J. DIFFERENTIAL GEOMETRY
**32** (1990) 429-455

# SPACE OF SOULS IN A COMPLETE OPEN MANIFOLD OF NONNEGATIVE CURVATURE

JIN-WHAN YIM

## 0. Introduction

Let $M$ be a complete open Riemannian manifold of nonnegative curvature. The most significant result in the study of the differential structure of this type of manifold is due to Cheeger and Gromoll. In [3] they produced a totally geodesic submanifold $S_0$, a *soul* of $M$, and showed that $M$ is diffeomorphic to the normal bundle $\nu(S_0)$ of $S_0$. Following this work, Sharafutdinov and, independently, Croke and Schroeder showed that there exists a *strong deformation retraction* $f\colon M \to S_0$ which is distance nonincreasing [4, 8]. Using this retraction one can show that if a soul is not unique, then they are all isometric and homologous to each other. Moreover, there are infinitely many isometric copies of a soul in $M$, which are not necessarily souls. This observation leads us to the following definition.

**Definition.** A subset $S \subset M$ is called a *pseudo-soul* if it is homologous and isometric to a soul $S_0$ with respect to the induced metric.

In particular, it is clear that all souls are pseudo-souls, and the definition is independent of a soul $S_0$. If a soul is not unique, then there are infinitely many pseudo-souls. The purpose of this paper is to investigate the union $\mathscr{H}$ of all pseudo-souls in $M$. In fact, we will prove the following theorem.

**Theorem.** *$\mathscr{H} \subset M$ is a totally geodesic embedded submanifold which is isometric to a product manifold $S_0 \times N$, where $N$ is a complete manifold of nonnegative curvature diffeomorphic to a Euclidean $k$-space $\mathbf{R}^k$ and $k$ is the dimension of the space of all parallel normal vector fields along the soul $S_0$. Furthermore any pseudo-soul in $M$ is of the form $S_0 \times \{p\}$ for some $p \in N$.*

As an immediate corollary of this theorem, if the normal bundle itself is parallel, we obtain the splitting $M = S_0 \times N$. This special case has been independently studied in [6].

There are two trivial examples of $M$ for which one can easily find pseudo-souls and the space $\mathscr{H}$. If $M$ is a paraboloid, then every point

Received April 8, 1988 and, in revised form February 10, 1989.

$p \in M$ is a pseudo-soul, and hence $\mathscr{H} = M$. The other case is a flat cylinder $S^1 \times \mathbf{R}$, in which $S^1 \times \{t\}$, $t \in \mathbf{R}$, is a soul (and hence a pseudo-soul) and $\mathscr{H} = M$. Until recently, partially due to insufficient examples, it was suspected that if $\mathscr{H}$ is not trivial (i.e., $\mathscr{H} \neq S_0$ or $M$), then $M$ should be a product $M_1 \times N$, where $M_1$ has a unique soul $S_0$ and $\mathscr{H} = S_0 \times N$. However, as pointed out by M. Strake, there does exist an example which is not a product but has nontrivial $\mathscr{H}$. It is still unanswered what is the best metric structure of $M$ one can expect when $\mathscr{H}$ is not trivial.

The definition of a pseudo-soul was first introduced by C. Croke and V. Schroeder by whom the statement in the theorem was conjectured and has been studied. Some of the techniques used in this paper are due to them.

The author would like to express his thanks to Professor Croke for his assistance in the preparation of this paper.

## 1. Preliminaries

The proof of our main theorem is rather technical and requires a knowledge of the geometry of convex sets, which may not be familiar to some readers. In the present section, for this reason, we provide a brief outline of the proof we will establish, and recall some notation and results from [3], [8].

As might be expected from the statement of the theorem, a parallel section in the normal bundle $\nu(S_0)$ of a soul is a key tool in the construction of the space $\mathscr{H}$. In fact, it was shown in [8] that the *exponential image* $\exp_{S_0} F$ of any parallel normal vector field $F$ along $S_0$ is a pseudo-soul. We aim to prove that in this fashion one can produce all the pseudo-souls in $M$, and then use a local argument to accomplish the final goal.

By the construction of the Sharafutdinov retraction $f$ (Theorem A.1), there exists a homotopy $H: M \times [0, 1] \to M$ such that $H(\cdot, 0) = \text{id}$ on $M$ and $H(\cdot, 1) = f$. One can further show that for any pseudo-soul $S \subset M$, $\{H(S, t)\}_{t \in [0, 1]}$ is a family of pseudo-souls continuously parametrized by $t \in [0, 1]$ such that $H(S, 0) = S$ and $H(S, 1) = S_0$ is a soul. We first replace this continuous family by a broken geodesic's worth of pseudo-souls, and then use a *curve shortening process* (Lemma 3.2) to prove that the connection between $S_0$ and $S$ can be done by a family $\{\gamma_p\}_{p \in S_0}$ of geodesics emanating from $S_0$ such that $\gamma_p(t) = \exp_p F(p)$, where $F$ is a parallel normal vector field along $S_0$.

Although the homotopy $H$ is only continuous, it can be shown in Proposition A.8 that for each $p \in M$ the curve $t \to \varphi_p(t) = H(p, t)$, $t \in [0, \varepsilon)$, has a *right tangent vector* $\nabla\psi(p) = \varphi_p'(0+)$. Therefore, $\nabla\psi$ is by definition a (not necessarily smooth) vector field on $M$. Along any pseudo-soul $S$, since $H(S, t)$, $t \in [0, \varepsilon)$, is an isometric variation through pseudo-souls, the variational vector field $\nabla\psi$ is a *global normal Jacobi field*, which we will show to be parallel. In §2, we will prove one of the most important properties of a pseudo-soul. Along a pseudo-soul the *mixed curvatures* vanish (Corollary 2.7), which implies that a global normal Jacobi field along a pseudo-soul is in fact parallel (Corollary 2.5). For any fixed pseudo-soul $S$, once we have this parallel normal vector field along each pseudo-soul $S_t = H(S, t)$, $t \in [0, 1]$, we can locally approximate the continuous 1-parameter family of pseudo-souls by the exponential images $exp_{S_t} s\nabla\psi$, $0 \leq s < \varepsilon$, which are pseudo-souls as well. We thus obtain a connection between $S_0$ and $S$ by a parallel family of broken geodesics (a (P)-connection), which will then be followed by curve shortening.

As we have to apply the curve shortening process to a family of broken geodesics, we need a local product structure in a neighborhood of every pseudo-soul. It is shown in §2 as another application of the vanishing mixed curvatures along pseudo-souls. Let $F$ be a parallel normal vector field along a pseudo-soul $S$, and let $S_1 = \exp_S F$ be a pseudo-soul connected to $S$ by $F$. If $F_1$ is another parallel normal vector field along $S$, then $\exp_S(F + sF_1)$, $0 \leq s < \varepsilon$, is an isometric variation of $S_1$, and hence there exists a corresponding global Jacobi field along $S_1$, which is again parallel. Consequently one can see that the dimensions of the spaces of all parallel normal vector fields along $S$ and $S_1$ are the same (Corollary 2.7), which implies the dimension is constant in $\mathscr{H}$ because every pseudo-soul is connected to a fixed soul by a family of broken geodesics generated by parallel vector fields as above. This number will be of course the dimension $k$ of the submanifold $N$ in the main theorem, and the tangent space of $N$ is the set of vectors which can be extended to parallel normal vector fields along pseudo-souls. It also proves that for any pseudo-soul $S \subset \mathscr{H}$, the orthogonal decomposition $T_pS \oplus T_pN$ into the tangent space of $S$ and its orthogonal complement in $\mathscr{H}$ is invariant under parallel translation, which will give us a local product structure (Proposition 2.8).

In the remainder of this section, we will recall the construction of a soul $S_0$ [3], and formally introduce the concept of a pseudo-soul.

**Definition 1.1.** A nonempty subset $C$ of $M$ will be called *totally convex* if for any $p, q \in C$ and any geodesic $\gamma: [0, 1] \to M$ from $p$ to $q$, we have $\gamma[0, 1] \subset C$.

For any compact subset $D$ of $M$ let $K$ be the supremum of sectional curvatures at points of $D$ and let $R$ denote the infimum of injectivity radii of points in $D$. Let $\varepsilon_D > 0$ be a number such that $\varepsilon_D < \frac{1}{2}\min\{\pi/\sqrt{K}, R\}$. Then, by [2, Theorem 5.14, Lemma 5.15], for all $x \in D$ and $r \in (0, \varepsilon_D]$ the metric ball $B_r(x)$ is strongly convex, i.e., for any $p, q \in \overline{B}_r(x)$ there is a unique minimal geodesic $\sigma_{pq}$ between $p$ and $q$ such that the interior of $\sigma_{pq}$ is contained in $B_r(x)$. Moreover, for any geodesic segment $\tau\colon [0, 1] \to B_r(x)$, $d(\tau(s), p)$ has at most one critical point, and such a critical point must be a minimum.

This number $\varepsilon_D > 0$ has been used for the construction of a soul $S_0$ and the deformation retraction $f$, and will be used again throughout this paper.

In [3, Proposition 1.3], it was shown that for any $p \in M$ there exists a family of compact t.c.s. (totally convex sets) $C_t$, $t \geq 0$, such that

(1) $t_2 \geq t_1$ implies $C_{t_2} \supset C_{t_1}$, and

$$C_{t_1} = \{q \in C_{t_2} | (q, \partial C_{t_2}) \geq t_2 - t_1\};$$

in particular, $\partial C_{t_1} = \{q \in C_{t_2} | d(q, \partial C_{t_2}) = t_2 - t_1\}$.

(2) $\bigcup_{t \geq 0} C_t = M$.

(3) $p \in \partial C_0$.

Put $C = C_0$, and let $a_0 = \sup\{d(q, \partial C) | q \in C\}$. Then $C^{a_0} = \{q \in C | d(q, \partial C) = a_0\}$ is *totally convex* and $\dim C^{a_0} < \dim C$ [3, Theorem 1.9]. This contraction can be iterated until we obtain a totally convex set without boundary, and therefore we may construct a flag of t.c.s.

$$C_0 = C(0) \supset C(1) \subset \cdots \supset C(k) = S_0,$$

where $C(i+1) = C(i)^{a_i}$, and $a_i = \sup\{d(q, \partial C(i)) | q \in C(i)\}$. Thus we have:

**Theorem** [3, *Theorem* 1.11]. *$M$ contains a compact totally geodesic submanifold $S_0$ (a soul) without boundary which is totally convex, $0 \leq \dim S_0 < \dim M$.*

Note that the basic construction of a soul may depend on the starting point $p \in \partial C_0$ in [3, Proposition 1.3], which means a soul may not be unique. In any case, if $S_0$ is a soul of $M$, we can construct a deformation retraction $f\colon M \to S_0$.

