{vol}_{2m+1}(\mathcal{Q}) = \mathrm{vol}(S^m)\mathrm{vol}_{m+1}(A).$$

The above two inequalities prove our claim.

Remark. The estimate in Proposition 1 is sharp for A = a standard hemisphere and $G = B = \partial A$, but not in general. It can be sharpened, for example, by introducing the infimum of the lengths of chords of A joining G and $B - G$ as an additional parameter.

Theorem 1. *Let $C \subseteq M$ be totally convex in $D \subseteq M$. Suppose $\mathring{C} \neq \emptyset$, $\mathrm{vol}_{m+1}(D - \mathring{C}) < \infty$ and D has strong Lipschitz boundary. Then* $\mathrm{vol}_m(\partial C) \leqslant \mathrm{vol}_m(\partial D)$.

Proof. By assumption we have $C \subseteq \mathring{D}$. Hence $A := D - \mathring{C}$ has strong Lipschitz boundary $B = \partial C \cup \partial D$; cf. [16, Theorem 6.1]. Since C is totally convex in D, there does not exist a chord of A with both end-points on ∂C. Hence the assumption of Proposition 1 cannot be satisfied for $G := \partial C$, i.e.,

$$\mathrm{vol}_m(\partial C) \leqslant \mathrm{vol}_m(B - \partial C) = \mathrm{vol}_m(\partial D).$$

Remark. If $\mathring{C} = \emptyset$, then $C = \partial C$, and one can even prove $2\,\mathrm{vol}_m(\partial C) \leqslant \mathrm{vol}_m(\partial D)$.

4. The splitting theorems

In this section we investigate the case of equality $\mathrm{vol}_m(\partial C) = \mathrm{vol}_m(\partial D)$ in Theorem 1. Examples on surfaces of revolution show that in this case the Riemannian structure of $D - \mathring{C}$ can still vary quite freely. Under suitable assumptions, however, every component of $D - \mathring{C}$ splits isometrically into a product of a real interval with a component of ∂C.

Proposition 2. *Let $C \subseteq M$ be totally convex in $D \subseteq M$. Suppose $\mathring{C} \neq \emptyset$, $D - \mathring{C}$ is compact, and* $\mathrm{vol}_m(\partial C) = \mathrm{vol}_m(\partial D) > 0$.

(i) *If D has a strong Lipschitz boundary, then D is locally concave, i.e., $M - \mathring{D}$ is locally convex.*

(ii) *If D is locally convex, then ∂D is a totally geodesic hypersurface.*

Remark. In case (ii) the set $M - \mathring{D}$ is even totally convex in $M - \mathring{C}$. It is rather doubtful that this should be true in the general case as well.

Proof. Let $E \subseteq T^1M$ denote the set of vectors $v \in \sigma^{-1}(\partial D)$ such that there exists $t_v > 0$ with $\exp(tv) \in \mathring{D} - C$ for all $t \in (0, t_v)$. Choose $t_v \in (0, \infty]$ to be maximal with this property. Proposition 1 and $\mathrm{vol}_m(\partial C) = \mathrm{vol}_m(\partial D)$ imply $t_v < \infty$ and $\exp(t_v v) \in \partial C$ for almost all $v \in E$. Now suppose (i) is not true. Then there exist points $p_0, q_0 \in M - \mathring{D}$ at arbitrarily small distance, in particular $d(p_0, q_0) < d(\partial C, \partial D)$, such that the geodesic segment $\overline{p_0q_0}$ intersects $\mathring{D}$. Hence there exist open sets $U, V \subseteq M - \mathring{D}$ such that for all $p \in U$, $q \in V$ the geodesic segment $\overline{pq}$ intersects $\mathring{D}$ while $\overline{pq} \cap C = \emptyset$. Let $\mathcal{O} \subseteq T^1M$ be the set of tangent vectors to such segments. Then $\mathcal{O}$ is open, $\mathcal{O} \cap E \neq \emptyset$, and $\exp(t_v v) \in \partial D$ for all $v \in \mathcal{O} \cap E$. Since for every $p \in (\partial D)'$ the set $E \cap T_p^1M$ contains an open hemisphere of T_p^1M, we get $\mathrm{vol}_{2m}(\mathcal{O} \cap E) > 0$ contradicting $\exp(t_v v) \in \partial C$ for almost all $v \in E$. Thus (i) is proved. (ii) is an obvious consequence of (i).

Theorem 2. *Let $C \subseteq M$ be totally convex in the locally convex set $D \subseteq M$. Suppose $\mathring{C} \neq \varnothing$, $D - \mathring{C}$ is compact, and $\mathrm{vol}_m(\partial C) = \mathrm{vol}_m(\partial D) > 0$. If the sectional curvature of M is nonnegative on $D - \mathring{C}$, then the components of $D - \mathring{C}$ are isometric to the products of real intervals and the boundary components of C.*

Remarks. 1. Examples on a paraboloid of revolution show that it does not suffice to assume that C and D are locally convex and $C \subseteq \mathring{D}$.

2. The author does not know if the assumption "D locally convex" is necessary. It can be omitted if, in addition to the other assumptions, $D - \mathring{C}$ is contained in the domain of a convex function which is nowhere constant on $D - \mathring{C}$, e.g., if M is noncompact and of nonnegative sectional curvature.

Proof. Denote by $\rho: D \to \mathbf{R}$, $\rho(p) := d(p, \partial D)$ the inner distance function from ∂D. Since the curvature is nonnegative on $D - \mathring{C}$, the function $\rho | D - \mathring{C}$ is concave, i.e., $-\rho | D - \mathring{C}$ is convex; cf. [7]. Set $a := \min(\rho | C)$. Then $a > 0$ since $C \subseteq \mathring{D}$. For $0 < r \leqslant a$ the sets ${}^rD := \rho^{-1}([r, \infty))$ are totally convex in D and $\partial({}^rD) = \rho^{-1}(r)$. Hence $\mathrm{vol}_m(\rho^{-1}(r)) \leqslant \mathrm{vol}_m(\partial D)$ by Theorem 1. Since C is totally convex in ${}^rD \subseteq D$ for $0 \leqslant r < a$, we obtain $\mathrm{vol}_m(\partial C) \leqslant \mathrm{vol}_m(\rho^{-1}(r))$. Hence

$$\mathrm{vol}_m(\rho^{-1}(r)) = \mathrm{vol}_m(\partial D)$$

for all $0 < r < a$. By Proposition 2 the hypersurface ∂D and its parallel hypersurfaces $\rho^{-1}(r)$ $(0 < r < a)$ are all totally geodesic. Hence the flow of $-\nabla\rho$ induces an isometry between $\rho^{-1}([0, a])$ and $\rho^{-1}(a) \times [0, a]$.

Now provided $D - \mathring{C}$ is connected we prove $\partial C = \rho^{-1}(a)$, hence $D - \mathring{C} = \rho^{-1}([0, a])$. Since $\rho^{-1}(a)$ is totally geodesic and $C \subseteq {}^aD$ is totally convex in D, the set $C \cap \rho^{-1}(a) = \partial C \cap \rho^{-1}(a)$ is totally convex in the compact Riemannian manifold $\rho^{-1}(a)$. By Corollary 1 in §5 the set $\partial C \cap \rho^{-1}(a)$ contains a component N of $\rho^{-1}(a)$. Then $N \times [0, a] \subseteq \rho^{-1}(a) \times [0, a]$ corresponds to an open and closed set in $D - \mathring{C}$. Hence $N = \rho^{-1}(a) = \partial C$ and $D - \mathring{C} \simeq \partial C \times [0, a]$. If $D - \mathring{C}$ is not connected, one can apply these arguments to each component of $D - \mathring{C}$ separately.

We are now going to prove a splitting theorem in the case of nonpositive sectional curvature. At no extra cost we get a slight generalization of Theorem 1. For a path-connected subset A of M we denote by $\pi_1(A, p)$ the fundamental group of A at $p \in A$, and by $i_{A*}(\pi_1(A, p))$ its image under the inclusion $i_A: A \to M$.