**Theorem** [8, *Theorem* 2.3]. *For any soul $S_0$ of $M$ there exists a homotopy $H\colon M \times [0, 1] \to M$ such that $H(\cdot, 0) = \mathrm{id}$ on $M$ and $H(\cdot, 1) = f$, where $f\colon M \to S_0$ is a strong deformation retraction. Further, for each $t \in [0, 1]$, $H(\cdot, t)$ is distance nonincreasing.*

If a soul is not unique, we may construct a deformation retraction for each soul in $M$. Let $S_0$, $S_1$ be two different souls, and let $(f_0, H_0)$, $(f_1, H_1)$ be the corresponding retractions. Then it was shown in [8] that

$f_0(S_1) = S_0$ and $f_1(S_0) = S_1$, and for each $t \in [0, 1]$, $H_0(S_1, t)$ and $H_1(S_0, t)$ are isometric and homologous to $S_0$ (and also $S_1$). Therefore, we have the following definition.

**Definition 1.2.** A subset $S \subset M$ is called a *pseudo-soul* if it is homologous and isometric to a soul with respect to the induced metric.

With this definition we further have:

**Proposition 1.3** [8, *Proposition* 3.1]. *Let $f\colon M \to S_0$ be given as above.*

(1) *For any pseudo-soul $S$ we have $f(S) = S_0$, and for any $t \in [0, 1]$, $H(S, t)$ is a pseudo-soul; hence there is a continuous 1-parameter family of pseudo-souls between $S$ and the soul $S_0$.*

(2) *For any pseudo-soul $S$ and $p, q \in S$, the distance between $p$ and $q$ in $S$ is the same as in $M$. In particular $S$ is totally geodesic.*

**Corollary 1.4.** *If $S_1$, $S_2$ are two pseudo-souls such that $S_1 \cap S_2 \neq \emptyset$, then we have $S_1 = S_2$.*

*Proof.* Suppose $S_1 \neq S_2$ and $q \in S_1 \cap S_2$. Since $S_1$, $S_2$ are complete totally geodesic submanifolds, we have $T_qS_1 \neq T_qS_2$ (otherwise, $S_1 = S_2$). However, $f\colon M \to S_0$ is a distance nonincreasing retraction, and hence $f\colon S_i \to S_0$, $i = 1, 2$, is an isometry. Thus $df\colon T_qS_i \to T_{f(q)}S_0$ is a linear isometry. Let $v_1 \in T_qS_1$, $v_2 \in T_qS_2$ be two vectors such that $v_1 \neq v_2$ and $df(v_1) = df(v_2) = v \in T_{f(q)}S_0$. Let $\gamma_1, \gamma_2$ be the geodesics such that $\gamma_i(0) = q$ and $\gamma_i'(0) = v_i$, $i = 1, 2$. Then it is obvious that $f(\gamma_i(t)) = \gamma(t)$ for the geodesic $\gamma$ in $S_0$ with $\gamma_0 = f(q)$ and $\gamma'(0) = v$. Since $v_1 \neq v_2$, it follows that $\angle(v_1, -v_2) \neq \pi$, and hence for any $\tau > 0$ we have $d(\gamma_1(\tau), \gamma_2(-\tau)) < 2\tau$. On the other hand, in $S_0$, we have for small $\tau > 0$

$$d(f(\gamma_1(\tau)), f(\gamma_2(-\tau)) = d(\gamma(\tau), \gamma(-\tau)) = 2\tau,$$

which is a contradiction since $f$ is distance nonincreasing.

## 2. Properties of pseudo-souls

One of the most important properties of a soul $S_0$ is that the mixed curvature terms vanish along $S_0$, i.e., the sectional curvature $K(u, v) = 0$ for any tangent vector $u$ of $S_0$ and any normal vector $v$ of $S_0$ [3, Theorem 3.1]. In this section we will show that this property also holds for pseudo-souls, and see what this implies about the metric structure of pseudo-souls. We first require the following lemma which can be proved by a standard comparison theorem [8].

**Lemma 2.1.** *Let $M$ be a complete Riemannian manifold with sectional curvature $K_M$ bounded above by $K \geq 0$, and let $\gamma\colon [0, 1] \to M$ be a* *geodesic. If* $c\colon [0, 1] \to M$ *is a piecewise smooth curve from* $\gamma(0)$ *to* $\gamma(1)$ *such that* $d(c(t), \gamma[0, 1]) < r$, $r \in (0, \varepsilon_{\gamma[0,1]}]$, *then the lengths of* $c$ *and* $\gamma$ *satisfy*

$$L[c] \geq (\cos\sqrt{K}r)L[\gamma].$$

*Proof.* By [2, Lemma 5.15], for each $t \in [0, 1]$, there exists a unique number $s_t \in [0, 1]$ such that $d(c(t), \gamma[0, 1]) = d(c(t), \gamma(s_t))$. Then it is easy to see that $s(t) = s_t$ is a piecewise smooth function of $t$ and we can apply [8, Corollary 2.2] to obtain the above inequality.

**Lemma 2.2.** *Let* $M$ *be a Riemannian manifold with nonnegative curvature, and let* $p, q \in M$ *be such that* $d(p, q) = d$ *for a fixed number* $d > 0$. *If* $\gamma_0$ *is a minimal geodesic from* $p$ *to* $q$, *then for any geodesic* $\gamma$ *with* $\gamma(0) = p$, $\gamma'(0) \perp \gamma_0'(0)$, *and* $\|\gamma'(0)\| \leq D$, *there are positive numbers* $A$ *and* $s_0$, *which depend only on* $D$, *such that* $d(q, \gamma(s)) < d + As^2$ *for all* $s \in [0, s_0]$.

*Proof.* For each $s \geq 0$ let $\gamma_s$ be a minimal connection from $\gamma(s)$ to $q$. Consider the geodesic triangle $(\gamma, \gamma_0, \gamma_s)$ with the angle $\angle(\gamma'(0), \gamma_0'(0)) = \pi/2$. By Toponogov's theorem, the length $L[\gamma_s]$ is not larger than the corresponding length of the Euclidean triangle. Therefore it follows that

$$d(q, \gamma(s)) = L[\gamma_s] \leq \sqrt{L[\gamma_0]^2 + L[\gamma|_{[0,s]}]^2} \leq \sqrt{d^2 + D^2s^2}.$$

Hence we can find $A$ and $s_0$ depending only on $D$ such that $d(q, \gamma(s)) < d + As^2$ for all $s \in [0, s_0]$.

**Theorem 2.3.** *Let* $M$ *be a complete open manifold of nonnegative curvature, and let* $S_0$ *and* $S$ *be a soul and a pseudo-soul of* $M$, *respectively. If* $\gamma\colon [0, 1] \to S$ *is a geodesic in* $S$ *and* $V$ *is a piecewise smooth vector field along* $\gamma$ *such that* $V(t)$ *is perpendicular to* $S$ *and vanishes at the end points, then the index form* $I(V, V)$ *is nonnegative.*

*Proof.* Suppose there exists a vector field $V$ along $\gamma$ such that $I(V, V) < 0$. Then clearly $V$ is not identically zero since otherwise $I(V, V) = 0$. Let $c\colon [-\delta, \delta] \times [0, 1] \to M$ be the variation of the geodesic $\gamma$ such that for each $t \in [0, 1]$ the curve $\sigma_t$ defined by $\sigma_t(s) = c(s, t)$ is a geodesic with $\sigma_t'(0) = V(t)$. For each $s \in [-\delta, \delta]$, let $c_s\colon [0, 1] \to M$ be the curve defined by $c_s(t) = c(s, t)$. Then $c_0(t) = \gamma(t)$ and $c_s$ is a piecewise smooth curve from $\gamma(0)$ to $\gamma(1)$. Since the index form $I(V, V)$ is negative and $V(t)$ is perpendicular to $\gamma$ with $V(0) = V(1) = 0$, it follows from the first and second variational formulas that

$$\left.\frac{d}{ds}\right|_{s=0} L[c_s] = 0, \qquad \left.\frac{d^2}{ds^2}\right|_{s=0} L[c_s] < 0,$$

where $L[c_s]$ is the arclength of $c_s$. Hence there are two positive numbers $B$ and $s_1$ such that $L[c_s] \leq L[\gamma] - Bs^2$ for all $s \in [0, s_1]$.

We now consider the deformation retraction $f: M \to S_0$ and the image $f(c_s)$ of the variation under the retraction $f$. Since $f|s: S \to S_0$ is an isometry, the curve $f \circ \gamma: [0, 1] \to S_0$ is a geodesic in $S_0$. Denote $\tilde{\gamma} = f \circ \gamma$, $\tilde{c} = f \circ c$, and $\tilde{c}_s = f \circ c_s$. Then $\tilde{c}_s: [-\delta, \delta] \times [0, 1] \to S_0$ is a continuous variation of the geodesic $\tilde{\gamma}$. We first claim that there are positive numbers $A$ and $s_0$ such that, for each $s \in [0, s_0]$, $\tilde{c}_s$ is contained in $B_\rho(\tilde{\gamma})$, where $B_\rho(\tilde{\gamma})$ is a $\rho$-tubular neighborhood of $\tilde{\gamma}$ and $\rho(s) = As^2$. One can verify this claim using the fact that $f: M \to S_0$ is distance nonincreasing. Put $D = \sup \|V(t)\|$, and let $d > 0$ be such that $d < \frac{1}{2}\min\{\varepsilon_S, \varepsilon_{S_0}\}$. It follows from Proposition 1.3(2) that for any $p, q \in S$ with $d(p, q) = d$ the minimal geodesic from $p$ to $q$ is contained in $S$. By Lemma 2.2 we can find positive numbers $A$ and $s_0$ such that for any $t \in [0, 1]$ and $q \in S$ with $d(q, \gamma(t)) = d$ we have $d(q, \sigma_t(s)) < d + As^2$ for all $s \in [0, s_0]$. In particular, for the fixed number $d$, we choose $s_0 > 0$ so that $Ds_0$ (and hence $As_0^2$) is smaller than $d$.

We are now ready to prove our claim. Suppose that $d(\tilde{c}_s(t_1), \tilde{\gamma}[0, 1]) \geq \rho = As^2$ for some $t_1 \in [0, 1]$. Then we have $d(\tilde{c}_s(t_1), \tilde{\gamma}(t_1)) = r_1 \geq \rho$. If $\gamma_1: [0, r_1] \to S_0$ is the minimal geodesic from $\tilde{c}_s(t_1)$ to $\tilde{\gamma}(t_1)$, we then extend $\gamma_1$ on $S_0$ and let $\tilde{q} \in S_0$ be such that $\tilde{q} = \gamma_1(r_1 + d)$. Since $r_1 \leq d(c_s(t_1), \gamma(t_1)) \leq Ds < d$, we see that $\gamma_1[0, r_1 + d]$ is contained in the strongly convex set $B_{2d}(\gamma_1(r_1))$, and it follows that $d(\tilde{q}, \tilde{c}_s(t_1)) = r_1 + d$. If $q = f^{-1}(\tilde{q})$ is a preimage of $\tilde{q}$ in $S$, then $d(q, \gamma(t_1)) = d$ and $d(q, c_s(t_1)) = d(q, \sigma_{t_1}(s)) < d + \rho$. Hence,

$$d(\tilde{q}, \tilde{c}_s(t_1)) = d(f(q), f \circ c_s(t_1)) \leq d(q, c_s(t_1)) < d + \rho,$$

which is a contradiction since $d(\tilde{c}_s(t_1), \tilde{\gamma}(t_1)) = r_1 \geq \rho$.