Theorem 3. *Let M be a complete Riemannian manifold of nonpositive curvature. Suppose that D is a connected subset of M with strong Lipschitz* *boundary containing a locally convex set C, and that* $D - \mathring{C}$ *is compact. If*

$$(*) \qquad i_{C*}(\pi_1(C, p)) = i_{D*}(\pi_1(D, p)),$$

then $\mathrm{vol}_m(\partial C) \leqslant \mathrm{vol}_m(\partial D)$. *If in addition* $C \subseteq \mathring{D}$ *and* $\mathrm{vol}_m(\partial C) = \mathrm{vol}_m(\partial D) > 0$, *then the components of* $D - \mathring{C}$ *are isometric to the products of real intervals and the boundary components of C.*

Remark. If C is locally convex and $C \subseteq \mathring{D}$, then $(*)$ implies that C is totally convex in D. The converse is true, if D is locally convex.

Proof. We use the universal Riemannian covering $f\colon M' \to M$ to construct a retraction $P\colon D \to C$ which is locally distance-nonincreasing. Choose connected components C' and D' of $f^{-1}(C)$ and $f^{-1}(D)$ such that $C' \subseteq D'$. Then C' is totally convex; cf. [2, (2.12)]. By [6, Proposition (3.4)] the metric projection $P'\colon M' \to C'$ onto C' is uniquely defined and distance-nonincreasing. Because of $(*)$ a map $P\colon D \to C$ can be defined by $P \circ (f|D') := f \circ (P'|D')$. Then P is locally distance-nonincreasing. Hence $\mathrm{vol}_m(\partial C) \leqslant \mathrm{vol}_m(\partial D)$ follows from (2.2) as soon as we have proved $P(\partial D) = \partial C$. As a consequence of $(*)$ the distance function $\rho'\colon M' \to \mathbf{R}$ from C' induces a function $\rho\colon D \to \mathbf{R}$ by $\rho \circ (f|D') = \rho'|D'$. Since $D - \mathring{C}$ is compact, ρ is bounded. Now for every $p \in \partial C'$ there exists a geodesic $c\colon [0, \infty) \to M'$ such that $c(0) = p$, $P' \circ c(t) = p$ and $\rho' \circ c(t) = t$ for all $t > 0$; cf. [6, Lemma (3.2)]. Since $\rho'|D'$ is bounded, we get $P'(\partial D') = \partial C'$, hence $P(\partial D) = \partial C$. If $\mathrm{vol}_m(\partial C) = \mathrm{vol}_m(\partial D)$, we conclude that $(P|\partial D)_{*p}$ is isometric for almost all $p \in \partial D$. Now ρ is C^1 on $D - C$, and $P_*(\nabla\rho) = 0$ whenever P_* is defined. This implies that $\nabla\rho|_p$ is orthogonal to $T_p\partial D$ if $(P|\partial D)_{*p}$ is isometric. Hence the derivative of the Lipschitz function $\rho|\partial D$ vanishes almost everywhere. Thus ρ is constant on connected components of ∂D. Since ρ' is convex, we conclude that D is locally convex. Then C is totally convex in D provided $C \subseteq \mathring{D}$. Since $P_* \mid T_p\partial D$ is isometric for almost all $p \in \partial D$ and since $\mathrm{vol}_m(\partial C) = \mathrm{vol}_m(\partial D) > 0$, we have $\mathring{C} \neq \varnothing$. By Proposition 2 the hypersurface ∂D is totally geodesic. Since the components of ∂C are parallel surfaces of components of ∂D, our claim easily follows from the curvature assumption.

5. Topological properties of totally convex sets

In [7] Cheeger and Gromoll proved that the inclusion of a compact totally convex submanifold into a complete connected Riemannian manifold is a homotopy equivalence. By the "Soul Theorem" [7] this statement is also true for compact totally convex sets in a complete Riemannian manifold of nonnegative curvature. We are going to generalize these results in several respects.

Theorem 4. *Let C, D be subsets of a complete Riemannian manifold M. If C is totally convex in D, and D is locally convex, then the inclusion i: $C \to D$ is a homotopy equivalence.*

Corollary 1. *The only totally convex set in a compact connected Riemannian manifold M is M itself.*

The main step in the proof of Theorem 4 is the following.