For each $s \in [0, s_0]$, we choose a partition $0 = t_0 < t_1 < \cdots < t_m = 1$ such that the broken geodesic $\tilde{\gamma}_s: [0, 1] \to S_0$, which is defined to minimize the distance between $\tilde{c}_s(t_{i-1})$ and $\tilde{c}_s(t_i)$ for each $i$ is contained in $B_{2\rho}(\tilde{\gamma})$. By construction, $f$ is distance nonincreasing, and therefore it follows that $L[\tilde{\gamma}_s] \leq L[c_s]$ for each $s \in [0, s_0]$. Set $K = \sup\{K_{S_0}\}$. Since $2As_0^2 < 2d < \varepsilon_{S_0}$, by Lemma 2.1 we obtain for all $s \in [0, s_0]$:

$$\begin{aligned} L[\tilde{\gamma}_s] &\geq (\cos 2\sqrt{K}\rho)L[\tilde{\gamma}] \\ &= (\cos 2\sqrt{K}As^2)L[\gamma] \\ &\geq (1 - 2KA^2s^4)L[\gamma], \end{aligned}$$

where the last inequality is from $\cos x \geq 1 - \frac{1}{2}x^2$. Hence, for each $s \leq \min\{s_0, s_1\}$, we have

$$L[\gamma] - 2KA^2L[\gamma]s^4 \leq L[\tilde{\gamma}_s] \leq L[c_s] \leq L[\gamma] - Bs^2.$$

The last inequality implies $Bs^2 \leq 2KA^2L[\gamma]s^4$, or

$$1 \leq 2KA^2B^{-1}L[\gamma]s^2 = Cs^2 \quad \text{for } C > 0,$$

which is a contradiction.

**Corollary 2.4** [3, *Lemma* 3.3]. *With $M$ and $S$ as above, all sectional curvatures vanish for planes spanned by a tangent vector of $S$ and a normal vector of $S$ in $M$. Equivalently,*

$$R(u, v)v = R(v, u)u = 0,$$

*where $u$ is any tangent vector of $S$, and $v$ is any normal vector of $S$ in $M$.*

With this corollary, we obtain the following, as was shown in the proof of [5, Theorem 2].

**Corollary 2.5.** *Let $M$ and $S$ be as above. If $J: S \to TM$ is a global normal Jacobi field, i.e., if $J$ is a Jacobi field along any geodesic in $S$, then $J$ is a parallel vector field along $S$.*

*Proof.* Suppose $\nabla_u J \neq 0$ for $u \in T_pS$. Consider the geodesic $\gamma: (-\infty, \infty) \to S$ with $\gamma(t) = \exp_p(tu)$, and the Jacobi field $J$ along $\gamma$. From the Jacobi equation, $J'' + R(J, \dot{\gamma})\dot{\gamma} = 0$, and Corollary 2.4, it follows that

$$\begin{aligned}\langle J, J\rangle'' &= 2\langle J', J'\rangle + 2\langle J'', J\rangle \\ &= 2\langle J', J'\rangle - 2\langle R(J, \dot{\gamma})\dot{\gamma}, J\rangle = 2\|J'\|^2 \geq 0,\end{aligned}$$

so that $\|J(t)\|^2$ is a convex function from $\mathbf{R}$ to $\mathbf{R}$, and

$$\left.\frac{d^2}{dt^2}\right|_{t=0} \|J(t)\|^2 = 2\|\nabla_u J\|^2 > 0.$$

Therefore the function $\|J(t)\|^2$ is unbounded, which is a contradiction since $S$ is compact.

Let $S$ be a pseudo-soul and let $F$ be a unit parallel normal vector field along $S$. In [8, Proposition 3.6], it was shown that for each $t \in \mathbf{R}$ the function $\phi_t: S \to M$, defined by $\phi_t(p) = \exp_p tF(p)$ ($F(p) \in T_pM$ will denote the restriction of $F$ at $p$), is an isometric embedding, and the union $\bigcup_{t\in\mathbf{R}} \phi_t(S) \subset M$ is an immersed totally geodesic submanifold which is isometric to $S \times \mathbf{R}$.

For any pseudo-soul $S$ let $\mathscr{P}(S) \subset \Gamma(\nu(S))$ denote the subspace of all parallel sections of $\nu(S)$, and let $\Phi(S) \subset \nu(S)$ be the subbundle with the fiber $\Phi_p(S) = \{F(p) \in T_pM | F \in \mathscr{P}(S)\}$ at $p \in S$. With this observation we prove the following application of Corollary 2.5.

**Proposition 2.6.** *Let $S$ and $S_1$ be pseudo-souls such that $S_1 = \exp_S F$ for some $F \in \mathscr{P}(S)$. If $d\exp_{p_0}|_{F(p_0)}$ is nonsingular for some $p_0 \in S$, then $d\exp$ preserves parallel sections, i.e.,*

$$d\exp_p(T_{F(p)}\Phi_p(S)) = \Phi_q(S_1),$$

*where $q = \exp_p F(p)$.*

*Proof.* It suffices to show that for any $p \in S$ the linear map $d\exp_p$: $tT_{F(p)}\Phi_p(S) \to T_qM$ is an injection with its image in $\Phi_q(S_1)$ since we can then interchange the roles of $S$ and $S_1$ to prove that it is an isomorphism.

Let $\{F_1, F_2, \cdots, F_m\}$ be an orthonormal basis of $\mathscr{P}(S)$ with $F_m = F/\|F\|$, i.e., for each $i = 1, \cdots, m-1$, $F_i$ is a unit parallel normal vector field along $S$ such that $F_i(p) \perp F(p)$. For each $i$ consider the family of vector fields $c_i(s) \in \mathscr{P}(S)$, $s \in (-\varepsilon, \varepsilon)$, such that $c_i(s) = F + sF_i$. Since $c_i(0) = F$ and $c_i'(0) = F_i$ (when $T_{F(p)}\Phi_p(S)$ is identified with $T_0\Phi_p(S)$ for each $p \in S$), we have to show that the set $\{J_i = d\exp|_F(F_i)\}$ is linearly independent and contained in $\mathscr{P}(S_1)$. However, the linear map $d\exp$: $T_F(T_{p_0}M) \to T_{q_0}M$ is nonsingular, and hence the set of vectors $\{J_i(q_0)\}$ is linearly independent in $T_{q_0}M$. Thus $\{J_i\}$ is linearly independent as vector fields on $S_1$. Moreover, for each $i$ and $s \in (-\varepsilon, \varepsilon)$, $c_i(s)$ is a parallel normal vector field along $S$, and hence $\exp_S c_i(s)$ is a smooth isometric variation of $S_1$ through pseudo-souls. Therefore, the variational vector field $J_i = \frac{\partial}{\partial s}|_{s=0} \exp_S c_i(s)$ is a global Jacobi field along the pseudo-soul $S_1$. Since every normal Jacobi field along a pseudo-soul is parallel by Corollary 2.4, it is sufficient to show that $J_i$ is perpendicular to $S_1$. In our case, since it is obvious that $J_m$ is perpendicular to $S_1$ and $J_i$, $i \le m-1$, we will only consider $\{J_i\}_{i \le m-1}$.

For each $p \in S$ define a geodesic $\gamma_p(t) = \exp_p tF(p)$, and for each $i \le m-1$ let $J_{i,p}$ be the normal Jacobi field along $\gamma_p$ defined by

$$J_{i,p}(t) = \frac{\partial}{\partial s}\bigg|_{s=0} \exp_p tc_i(s).$$

For any $v \in T_pS$ let $v$ be extended to the Jacobi field $J(t)$, $t \in [0, 1]$, along $\gamma_p$ such that $J(0) = v$ and $J'(0) = 0$. Then by the product and totally geodesic structure $\bigcup \phi_t(S) = S \times \mathbf{R}$, we know that $J'(t) = 0$ for all $t \in [0, 1]$. We now consider the two Jacobi fields $J$ and $J_{i,p}$ along $\gamma_p$.

Since $\langle J', J_{i,p}\rangle - \langle J, J'_{i,p}\rangle$ is constant along $\gamma_p$, and $\langle J'(0), J_{i,p}(0)\rangle = \langle J(0), J'_{i,p}(0)\rangle = 0$ when $t = 0$, we have $\langle J'(t), J_{i,p}(t)\rangle = \langle J(t), J'_{i,p}(t)\rangle$ for any $t \in [0, 1]$. Thus

$$\langle J, J_{i,p}\rangle' = \langle J', J_{i,p}\rangle + \langle J, J'_{i,p}\rangle = 2\langle J', J_{i,p}\rangle = 0.$$

In particular, when $t = 1$, $J_i(q) = J_{i,p}(1)$ is perpendicular to $J(1) \in T_qS_1$, where $q = \exp_p F(p) \in S_1$. This is true for an arbitrary vector $v \in T_pS$, and hence $J_i$ is perpendicular to $S_1$, which implies that $J_i \in \mathscr{P}(S_1)$ for each $i$.

**Corollary 2.7.** *If $S$ and $S_1$ are two pseudo-souls such that $S_1 = \exp_S F$ for some parallel normal vector field $F \in \mathscr{P}(S)$, then* $\dim(\mathscr{P}(S)) = \dim(\mathscr{P}(S_1))$.

*Proof.* With the same notation as above, if there is a pair $(p, q)$ such that $q \in S_1$ is not conjugate to $p \in S$ along $\gamma_p$, then the proposition obviously implies the corollary. If not, choose a number $t_0 \in (0, 1)$ such that $\gamma_p(t_0)$ is not conjugate to both $p$ and $q$. Put $S_{t_0} = \exp_S t_0 F$. Then by the proposition we have $\dim(\mathscr{P}(S_1)) = \dim(\mathscr{P}(S_{t_0})) = \dim(\mathscr{P}(S))$.

For any pseudo-soul $S$ and any $p \in S$, denote by $N_r(p)$ the set $\{\exp_p F(p) | F \in \mathscr{P}(S), \|F\| < r\}$. If $r > 0$ is small enough (e.g., smaller than InjRad$(p)$), then the map $h_p\colon S \times N_r(p) \to M$, $h_p(q, \exp_p F(p)) = \exp_q F(q)$, is well defined. We will denote its image by $\mathscr{H}_r(S) \subset M$. In fact, we can prove the following proposition.