Lemma 1. *Under the assumptions of Theorem 4 any Lipschitz map g: $(D^n, \partial D^n) \to (D, C)$ is (D, C)-homotopic to a map into C.*

Proof. Since D is connected by assumption, the statement is true for $n = 0$. For $n \geqslant 1$ we consider g as a map from D^{n-1} into an appropriate space of curves. Applying a special energy-decreasing deformation $\mathfrak{D}$ to g we can deform g to a map into C since our assumptions imply that $\mathfrak{D}$ has no stationary points outside C. Similar deformations have been used on the space of closed curves to construct closed geodesics; cf. the appendix to [12]. Here we consider the space

$$\Omega = \{\gamma | \gamma: [-1, 1] \to D \text{ Lipschitz}, \gamma(-1) \in g(\partial D^n), \gamma(1) \in C\}$$

endowed with the metric $d_\infty(\gamma_0, \gamma_1) := \max d(\gamma_0(t), \gamma_1(t))$. On Ω we have the lower semicontinuous functional "energy" E: $\Omega \to \mathbf{R}$, $E(\gamma) := \int_{-1}^{1} |\dot\gamma|^2(t)\, dt$. For $\kappa \geqslant 0$ we set $\Omega^\kappa = E^{-1}([0, \kappa]) \subseteq \Omega$. The Lipschitz map g: $(D^n, \partial D^n) \to (D, C)$ gives rise to a continuous map G: $(D^{n-1}, \partial D^{n-1}) \to (\Omega, \Omega^0)$ which maps $x \in D^{n-1}$ to the curve $G(x) \in \Omega$, $G(x)(t) := g(x, t\sqrt{1 - |x|^2}\,)$. A (Ω, Ω^0)-homotopy of G induces a (D, C)-homotopy of g. Hence it suffices to prove that G is (Ω, Ω^0)-homotopic to a map into Ω^0. We are now going to define the deformation $\mathfrak{D}$. If L is a Lipschitz constant for g, choose $\kappa_0 > 2L^2$. Then $G(x) \in \Omega^{\kappa_0}$ for all $x \in D^{n-1}$. Choose a neighborhood U of C such that the metric projection P: $U \to C$ onto C is uniquely defined and locally Lipschitz; cf. [15, Theorem 1]. Set

$$K = \{p \in M | \text{There exists } x \in D^n \text{ such that } d(p, g(x)) \leqslant \sqrt{2\kappa_0}\,\},$$

and choose $\varepsilon > 0$ with the following properties:

(i) ε is smaller than the minimum of the injectivity radius on K,

(ii) if $p, q \in D \cap K$ and $d(p, q) < \varepsilon$, then the shortest geodesic segment $\overline{pq}$ from p to q is contained in D; cf. [2, (1.2)],

(iii) $\{q \in K | d(q, C) < \varepsilon\} \subseteq U$.

Finally we choose a partition $-1 = s_0 < s_1 < \cdots < s_{2k-1} < s_{2k} = 1$ of $[-1, 1]$ such that $s_{i+2} - s_i < \varepsilon^2/\kappa_0$ for $i = 0, \cdots, 2k - 2$. Note that $\gamma \in \Omega^{\kappa_0}$ implies $\gamma([-1, 1]) \subseteq K$ and $d^2(\gamma(s_i), \gamma(s_{i+2})) \leqslant (s_{i+2} - s_i)\kappa_0 < \varepsilon^2$. Hence we can define continuous energy-decreasing maps $\mathfrak{D}_1: \Omega^{\kappa_0} \to \Omega^{\kappa_0}$ and $\mathfrak{D}_2: \Omega^{\kappa_0} \to \Omega^{\kappa_0}$ respectively by

$$\begin{aligned}
\mathfrak{D}_1\gamma|[s_{2i}, s_{2i+2}] &= \text{shortest geodesic from } \gamma(s_{2i}) \\
&\quad \text{to } \gamma(s_{2i+2}), \quad 0 \leqslant i < k; \\
\mathfrak{D}_2\gamma|[s_0, s_1] &= \gamma|[s_0, s_1], \\
\mathfrak{D}_2\gamma|[s_{2i-1}, s_{2i+1}] &= \text{shortest geodesic from } \gamma(s_{2i-1}) \\
&\quad \text{to } \gamma(s_{2i+1}), \quad 1 \leqslant i < k, \\
\mathfrak{D}_2\gamma|[s_{2k-1}, s_{2k}] &= \text{shortest geodesic from } \gamma(s_{2k-1}) \\
&\quad \text{to } P(\gamma(s_{2k})).
\end{aligned}$$