**Proposition 2.8.** *If $r < \varepsilon_S$, then for any fixed $p_0 \in S$, $h_{p_0}\colon S \times N_r(p_0) \to M$ is a totally geodesic isometric embedding onto its image $\mathscr{H}_r(S)$. In particular, for any $p \in S$, $h_{p_0}(p, N_r(p_0)) = N_r(p) \subset M$ is a totally geodesic embedded submanifold of $M$.*

*Proof.* Let $\Phi^r(S)$ be the subset $\{F(p) \in TM | p \in S, F \in \mathscr{P}(S), \|F\| < r\} \subset \nu(S)$. By the definition of the number $\varepsilon_S$, the map $\exp\colon \Phi^r(S) \to M$ is an immersion. Furthermore, since $r$ is smaller than the convexity radius, we see that it is injective. In fact, if $\exp_p F_1(p) = \exp_q F_2(q) = x$, then the minimal connection $\sigma_{pq}$ between $p$ and $q$ is contained in the strongly convex ball $B_r(x)$. By Proposition 1.3(2), we have $\sigma_{pq} \subset S$, and hence $F_1(p)$ and $F_2(q)$ are both perpendicular to $\sigma_{pq}$, which is impossible in $B_r(x)$. Therefore it easily follows that $h$ $(= h_p)$ is an embedding.

We first claim that for any $p \in S$ the subset $N_r(p) \subset M$ is totally geodesic and the map $h|_{(p, N_r(p_0))}\colon \{p\} \times N_r(p_0) \to N_r(p)$, which was shown to be a diffeomorphism, is an isometry. First of all, we note that by Proposition 2.6 for any $(p, x) \in S \times N_r(p_0)$ such that $x = \exp_p F(p_0)$,

$F \in \mathscr{P}(S)$, and $y = \exp_p F(p) \in N_r(p)$, we have

$$T_y N_r(p) = dh|_{(p,x)} T_x N_r(p_0) = d\exp_p(T_{F(p)}\Phi_p(S)) = \Phi_y(S_y),$$

where $S_y = S_x = \exp_S F_0$. Let $v \in T_x N_r(p_0)$ be an arbitrary vector and let $F_t \in \mathscr{P}(S)$, $t \in (-\varepsilon, \varepsilon)$, be a smooth curve such that $\frac{\partial}{\partial t}|_{t=0} \exp_{p_0} F_t(p_0) = v$. Consider the variation $c: S \times (-\varepsilon, \varepsilon) \to M$ defined by $c(p, t) = \exp_p F_t(p) = h(p, \exp_{p_0} F_t(p_0))$ for all $p \in S$. For all $t \in (-\varepsilon, \varepsilon)$ we know that $\partial c/\partial t = d\exp|_{F_t}(\partial F_t/\partial t)$ is a parallel normal vector field along the pseudo-soul $\exp_S F_t$. Denote this vector field by $V(\cdot, t)$. When $t = 0$, we have

$$\|dh|_{(p,x)}(v)\| = \left\| \frac{\partial}{\partial t}\right|_{t=0} \exp_p F_t(p) \left\| = \|V(p, 0)\| = \|v\|.\right.$$

Thus $h|_{(p, N_r(p_0))}: \{p\} \times N_r(p_0) \to N_r(p)$ is an isometry. To complete the proof of our claim, we will show that $B_y(v, v) = 0$, where $B_y$ is the *second fundamental form* of $N_r(p)$ at $y$. Let $c_p$ be a curve such that $c_p(t) = c(p, t) = \exp_p F_t(p)$, and define for any vector $X_y \in T_y S_y$ a vector field $X: (-\varepsilon, \varepsilon) \to TM$ along the curve $c_p$ by $X(t) = dc|_{(p,t)} \circ dc^{-1}(X_y)$. Since $dc^{-1}(X_y) \in T_p S$ we have $[dc^{-1}(X_y), \frac{\partial}{\partial t}] = 0$ on $S \times (-\varepsilon, \varepsilon)$, and so $[X, V] = 0$ where they are defined. Moreover, for each $t \in (-\varepsilon, \varepsilon)$, $X(t)$ is a tangent vector of the pseudo-soul $\exp_S F_t$, along which $V(\cdot, t)$ is parallel and perpendicular. Thus

$$\begin{aligned} 0 = V\langle V, X\rangle &= \langle X, \nabla_V V\rangle + \langle \nabla_V X, V\rangle \\ &= \langle X, \nabla_V V\rangle - \langle \nabla_X V, V\rangle = \langle X, \nabla_V V\rangle. \end{aligned}$$

Hence $\nabla_V V$ is a normal vector field along $S_y$. Furthermore, by Corollary 2.4, we have $\nabla_X \nabla_V V - \nabla_V \nabla_X V = R(X, V)V = 0$, which implies $\nabla_X \nabla_V V = \nabla_V \nabla_X V = 0$. Thus $\nabla_V V(y) \in \Psi(y) = T_y N_r(p)$. Since, for any fixed $p \in S$, $c_p$ is a curve in $N_r(p)$ for $\varepsilon$ small enough and $V(p, t) = c_p'(t)$, we may conclude that $B_y(v, v) = 0$. Here $v = V(p, 0)$ is an arbitrary vector in $T_y N_r(p)$ and $B_y$ is a symmetric tensor. Therefore, $B_y = 0$ for all $y \in N_r(p)$, and hence $N_r(p)$ is totally geodesic in $M$.

For any $p \in S$, $x \in N_r(p_0)$, and $Y \in T_{(p,x)}(S \times N_r(p_0))$, let $v \in T_x N_r(p_0)$ and $X_p \in T_p S$ be such that $Y = v + X_p$. Then we have $dh|_{(p,x)}(Y) = dh(v) + dh(X_p) = V(p, 0) + X_y \in T_y \mathscr{H}_r(S)$ with the same notation as above. Hence $h$ is clearly an isometry. Let $B_y$ be the second fundamental form of $\mathscr{H}_r(S)$ at $y$. Then we use the same extension of $V$ and $X$ to show $B_y = 0$. In fact, since any vector in $T_y \mathscr{H}_r(S)$ can be orthogonally decomposed in the form of $V + X_y$, we only consider $B_y(V+X, V+X)$. Since $\nabla_X V = 0$ from above, we have $B_y(V, X) = 0$. Moreover, $N_r(p)$ and $S_y$ are both totally geodesic, and therefore

$$B_y(V+X, V+X) = B_y(V, V) + 2B_y(V, X) + B_y(X, X) = 0.$$

Hence $\mathscr{H}_r(S)$ is totally geodesic in $M$.

## 3. Proof of theorem

In this section, we will prove our main theorem combining all of the previous results. The following facts are crucial for the remaining part of our argument, but they are somewhat technical to be discussed here. We have put an appendix at the end to study these facts in detail.

For any compact t.c.s. $C$, $\partial C \neq \emptyset$, let $\psi: C \to \mathbf{R}$ be such that $\psi(q) = d(q, \partial C)$. Then $\psi$ is a *convex function*, and hence for each $b \in [0, a_0 = \sup\{\psi\}]$ the subset $C^b = \{q \in C | \psi(q) \geq b\}$ is totally convex as well. With this property of a totally convex set and the flag of t.c.s. in the construction of a soul, we obtain an *exhaustion* of $M$ by t.c.s.

If $H: M \times [0, 1] \to M$ is the homotopy of the Sharafutdinov retraction $f$ corresponding to a soul $S_0$, define for each $p \in M$ a continuous curve $\varphi_p: [0, 1] \to M$ by $\varphi_p(p) = H(p, t)$. Let $C$ be the t.c.s. of the totally convex exhaustion of $M$ such that $p \in \partial C$. We then reparametrize $\varphi_p$ so that $\psi(\varphi_p(t)) = t \leq a_0$ for $\psi = d(\cdot, \partial C)$ and $a_0 = \sup\{\psi\}$. In the appendix we show the following.

(A1) For each $t \in [0, a_0)$, $\varphi_p(t)$ has a *right tangent vector* $\nabla\psi(\varphi_p(t))/\|\nabla\psi\|^2$ (Proposition A.8), where $\nabla\psi$ is a (generalized) gradient of $\psi$ (Definition A.4).

Note that $\nabla\psi$ is independent of $C$ in the totally convex exhaustion, and hence is a well-defined (not even continuous in general) vector field on $M$.

(A2) If a pseudo-soul $S \subset M$ is not a unique soul, then it is completely contained in $\partial C$ for some t.c.s. $C$ in the totally convex exhaustion of $M$ (Proposition A.2(4)), and $\nabla\psi$ is a parallel normal vector field along $S$ (Corollary A.9).

(A3) (Theorem A.5(3)) For each fixed $a \in (0, a_0)$, there exists a number $A > 0$ such that for any $p \in \partial C^t$, $t \in [0, a]$, there is $\varepsilon > 0$ with

$$\psi\left(\exp_p \varepsilon \frac{\nabla\psi}{\|\nabla\psi\|}(p)\right) \geq t + A\varepsilon.$$

(A4) (Corollary A.6) For any $a, b \in [0, a_0)$, $\varphi_p|_{[a,b]}$ is a rectifiable curve.

According to our construction of $(H, f)$, we have in fact defined $\varphi_p$ as an "*integral curve*" of the vector field $\nabla\psi$, which is canonically defined by the totally convex exhaustion. The third claim (A3) is equivalent to saying that away from the maximal valued set $C^{a_0}$ of $\psi$, the gradient $\|\nabla\psi\|$ is bounded below by a positive number; (A4) is a consequence of (A3). Another interpretation of the vector $\nabla\psi(p)$, $p \in \partial C^t$, is the *center of the tangent cone* of $C^t$ at $p$, which implies for any vector $v \in T_pM$ in the tangent cone (i.e., $\exp_p tv$ is an interior point of $C^t$ for small $t > 0$), we have $\angle(\nabla\psi(p), v) < \pi/2$. All of the results above are more precisely stated in the appendix, which should be referred to for the details.

As mentioned earlier, we want to approximate the continuous family of pseudo-souls by a broken geodesic, and then apply a curve shortening process. We first make a definition for this type of connection, and show how the curve shortening process applies.

**Definition 3.1.** Two pseudo-souls $S_0$ and $S_1$ are called (P)-*connected* by a broken geodesic $\gamma\colon [0, 1] \to M$ if they satisfy the following properties:

(1) There is a partition $0 = t_0 < t_1 < \cdots < t_m = 1$ such that for each $i$, $\gamma|_{[t_{i-1}, t_i]}$ is a geodesic.

(2) For each $i$, $\gamma(t_i)$ is contained in a pseudo-soul $S_{t_i}$, and $\gamma'(t_i^+)$ and $\gamma'(t_i^-)$ are both perpendicular to $S_{t_i}$, where $\gamma'(t_i^\pm)$ denotes $\lim_{t\to t_i^\pm} \gamma'(t)$. In particular, $\gamma(0) \in S_0$ and $\gamma(1) \in S_1$.

(3) For each $i$, $\gamma|_{[t_{i-1}, t_i]}$ can be extended to a parallel connection between $S_{t_{i-1}}$ and $S_{t_i}$, i.e., $\gamma'(t_{i-1}^+)$ has a parallel extension $F$ along $S_{t_{i-1}}$ such that $S_{t_i} = \exp_{S_{t_{i-1}}}(t_i - t_{i-1})F$.