Now $\mathfrak{D}: \Omega^{\kappa_0} \to \Omega^{\kappa_0}$ is defined by $\mathfrak{D} := \mathfrak{D}_2 \circ \mathfrak{D}_1$. Obviously $\mathfrak{D}_1$, $\mathfrak{D}_2$ and $\mathfrak{D}$ are $(\Omega^{\kappa_0}, \Omega^0)$-homotopic to the identity. Hence $\mathfrak{D} \circ G$ is $(\Omega^{\kappa_0}, \Omega^0)$-homotopic to G. Our claim is an easy consequence of the following property of $\mathfrak{D}$:

(*) For every $\kappa \in (0, \kappa_0)$ there exists $\delta > 0$ such that $\mathfrak{D}(\Omega^{\kappa+\delta}) \subseteq \Omega^{\kappa-\delta}$.

Suppose (*) is not true. Then we can find a sequence c_n in Ω such that $\lim E(c_n) = \lim E(\mathfrak{D}c_n) = \kappa$. Since K is compact, we may assume that $\lim c_n(s_{2i}) =: c(s_{2i})$ exists for $0 \leqslant i \leqslant k$. We define $c \in \Omega$ by

$$\begin{aligned}
c|[s_{2i}, s_{2i+2}] &= \text{shortest geodesic from } c(s_{2i}) \\
&\quad \text{to } c(s_{2i+2}), \quad 0 \leqslant i < k.
\end{aligned}$$

Then we have $\mathfrak{D}_1 c = c$ and $c = \lim \mathfrak{D}_1 c_n$. Now

$$E(\gamma) = \sum_{i=0}^{2k-1} \frac{d^2(\gamma(s_{i+1}), \gamma(s_i))}{s_{i+1} - s_i},$$

if $\gamma|[s_i, s_{i+1}]$ is a shortest geodesic for $0 \leqslant i < 2k$. Hence $E(c) = \lim E(\mathfrak{D}_1 c_n) = \kappa$ and $E(\mathfrak{D}_2 c) = \lim E(\mathfrak{D}_2(\mathfrak{D}_1 c_n)) = \kappa$. Thus we get $E(c) = E(\mathfrak{D}c) = \kappa > 0$, which implies that $c \in \Omega$ is a nonconstant geodesic with $c(1) = P(c(s_{2k-1}))$, in contradiction to the total convexity of C in D. Hence (*) is proved.

Proof of Theorem 4. Theorem 4 follows from Lemma 1 by algebraic topology. It suffices to prove that $\pi_n(D, C)$ is trivial for all $n \geqslant 0$; cf. [14, p. 405, Corollary 24]. Hence we need to know that Lemma 1 is true for not only Lipschitz maps but also all continuous maps. Using the fact that there are neighborhoods U of C and V of D and locally Lipschitz deformation retractions of U (resp. V) onto C (resp. D), one can reduce the continuous case to the Lipschitz case by standard approximation arguments.

The following corollary replaces [4, Theorem 4] which has been referred to in [3].

Corollary 2. *In addition to the hypotheses of Theorem* 4 *assume that* $D - \mathring{C}$ *is connected. Then* $D - \mathring{C}$ *is compact if* ∂D *is nonvoid and compact.*

Remark. If $D - \mathring{C}$ is not connected, Corollary 2 applies to each component V of $D - \mathring{C}$ separately, i.e., V is compact if $V \cap \partial D$ is nonvoid and compact.