**Lemma 3.2.** *Let $S_1$ and $S_2$ be two pseudo-souls (P)-connected by a broken geodesic $\gamma\colon [0, 1] \to M$. Then they can be (P)-connected by a smooth geodesic $\gamma_0\colon [0, 1] \to M$ with the following property*:

$$(\mathrm{P}_1)\colon \gamma(0) = \gamma_0(0) \in S_1 \textit{ and } \gamma(1) = \gamma_0(1) \in S_2.$$

*Proof.* We will prove the existence of a smooth geodesic (P)-connecting $S_1$ and $S_2$ by a *curve shortening process.* Let $D \subset M$ be a compact set such that $B_L(\gamma(0)) \subset D$, where $L = L_{[\gamma]}$. Then clearly the whole curve shortening process of $\gamma$ will be contained in $D$. Let $r > 0$ be such that $r < \varepsilon_D$ and let $m$ be an integer such that $L/m < r$. All curves will be assumed to be parametrized on $[0, 1]$ proportional to arclength.

Divide the broken geodesic $\gamma$ into $m$ equal segments, each of length $L/m$, by the division points $p_0, p_1, \cdots, p_m$. For each $i = 1, 2, \cdots, m$ $\gamma[t_{i-1}, t_i]$ is a broken geodesic contained in $N_r(p_{i-1})$ by Proposition 2.8. Replace each $\gamma[t_{i-1}, t]$ by a minimal geodesic in $N_r(p_{i-1})$ which is, by strong convexity, a minimal geodesic in $M$. Clearly $S_1$ and $S_2$ are (P)-connected by this new broken geodesic $\tilde{\gamma}$ with the property $(\mathrm{P}_1)$, and its length is strictly smaller than that of $\gamma$, except when $\gamma = \tilde{\gamma}$. Now take the $m$ midpoints of the segments of $\tilde{\gamma}$. Each pair of successive midpoints are at distance $< r$ apart, so it may be connected by a unique minimal geodesic as above. Denote this new broken geodesic by $D(\gamma)$ which is another connection from $S_1$ to $S_2$ with the property $(\mathrm{P}_1)$. The curve shortening process can be iterated to yield a sequence of broken geodesics:

$$\gamma_0 = \gamma, \gamma_1 = D(\gamma), \cdots, \gamma_i = D(\gamma_{i-1}), \ldots$$

For each $i$, $S_1$ and $S_2$ are $(\mathrm{P}_1)$-connected by $\gamma_i$, and the length of $\gamma_i$ is strictly less than that of $\gamma_{i-1}$ unless $\gamma_{i-1}$ is already a smooth geodesic. The existence of some subsequence of $\{\gamma_i\}$ converging to a smooth geodesic (P)-connecting $S_1$ and $S_2$ is guaranteed by Birkhoff (cf., [1]).

**Theorem 3.3.** *Let $C$ be a compact totally convex set and let $S_0$ be the soul of $C$. Then there is a number $\mathcal{N} > 0$ such that every pseudo-soul in $C$ can be (P)-connected to $S_0$ by a geodesic of length bounded above by $\mathcal{N}$.*

*Proof.* Since there is a flag, $C \supset C(1) \supset \cdots \supset C(k) = S_0$, for each $i < k$ we assume, by induction, that there is a number $\mathcal{N}_i > 0$ such that every pseudo-soul in $C(i)$ can be (P)-connected to $S_0$ by a geodesic of length $\leq \mathcal{N}_i$, and then show that there exists $\mathcal{N}_{i-1}$ for $C(i-1)$. Since we will use the same argument for each $i$, we only consider the case when $i = 1$ (i.e., $C(i-1) = C$).

Put $C(1) = C^{a_0}$. For each $t \in (0, a_0)$ let $p_t$ be a point in $\partial C^t$ such that $d(p_t, C(1)) = \sup\{d(p, C(1)) | p \in \partial C^t\}$. If $d(p_t, C_1))$ does not converge to zero as $t \to a_0$, we have a subsequence of $\{p_t\}$ converging to $p_0 \in C^{t_0}$ for some $t_0 < a_0$, which is an obvious contradiction since it implies $d(p_t, p_0) \geq t - t_0$ for all $t > t_0$. Therefore $d(p_t, C(1)) \to 0$ as $t \to a_0$, and one find a number $a \in (0, a_0)$ such that $d(p, C(1)) < r < \varepsilon_C$ for any $p \in C^a$.

Let $S$ be any pseudo-soul in $C$. We assume that $S$ is not contained in $C^a$ and claim that $S$ can be (P)-connected to a pseudo-soul in $\partial C^a$ be a geodesic of length at most $aA^{-1}$, where $A > 0$ is the number in (A3) corresponding to $a \in (0, a_0)$ chosen above. By (A2) one can find $b_0 \in [0, a_0)$ such that $S \subset \partial C^{b_0}$. Since $\nabla\psi$ is a parallel normal vector field along $S$, $\exp_S t\nabla\psi/\|\nabla\psi\|$ is a family of pseudo-souls (P)-connected to $S$ by the geodesic $\gamma(t) = \exp_p t\nabla\psi/\|\nabla\psi\|$, $p \in S$. Moreover, by (A3), we can find $\varepsilon > 0$ such that $\psi \circ \gamma(\varepsilon) \geq \psi(p) + A\varepsilon$. Therefore, $S$ can be (P)-connected to a pseudo-soul $S_b \subset \partial C^b$ by the geodesic $\gamma$ whose length is at most $(b - b_0)A^{-1} \leq bA^{-1}$. Let $\overline{b} \in (b_0, a]$ be the least upper bound of $b > 0$ such that $S$ can be connected to $S_b \subset \partial C^b$ be a geodesic of length at most $bA^{-1}$. By taking a limit of $S_b$ as $b \to \overline{b}$, one can see the $\overline{b}$ has the same property. If $\overline{b}$ is strictly smaller than $a$, we use the parallel field $\nabla\psi$ to (P)-connect the pseudo-soul $S_{\overline{b}}$ to a pseudo-soul $S_b$, $b > \overline{b}$, by a geodesic of length at most $(b - \overline{b})A^{-1}$. Then, by curve shortening, one can (P)-connect $S$ to $S_b$ to obtain a contradiction.

We now (P)-connect $S_a$ to a pseudo-soul in $C(1)$. For each $p \in S_a$ let $g(p) \in C(1)$ be such that $d(p, g(p)) = d(p, C(1))$. Then by the choice of the number $a$, $d(p, g(p)) < r$. Let $c_p: [0, 1] \to C$ be the minimal geodesic from $p$ to $g(p)$. Since $c_p'(0)$ is in the tangent cone of $C^a$ at $p$, we have $\angle(\nabla\psi(p), c_p'(0)) < \pi/2$, and hence $d(g(p), \gamma(t))$ is strictly decreasing for small $t > 0$, where $\gamma(t) = \exp_p t\nabla\psi/\|\nabla\psi\|$. Thus there exists $\varepsilon > 0$ such that $\gamma(\varepsilon) \in \partial C^b$, $b > a$, and $\gamma[0, \varepsilon]$ is contained in the strongly convex ball $B_r(g(p))$. We connect $S_a$ to $S_b$ along $\gamma$. Let $\overline{b} \in (a, a_0]$ be the least upper bound of $b$ with this property. It is clear by a limiting argument that $\overline{b}$ has the property too. If $\overline{b} < a_0$, let $\gamma_1: [0, \varepsilon] \to B_r(g(p))$ be the geodesic which (P)-connects $S_a$ to $S_{\overline{b}}$, and let $c$ be the minimal geodesic from $p_1 = \gamma_1(\varepsilon_1)$ to $g(p)$. Then, for the same reason as above, we have $\angle(\nabla\psi(p_1), c'(0)) < \pi/2$, and hence $d(g(p), \gamma_2(t))$ is strictly decreasing for small $t > 0$, where $\gamma_2(t) = \exp_{p_1} t\nabla\psi$. Thus there exists $\varepsilon_2 > 0$ such that $\gamma_2[0, \varepsilon_2] \subset B_r(g(p))$. We then apply the curve shortening process for $\gamma_1[0, \varepsilon_1] \cup \gamma_2[0, \varepsilon_2]$, which clearly takes place in $B_r(g(p))$ and we get a contradiction. Therefore $S_a$ is now (P)-connected to a pseudo-soul in $C(1)$ by a geodesic which is minimal since it is contained in the strongly convex set $B_r(g(p))$.

By assumption, every pseudo-soul in $C(1)$ is (P)-connected to $S_0$ by a geodesic of length at most $\mathcal{N}_1$. Put $\mathcal{N}_0 = \mathcal{N}_1 + aA^{-1} + 2r$. Then the theorem follows by a final application of curve shortening.

Since every pseudo-soul is (P)-connected to the soul, by Corollary 2.7, we have the following immediate consequence.

**Corollary 3.4.** $\dim(\mathscr{P}(S))$ *is constant for any pseudo-soul* $S$.

Let $S_0$ be a soul of $M$. For each $q \in S_0$ denote by $N(q)$ the set $N_\infty(q) = \{\exp_q F(q) | F \in \mathscr{P}(S_0)\} \subset \nu(S_0)$, and define a map $h: S_0 \times N(q) \to M$ by $h(p, \exp_p F(q)) = \exp_p F(p)$ for all $p \in S_0$ (a global version of $h_p$ in Proposition 2.8). Our final goal is to prove that $h$ is a totally geodesic isometric embedding with its image $\mathscr{H} \subset M$, which is by definition the set $\exp_{S_0} \Phi(S_0) = \{\exp_q F(q) | q \in S_0, F \in \mathscr{P}(S_0)\}$. As a result of the previous lemmas, along with the fact that all souls are pseudo-souls to each other, we know every pseudo-soul is contained in $\mathscr{H}$ and $\mathscr{H}$ is a union of pseudo-souls. We first prove the following:

**Lemma 3.5.** *$\mathscr{H} \subset M$ is embedded.*

*Proof.* For any $p \in \mathscr{H}$ let $S_p$ be the pseudo-soul containing $p$, and let $r > 0$ be such that $h_p : S_p \to N_r(p) \to \mathscr{H}_r(S_p) \subset M$ is an isometric embedding. Then it will suffice to show that there exists a metric ball $B_{r_1}(p)$, $0 < r_1 \le r$, such that

$$B_{r_1}(p) \cap \mathscr{H} = B_{r_1}(p) \cap \mathscr{H}_r(S_p).$$

Suppose not. Then there exists a sequence $\{q_k\}$ of points in $\mathscr{H}$ such that $q_k \notin \mathscr{H}_r(S_p)$ and $q_k \to p$ as $k \to \infty$. For each $k$ let $S_k = \exp_{S_0} F_k$, $F_k \in \mathscr{P}(S_0)$, be the pseudo-soul containing $q_k$. Since all $S_k$'s are contained in some compact t.c.s., by Theorem 3.3 we may assume that $\|F_k\| \le \mathscr{N}$ for some $\mathscr{N} > 0$, and hence $\{F_k\}$ is uniformly convergent for some subsequence. Assume that $F_k \to F \in \mathscr{P}(S_0)$ as $k \to \infty$. Since $q_k \in S_k$ and $S_k \to \exp_{S_0} F$ as $k \to \infty$, it is easy to see by Corollary 1.4 that $\exp_{S_0} F = S_p$. We are going to show that if $k$ is large enough, then $S_k \subset \mathscr{H}_r(S_p)$, which is a contradiction and the lemma will follow.

If $p = \exp_q v$ for $q \in S_0$ and $v = F(q) \in \Phi_q(S_0)$, let $v_k \in \Phi_q(S_0)$ be such that $F_k(q) = v_k$. Then clearly $v_k \to v$ as $k \to \infty$. If $p$ is not conjugate to $q$ along the geodesic $\exp_q tv$, then by Proposition 2.6 we can find a ball $\widetilde{B}_{r_1}(v) \subset T_q M$ such that $\exp_q : \widetilde{B}_{r_1}(v) \to M$ is an embedding and $\exp_q(\widetilde{B}_{r_1}(v) \cap \Phi_q(S_0)) \subset N_r(p)$. Therefore, if $\|v_k - v\| < r_1$, then $S_k \subset \mathscr{H}_r(S_p)$. If $p$ is conjugate to $q$ along $\exp_q tv$, pick $w = t_1 v$ such that $p_1 = \exp_q w$ is not conjugate to $q$ and $p_1 \in N_r(p)$. Let $r_1 > 0$ be such that $N_{r_1}(p_1) \subset N_r(p)$. Since $v_k \to v$, it is clear that $w_k = t_1 v_k \to w$. By the same argument as above we can find a number $r_2 > 0$ such that if $\|w_k - w\| < r_2$, then $\exp_q w_k \in N_{r_1}(p_1) \subset N_r(p)$, for which we clearly have $\exp_q v_k = \exp_q(w_k / t_1) \in N_r(p)$ and $S_k \subset \mathscr{H}_r(S_p)$.

**Lemma 3.6.** *For any $p \in \mathscr{H}$ there exists $r > 0$ such that $N_r(p) \subset N(q)$ for some $q \in S_0$.*

*Proof.* Since there is a pseudo-soul $S$ containing $p$, we can find $q \in S_0$ and $F \in \mathscr{P}(S_0)$ such that $\exp_q F(q) = p$, and $\exp_{S_0} F = S$. Let $r > 0$ be such that $\mathscr{H}_r(S) \subset M$ is isometric to $S \times N_r(p)$. For any

$p_1 = \exp_p v \in N_r(p)$ let $v$ be extended to the parallel normal vector field $V$ along $S$. Then the pseudo-soul $\exp_S V$ is (P)-connected to $S_0$ by the broken geodesic $(\exp_q tF) \cup (\exp_p tV)$, and hence $p_1$ is contained in $N(q)$ by curve shortening.

**Proposition 3.7.** *If $f\colon M \to S_0$ and $H\colon M \times [0, 1] \to M$ are the canonical retraction and its homotopy, then for any $q \in S_0$ and any $t \in [0, 1]$ we have $H(N(q), t) \subset N(q)$. In particular, when $t = 1$, $f(N(q)) = q$.*

*Proof.* For any $p \in M$ find $q \in S_0$ and $F \in \mathscr{P}(S_0)$ such that $p = \exp_q F(q)$, and let $\varphi_p[0, 1] \to M$ be the curve such that $\varphi_p(t) = H(p, t)$. It then suffices to show that $\varphi_p[0, 1] \subset N(q)$. Let $C$ be a compact t.c.s. such that $p \in \partial C$. By induction over the flag of $C$, for each $i > 0$ we assume that $\varphi_p[0, t_{i-1}] \subset N(q)$, and then show that $\varphi_p[t_{i-1}, t_i] \subset N(q)$, where $t_i = \sup\{t \in [0, 1] | \varphi_p(t) \notin C(i)\}$. We may also assume that $i = 1$ since the argument will be the same for any compact t.c.s. Let $\varphi_p$ be reparametrized so that $\psi \circ \varphi_p(t) = t$ for the distance function $\psi = d(\cdot, \partial C)$.

Let $S = \exp_{S_0} F \subset \partial C$ be the pseudo-soul such that $p \in S$. By Lemma 3.5, there exists $r > 0$ such that $B_r(p) \cap \mathscr{H} = B_r(p) \cap \mathscr{H}_r(S)$, where $\mathscr{H}_r(S)$ is isometric to $S \times N_r(p)$. Let $\delta > 0$ be such that $\varphi_p[0, \delta] \subset B_r(p)$, and first try to prove that $\varphi_p[0, \delta] \subset N(q)$. Since $H(S, t)$ is a pseudo-soul for each $t \in [0, 1]$, we have $\varphi_p[0, 1] \subset \mathscr{H}$, and hence $\varphi_p[0, \delta] \subset H_r(S)$. Therefore, by lemma 3.6, it suffices to show $\varphi_p[0, \delta] \subset N_r(p)$. We now consider $\varphi_p$ as a curve in the product space $S \times N_r(p)$, in which $\varphi_p$ can be expressed in the following form:

$$\tilde{\varphi}_p(t) = h_p^{-1} \circ \varphi_p(t) = (c(t), \sigma(t)) \in S \times N_r(p).$$

For any $\tau > 0$ let $\gamma_\tau\colon [0, \tau] \to C$ be the minimal geodesic from $\varphi_p(0)$ to $\varphi_p(\tau)$. By (A1), we have

$$\lim_{r \to 0^+} \gamma_\tau'(0) = \frac{\nabla \psi}{\|\nabla \psi\|^2}(p).$$

For any $\varepsilon > 0$ there is $\tau > 0$ such that $\angle(\gamma_\tau'(0), \nabla\psi(p)) < \varepsilon$. However, since $\nabla\psi$ is parallel along $S$, the geodesic $\exp_p t\nabla\psi$ is contained in $N_r(p)$ for small $t > 0$. If we have chosen $\tau$ small enough, the map $d(\gamma_\tau(\tau), \exp_p t\nabla\psi)$ attains its minimum when $t = t'$ such that $\exp_p t'\nabla\psi \in N_r(p)$, and then by Toponogov we have

$$d(\gamma_\tau(\tau), \exp_p t'\nabla\psi) \le d(\gamma_\tau(\tau), p) \sin\varepsilon \le L[\varphi_p|_{[0,\tau]}] \sin\varepsilon.$$

Since $d(c(\tau), p) = d(\gamma_\tau(\tau), N_r(p)) \le d(\gamma_\tau(\tau), \exp_p t'\nabla\psi)$, it follows that $d(c(\tau), p) \le L[\varphi_p|_{[0,\tau]}]\sin\varepsilon$. Let $\overline{\tau} \le \delta$ be the least upper bound of $\tau \in [0, \delta]$ with this property. By a limiting argument we know $\overline{\tau}$ has that property. If $\overline{\tau} < \delta$, by the same argument as above, we can find $\tau > 0$ such that $d(c(\overline{\tau}+\tau), c(\overline{\tau})) \le L[\varphi_p|_{[\overline{\tau},\overline{\tau}+\tau]}]\sin\varepsilon$. Then

$$\begin{aligned} d(c(\overline{\tau}+\tau), p) &\le d(c(\overline{\tau}+\tau), c(\overline{\tau})) + d(c(\overline{\tau}), p) \\ &\le (L[\varphi_p|_{[\overline{\tau},\overline{\tau}+\tau]}] + L[\varphi_p|_{[0,\overline{\tau}]}])\sin\varepsilon \\ &= L[\varphi_p|_{[0,\overline{\tau}+\tau]}]\sin\varepsilon, \end{aligned}$$

which is a contradiction. Therefore, we may conclude that for any $\varepsilon > 0$ and any $t \in [0, \delta]$ we have $d(c(t), p) \le \varepsilon L[\varphi_p|_{[0,\delta]}]$, which implies $c(t) = p$ for all $t \in [0, \delta]$, and hence $\varphi_p[0, \delta]$ is contained in $N(q)$. Let $\overline{\delta} \in [0, t_1]$, $t_1 = \sup\{t | \varphi_p(t) \notin C(1)\}$, be the least upper bound of $\delta$ such that $\varphi_p[0, \delta] \subset N(q)$. If $\overline{\delta} < t_1$, consider a sequence $s_k$ such that $s_k \to \overline{\delta}$ as $k \to \infty$ and $\varphi_p(s_k) \in N(q)$. Let $F_k \in \mathscr{P}(S_0)$ be such that $\exp_p F_k(q) = \varphi_p(s_k)$. By Theorem 3.3, $\|F_k\|$ is bounded by some number, and hence $\{F_k\}$ has a subsequence converging to $F \in \mathscr{P}(S_0)$. Clearly $\exp_q F(q) = \varphi_p(\overline{\delta})$, and it follows that $\varphi_p(\overline{\delta}) \in N(q)$. By the same argument as above, we can extend $\overline{\delta}$ to be a larger number to obtain a contradiction.

**Corollary 3.8.** *For any $q \in S_0$ the subset $N(q) \subset M$ is a totally geodesic embedded submanifold of $M$, which is diffeomorphic to $\mathbf{R}^k$, $k = \dim(\mathscr{P}(S_0))$.*

*Proof.* Let $p \in N(q)$ be such that $p = \exp_q F(q)$, $F \in \mathscr{P}(S_0)$, and let $S = \exp_{S_0} F$. Choose a metric ball $B_r(p) \subset M$ such that $B_r(p) \cap \mathscr{H} = B_r(p) \cap \mathscr{H}_r(S)$, where $\mathscr{H}_r(S) = h_p(S \times N_r(p))$. For any $p_1 \in S$, $p_1 = \exp_{q_1} F(q_1)$, we have $h_p(p_1 \times N_r(p)) = N_r(p_1)$. Thus, by Lemma 3.6 and Proposition 3.7, we have $N_r(p_1) \subset N(q_1)$ and $f(N_r(p_1)) = q_1$. Since $f$ is a well-defined function, we have $N(q) \cap \mathscr{H}_r(p) = N_r(p)$. Then $N_r(p) \subset N(q)$ and $B_r(p) \cap \mathscr{H} = B_r(p) \cap \mathscr{H}_r(S)$ implies that $B_r(p) \cap N(q) = B_r(p) \cap \mathscr{H}_r(S) \cap N(q) = N_r(p)$. Therefore $N(q)$ is embedded in $M$. Furthermore, since each $N_r(p) \subset M$ is totally geodesic, so is $N(q)$.

By Proposition 3.7, for each $q \in S_0$ and any $t \in [0, 1]$ we have $H(N(q), t) \subset N(q)$, which means $H|_{N(q)}: N(q) \times [0, 1] \to N(q)$ is a homotopy such that $H(\cdot, 0) = \mathrm{id}$ on $N(q)$ and $H(N(q), 1) = q$. Thus $N(q)$ is contractible. In fact, since $N(q) \cap S_0 = \{q\}$ and $N(q)$ is totally geodesic, it is easy to see that $N(q)$ has a point soul $q$, and hence is diffeomorphic to $\mathbf{R}^k$.