Proof. Theorem 4 implies Corollary 2 by algebraic topology. We first show that the inclusion $H_m(\partial C) \to H_m(D - \mathring{C})$ is an isomorphism. Let $U \subseteq D$ be a neighborhood of C such that the metric projection P: $U \to C$ is uniquely defined and such that for every $q \in U$ the geodesic segment $\overline{qP(q)}$ is contained in U; cf. [15, Theorem 1]. Using a partition of unity one can construct a continuous function f: $\partial C \to (0, \infty)$ such that the set $\{p \in M | d(p, C) < 2f(p)\}$ is contained in U. Set $V := \{p \in M | d(p, C) < f(p)\}$. Then (D, C) (resp. $(D - \mathring{C}, \partial C)$) are deformation retracts of (D, V) (resp. $(D - \mathring{C}, V - \mathring{C})$). Hence $\mathring{C}$ can be excised from (D, C), and Theorem 4 implies that $H_m(\partial C) \to H_m(D - \mathring{C})$ is an isomorphism. We will use homology with $\mathbf{Z}_2$-coefficients. Then $H_m(\partial D) \neq 0$ since ∂D is a compact m-dimensional manifold. Hence the inclusion

$$H_m(\partial C) \oplus H_m(\partial D) \to H_m(D - \mathring{C})$$

has nontrivial kernel. By the long exact sequence of the pair $(D - \mathring{C}, \partial C \cup \partial D)$, this implies $H_{m+1}(D - \mathring{C}, \partial C \cup \partial D) \neq 0$. But for a connected $(m + 1)$-dimensional manifold $X = D - \mathring{C}$ with boundary $\partial X = \partial C \cup \partial D$, we have $H_{m+1}(X, \partial X) \neq 0$ only if X is compact. This can be seen by applying [9, Corollary 22.25] to the double of X.

References

[1] V. Bangert, *Konvexität in riemannschen Mannigfaltigkeiten*, Thesis, Dortmund, 1977.

[2] ______, *Konvexe Mengen in Riemannschen Mannigfaltigkeiten*, Math. Z. **162** (1978) 263–286.

[3] ______, *Riemannsche Mannigfaltigkeiten mit nicht-konstanter konvexer Funktion*, Arch. Math. **31** (1978) 163–170.

[4] ______, *On the monotony of the boundary volumes of geodesically convex sets*, preprint.

[5] M. Berger, *Lectures on geodesics in Riemannian geometry*, Tata Institute of Fundamental Research, Bombay, 1963.

[6] R. L. Bishop & B. O'Neill, *Manifolds of negative curvature*, Trans. Amer. Math. Soc. **145** (1969) 1–49.

[7] J. Cheeger & D. Gromoll, *On the structure of complete manifolds of nonnegative curvature*, Ann. of Math. **96** (1972) 413–443.

[8] H. Federer, *Geometric measure theory*, Grundlehren Math. Wiss. 153, Springer, Berlin-Heiderberg-New York, 1969.

[9] M. J. Greenberg, *Lectures on algebraic topology*, Benjamin, New York-Amsterdam, 1967.

[10] R. E. Greene & K. Shiohama, *Convex functions on complete noncompact manifolds: Topological structure*, Invent. Math. **63** (1981) 129–157.

[11] H. Karcher, *Schnittort und konvexe Mengen in vollständigen Riemannschen Mannigfaltigkeiten*, Math. Ann. **177** (1968) 105–121.

[12] W. Klingenberg, *Lectures on closed geodesics*, Grundlehren Math. Wiss. 230, Springer, Berlin-Heidelberg-New York, 1978.

[13] L. A. Santalo, *Integral geometry in general spaces*, Proc. Intern. Congress Math. (Cambridge, 1950), Amer. Math. Soc., Vol. I, 1952, 483–489.

[14] E. H. Spanier, *Algebraic topology*, McGraw-Hill, New York 1966.

[15] R. Walter, *On the metric projection onto convex sets in Riemannian spaces*, Arch. Math. **25** (1974) 91–98.

[16] ______, *Some analytical properties of geodesically convex sets*, Abh. Math. Sem. Univ. Hamburg **45** (1976) 263–282.

UNIVERSITY OF FREIBURG, FREIBURG