We now obtain the proof of our main theorem by combining all of our previous results.

**Theorem 3.9.** *For any $q_0 \in S_0$ the map $h: S_0 \times N(q_0) \to M$ is a totally geodesic isometric embedding with its image $\mathscr{H}$.*

*Proof.* We first show that $h$ is a well-defined injective map. Suppose there are two points $q_1, q_2 \in S_0$ and two parallel vector fields $F_1, F_2 \in \mathscr{P}(S_0)$ such that $\exp_{q_1} F_1(q_1) = \exp_{q_2} F_2(q_2)$. By Corollary 1.4, we have $\exp_{S_0} F_1 = \exp_{S_0} F_2 = S$. For any $p \in S$, if $p_1, p_2 \in S_0$ are such that $\exp_{p_1} F_1(p_1) = \exp_{p_2} F_2(p_2) = p$, then by Proposition 3.7, $p_1 = f(\exp_{p_1} F_1) = f(\exp_{p_2} F_2) = p_2$, which implies that $q_1 = q_2$ and $\exp_q F_1(q) = \exp_q F_2(q)$ for any $q \in S_0$. Thus $h$ is injective. In particular, when $q_1 = q_2 = q_0$, we see that $h$ is well defined.

For any $p = \exp_{q_0} F(q_0)$, $F \in \mathscr{P}(S_0)$, the pseudo-soul $S = \exp_{S_0} F$ has a neighborhood $\mathscr{H}_r(S) = h_p(S \times N_r(p))$ such that $N_r(p) \subset N(q_0)$. Then, for any $p_1 \in N_r(p)$ we have $h(S_0 \times \{p_1\}) = h_p(S \times \{p_1\})$ by curve shortening. Therefore it follows that $h|_{S_0 \times N_r}(p) = h_p$ when $S$ is identified to $S_0$. By Lemma 3.5 and Proposition 2.8, $h$ is a totally geodesic isometric embedding.

If the holonomy group of the normal bundle of a soul is trivial, i.e., if every vector $v \in \nu(S_0)$ has a parallel extension over $S_0$, then we have the following immediate consequence of the theorem.

**Corollary 3.10.** *If the normal bundle $\nu(S_0)$ of a soul is a parallel, i.e., $\nu(S_0) = \Phi(S_0)$, then $M$ is isometric to $S_0 \times N$, where $N$ is a totally geodesic embedded submanifold of $M$, which is diffeomorphic to $\mathbf{R}^k$, $k = \operatorname{codim}(S_0)$.*

In [7, Corollary 5] it was shown that if $\operatorname{codim}(S_0) = 2$ and $\nu(S_0)$ is flat, then $M \to S_0$ is locally isometrically a product. Using the corollary above, one can easily generalize the argument to obtain the following.

**Corollary 3.11.** *If the normal bundle $\nu(S_0)$ of a soul $S_0$ is flat, i.e., if the normal holonomy group is locally trivial, then there is a Riemannian submersion $M \to S_0$ which splits locally isometrically.*

## Appendix

In this section, we will first review the construction of the Sharafutdinov retraction $f$, and investigate the geometry of t.c.s. (totally convex set) in nonnegatively curved manifolds. Some of the results discussed in this appendix may also be found in [4] or [5]. However, our approach here will be more geometric and sometimes simpler than Sharafutdinov's arguments. Since we have to use some results from §§1 and 2, the logical place where the content of this appendix could be inserted is between §§2 and 3.

As was shown in the construction of a soul, there exists a filtration $C_t$ of $M$ by t.c.s., and a flag of t.c.s. such that

$$\bigcup_{t\leq 0} C_t = M,$$
$$C_0 = C(0) \supset C(1) \supset \cdots \supset C(k) = S_0.$$

Furthermore, for any compact t.c.s $C$, $\partial C \neq \emptyset$, one can define a function $\psi: C \to \mathbf{R}$ by $\psi(q) = d(q, \partial C)$. Then $\psi$ is a *convex function* [3, Theorem 1.10], i.e., for any normal geodesic segment $c$ contained in $C$, we have

$$\psi \circ c(\alpha t_1 + \beta t_2) \geq \alpha\psi \circ c(t_1) + \beta\psi \circ c(t_2),$$

where $\alpha, \beta > 0$, $\alpha + \beta = 1$.

Put $a_0 = \sup\{\psi(q) | q \in C\}$. Then, for each $b \in [0, a_0]$, the subset $C^b = \{q \in C | \psi(q) \geq b\}$ is totally convex. Therefore, in fact, there exists an *exhaustion* of $M$ by t.c.s., which means for any $p \in M$ one can find a t.c.s. $C$ ($C_t$, or $C(i)^b$) such that $p \in \partial C$.

Because of this totally convex exhaustion of $M$, to construct a deformation retraction from $M$ to $S_0$, it suffices to show that for any compact totally convex set $C$ and any two numbers $a, b$, $0 \leq a < b \leq a_0$, there exists a retraction $f_a^b: C^a \to C^b$ such that $f_a^b|_{C^b} = \mathrm{id}$ [8, Theorem 2.3].

For any $a, b$, $0 \leq a < b \leq a_0$, let $P_k = \{a = t_0 < t_1 < \cdots < t_{2^k} = b\}$ be the partition of $[a, b]$ into $2^k$ equal segments. Define $f_k: C^a \to C^b$ by $f_k = g_{2^k} \circ g_{2^k-1} \circ \cdots \circ g_1$, where $g_i: C^{t_{i-1}} \to C^{t_i}$ is a projection, i.e., for each $q \in C^{t_i}$, $d(q, g_i(q)) = d(q, C^{t_i})$. Then, by Ascoli's theorem, a subsequence of $\{f_k\}$ converges to a continuous function $f_a^b: C^a \to C^b$ as $k \to \infty$. For any $p \in \partial C^\alpha$ and any partition $P_k$ let $\gamma_{p,k}$ be the broken geodesic which minimizes distance from $g_{i-1} \circ \cdots \circ g_1(p)$ to $g_i \circ \cdots \circ g_1(p)$ for each $i$. Each $\gamma_{p,k}$ is assumed to be parametrized so that $\gamma_{p,k}(t_i) = g_i \circ \cdots \circ g_1(p)$. Then, as $\{f_k\}$ converges to $f_a^b$, $\{\gamma_{p,k}\}$ converges to a continuous curve. Put $H_a^b(p, t) = \lim_{k\to\infty} \gamma_{p,k}(t)$. For any $q \in C^a$, if $q = H_a^b(p, t_0)$ for some $p \in \partial C^a$ and $t_0 \in [a, b]$, we define $H_a^b(q, t)$ as follows:

$$H_a^b(q, t) = \begin{cases} q & \text{if } t \leq t_0, \\ H_a^b(p, t) & \text{otherwise.} \end{cases}$$

Thus it is clear $H_a^b\colon C^a \times [0, 1] \to C^a$ is a homotopy such that $H_a^b(\cdot, 0) =$ id and $H_a^b(\cdot, 1) = f_a^b$.

We now use the totally convex exhaustion of $M$ to extend this partial construction to a deformation retraction $f\colon M \to S$ and its homotopy $H\colon M \times [0, 1] \to M$. In fact, for any $p \in M$ let $C$ be the t.c.s. ($C_t$, $t \geq 0$, or $C(i)^b$, $i \geq 0$, $b \in [0, a_i]$) such that $p \in C$. Then define $H(p, t)$ as a composition of the $H_a^b$'s constructed above for $C$ with a suitable change of parametrization. According to our construction, the choice of a homotopy $H$ (and hence $f$) may not be unique. We will, however, make a choice and call it canonical. Then the following was shown in [8].

**Theorem A.1.** *$H\colon M \times [0, 1] \to M$ is a continuous map such that $H(\cdot, 1) = \mathrm{id}$ on $M$ and $H(\cdot, 1) = f$, where $f\colon M \to S_0$ is a strong deformation retraction. Furthermore, for each $t \in [0, 1]$, $H(\cdot, t)$ is distance nonincreasing.*

For any compact t.c.s. $C$ and $p \in \partial C$ the *tangent cone* $C_p$ at $p$ is defined in [3]. For any $p \in C$ let $b \in [0, a_0]$ be such that $p \in \partial C^b$. We then use the same notation $C_p$ to denote the tangent cone of $C^b$ at $p$, which is by definition the set

$$\{v \in T_pM \mid \exp_p tv/\|v\| \in \mathring{C}^b \text{ for some positive } t < r(p)\} \bigcup \{0\},$$

where $\mathring{C}^b$, is the interior points of $C^b$, and $r(p)$ is the *convexity radius* of $C$ at $p$ [2, Theorem 5.14]. Let $\widehat{C}_p \subset T_pM$ be the subspace spanned by $C_p$. Then we have the following known facts.

**Proposition A.2** [8, §1]. (1) *If $a$ is a number such that $0 < a < a_0 = \sup\{\psi\}$, then there exists an angle $\theta > 0$ such that for any $t \in [0, a]$ and any $p \in \partial C^t$ the tangent cone $C_p$ contains a circular cone,*

$$C_p(v_p, \theta) = \{v \in \widehat{C}_p \mid \angle(v_p, v) \leq \theta\},$$

*for some $v_p \neq 0$.*

(2) *For any $b \in (0, a_0)$ and any $p \in \partial C^b$ we have*

$$\begin{aligned} C_p^* &\overset{\text{def}}{=} \{v \in \widehat{C}_p \mid \langle v, w\rangle < \text{ for all } w \in C_p - \{0\}\} \\ &= \{v \in \widehat{C}_p \mid d(\exp_p tv/\|v\|, \partial C^b) = t \text{ for small } t > 0\} \\ &= \text{the convex hull of } \{v_i\}_p, \end{aligned}$$

*where $\{v_i\}_p = \{v \in \widehat{C}_p \mid \exp_p tv/\|v\| \in \partial C^{b-t}$ for small $t > 0\}$.*

(3) *If a geodesic* $\gamma$ *is contained in* $\partial C^b$, $0 < b < a_0$, *then* $\{v_i\}_p$ *(and also* $C_p^*$ *) is orthogonal to* $\gamma'(t)$ *and invariant under parallel translation along* $\gamma$.

(4) *Any compact totally geodesic submanifold in* $C$ *is completely contained in* $\partial C^b$ *for some* $b \in [0, a_0]$.

**Theorem A.3.** *Let* $C$ *be a compact t.c.s.,* $\partial C \neq \emptyset$, *and let* $\psi(q) = d(q, \partial C)$. *For any* $p \in C$ *and any* $X \in \widehat{C}_p$ $(X \in C_p$ *if* $p \in \partial C)$, $\psi$ *has a right derivative* $X^+(\psi)$, *i.e., for any smooth curve* $\gamma: (-\varepsilon, \varepsilon) \to M$ *with* $\gamma(0) = p$, $\gamma'(0) = X$, *the right limit*

$$X^+(\psi) = \lim_{t \to 0^+} \frac{\psi \circ \gamma(t) - \psi(p)}{t}$$

*exists independently of* $\gamma$. *Furthermore, the following hold.*

(1) *If* $p \in \mathring{C}$ *and* $X \in \widehat{C}_p$, *then* $X^+(\psi) = -\|X\| \cos\alpha$, *where* $\alpha = \inf\{\angle(X, v) | v \in \{v_i\}_p\}$, *and* $\{v_i\}_p$ *is defined in Proposition* A.2(2).

(2) *If* $p \in C$ *and* $X \in C_p$, *then* $X^+(\psi) = \|X\| \sin\beta$, *where* $\beta = \inf\{\angle(X, v) | v \in \partial C_p\}$, *and* $\partial C_p = \overline{C}_p - C_p$ $(\overline{C}_p =$ *the closure of* $C_p)$.

*Proof.* We first observe that for any interior point $p$ and any $X \in C_p$, the two expressions for $X^+(\psi)$ above are consistent by Proposition A.2(2). On the unit sphere $S(1) \subset \widehat{C}_p$, if a set $(\{v_i\}_p \cap S(1))$ is strictly contained in the upper hemisphere, then the distances from the south pole $(X/\|X\|)$ to the set and to its convex hull are same. In our case, since $X \in C_p$ implies $\alpha > \pi/2$, and $C_p^*$ is the convex hull of $\{v_i\}_p$, we see that

$$\begin{aligned} \alpha &= \inf\{\angle(X, v) | v \in \{v_i\}_p\} \\ &= \inf\{\angle(X, v) | v \in C_p^*\} = \beta + \pi/2, \end{aligned}$$

and hence $-\cos\alpha = -\cos(\beta + \pi/2) = \sin\beta$. If the right limit $X^+(\psi)$ exists for any smooth curve, then it will be the same for all curves with the same initial conditions because $\psi$ is a *Lipschitz function.* We assume that $\gamma$ is a geodesic.

For any $p \in \mathring{C}$ and $X \in \widehat{C}_p$ let $b > 0$ be such that $p \in \partial C^b$. By the convexity of the function $\psi$ [3, Theorem 1.10] we have for any $v \in \{v_i\}_p$ that

$$\psi \circ \gamma(t) = d(\gamma(t), \partial C) \leq b - t\|X\| \cos(\angle(X, v)).$$

Thus

$$\limsup_{t \to 0^+} \frac{\psi \circ \gamma(t) - \psi(p)}{t} \leq -\|X\| \cos\alpha.$$

To show the opposite direction of the inequality, we suppose that there exist a strictly decreasing sequence $\{t_k\}$ converging to zero such that

$$\lim_{k\to\infty} \frac{\psi \circ \gamma(t_k) - \psi(p)}{t_k} < -\|X\| \cos\alpha.$$

Let $a \in (0, b)$ be such that $b - a = \delta < r/2$, where $r < \varepsilon_C$. Consider the sequence $\{p_k = \gamma(t_k)\}$ which clearly converges to $p$ as $k \to \infty$. For each $k$ let $q_k \in \partial C^a$ be such that $d(p_k, q_k) = d(p_k, \partial C^a)$. Since $\partial C^a$ is compact, there exists a convergent subsequence of $\{q_k\}$, and we assume (by abuse of notation) that $q_k \to q \in \partial C^a$. Then it is clear that $d(p, q) = \delta$ and there is a normal minimal geodesic $c\colon [0, \delta] \to C$ from $p$ to $q$ such that $w = c'(0) \in \{v_i\}_p$. Put $\alpha_0 = \angle(X, w)$. Then $\alpha \le \alpha_0$, so there exist $\mathcal{N} \in \mathbf{Z}$ and $\varepsilon > 0$ such that $\psi\circ\gamma(t_k) \le b - t_k\|X\|\cos(\alpha_0 - \varepsilon)$ for any $k \ge \mathcal{N}$. We now pick $k \ge \mathcal{N}$ large enough that $p_k = \gamma(t_k) \in B_r(q_k)$ and $\angle(c'(0), c_k'(0)) < \varepsilon/2$, where $c_k$ is the minimal geodesic from $p$ to $q_k$. Then $\angle(X, c_k'(0)) \ge \alpha_0 - \varepsilon/2$. Therefore, for any $t > 0$ with $\exp_p tX \in B_r(q_k)$, we have by strong convexity

$$\begin{aligned} d(q_k, \exp_p tX) &\ge d(q_k, p) - t\|X\|\cos(\alpha_0 - \varepsilon/2) \\ &\ge \delta - t\|X\|\cos(\alpha_0 - \varepsilon/2), \end{aligned}$$

which is a contradiction since

$$d(q_k, \exp_p t_k X) = d(q_k, pk) = \psi \circ \gamma(t_k) - a \le \delta - t_k\|X\|\cos(\alpha_0 - \varepsilon).$$

We may now conclude that

$$\liminf_{t\to 0^+} \frac{\psi \circ \gamma(t) - \psi(p)}{t} \ge -\|X\|\cos\alpha,$$

and hence the right limit $X^+(\psi)$ exists and (1) is satisfied.

The only remaining case is when $p \in \partial C$ and $X \in C_p$. By the same convexity of $\psi$, we obtain the following:

$$\limsup_{t\to 0^+} \frac{\psi \circ \gamma(t) - \psi(p)}{t} \le \|X\| \sin\beta.$$

Assume again that there exists a sequence $\{t_k\}$ for which we have a strict inequality. Let $q_k \in \partial C$ be defined such that $d(p_k = \gamma(t_k), q_k) = d(p_k, \partial C)$. Then $q_k \to p$ as $k \to \infty$. For each $k$ let $c_k\colon [0, 1] \to M$ be the minimal geodesic from $p$ to $q_k$, and set $w_k = \|c_k'(0)\|^{-1} c_k'(0) \in \overline{C}_p$. We assume (after taking a subsequence if necessary) that as $k \to \infty$, $\{w_k\}$ converges to a unit vector $w \in \overline{C}_p$. Since $q_k \to p$ as $k \to \infty$, it is easy to see that $w \notin C_p$, and hence $w \in \partial C_p$. We then use a similar argument for $w$ as in the first case to obtain a contradiction.

**Definition A.4.** For any $p \in C$ $(p \notin C^{a_0})$ a unit vector $v_p \in \widehat{C}_p$ will be called a *(generalized) gradient direction* if

$$v_p^+(\psi) = \sup\{v^+(\psi) | v \in \widehat{C}_p\,, \|v\| = 1\},$$

and denote $\nabla\psi = v_p^+(\psi)v_p$.

If $p \in C^{a_0}$, then we will define a gradient direction at $p$ as a point in the totally convex set $C^{a_0}$. Since $v^+(\psi)$ is positive only if $v \in C_p$, it is obvious that $v_p \in C_p$, and hence the definition still makes sense for $p \in \partial C$.

By Proposition A.2 we know several properties of $C_p$ $(\{v_i\}_p$, or $C_p^*)$. We will use these facts to examine the properties of $\nabla\psi$.

**Theorem A.5.** *With $C$ and $\psi$ as above, we have the following.*

(1) *For each $p \in C$, $\nabla\psi(p) \in C_p$ is unique.*

(2) *For any $b \in (0, a_0)$ and any geodesic $\gamma$ contained in $\partial C^b$, $\nabla\psi$ is perpendicular to $\gamma'(t)$ and parallel along $\gamma$.*

(3) *For any $a \in [0, a_0)$ there is an angle $\theta > 0$ such that for any $t \in [0, a]$ and any $p \in \partial C^t$ we have $\|\nabla\psi(p)\| \geq \sin\theta$.*

*Proof.* As $C_p^*$ is convex, it is easy to see that there exists a unique *minimal circular cone* containing $C_p^*$ with its center $-v_p \in C_p^*$. Then clearly the function $\inf\{\angle(\cdot\,, v) | v \in C_p^*\}$ attains its maximum at $v_p \in C_p$, and by definition we have $\nabla\psi = v_p^+(\psi)v_p$, where we assumed $\|v_p\| = 1$. Therefore $\nabla\psi(p)$ is unique for each $p \in C$. By Proposition A.2(3), $C_p^*$ is parallel normal to any geodesic in $\partial C^b$, $b \in (0, a_0)$. Thus the minimal cone is parallel along the geodesic, and so is the gradient $\nabla\psi$. Moreover, since $-v_p \in C_p^*$, $\nabla\psi(p)$ is perpendicular to the geodesic. By Proposition A.2(1) and Theorem A.3, the last claim (3) easily follows.

Let $S_0$ be a soul and let $H$ be the canonical homotopy corresponding to $S_0$. For any $p \in M$ put $\varphi_p(t) = H(p, t)$. If $C$ is the t.c.s. of the totally convex exhaustion of $M$ such that $p \in \partial C$, we then reparametrize the continuous curve $\varphi_p$ so that $\psi(\varphi_p(t)) = t$ for $t \leq a_0 = \sup\{\psi(x) | x \in C\}$. By the definition of $H$, for any $a, b$, $0 \leq a < b \leq a_0$, the curve $\varphi_p[a, b]$ can be obtained as a limit of the broken geodesic $\gamma_{p,k}$ with the partition $P_k = \{a = t_0 < \cdots < t_{2^k} = b\}$ (note that the set of dyadic numbers is dense in $[0, 1]$). For this continuous curve we have the following corollary.

**Corollary A.6.** *For any $a, b$, $0 \leq a < b < a_0$, $\varphi_p: [a, b] \to C$ is a rectifiable curve and*

$$L[\varphi_p] \leq (b - a)\sup\{\|\nabla\psi(p)\|^{-1} | p \in \partial C^t\,,\ a \leq t \leq b\}.$$

*Proof.* By construction, $\varphi_p$ is a uniform limit of the broken geodesic $\gamma_{p,k}: [a, b] \to C$, which is defined to be such that

$$L[\gamma_{p,k}|_{[t_{i-1}, t_i]}] = d(\gamma_{p,k}(t_{i-1}), \partial C^{t_i})$$

for the partition $P_k = \{a = t_0 < \cdots < t_{2^k} = b\}$. Hence it suffices to prove that for any $k$ the length $L[\gamma_{p,k}]$ is bounded by the above number, and then it is enough to show for each $i$

$$L[\gamma_{p,k}|_{[t_{i-1}, t_i]}] \le (t_i - t_{i-1}) \sup\{\|\nabla\psi\|^{-1}\}.$$

For any $t_{i-1}, t_i \in [a, b]$, to simplify notation put $t_i - t_{i-1} = \delta$ and let $\gamma: [0, \delta] \to C$ be such that $\gamma(s) = \gamma_{p,k}(t_i - s)$. Then we have $q = \gamma(0) \in \partial C^{t_i}$ and $d(\gamma(s), \partial C^{t_i}) = L[\gamma|_{[0,s]}]$, and hence by Proposition A.2(2) it follows that $\gamma'(0) \in C_q^*$. Let $\theta > 0$ be the angle such that $\|\nabla\psi(q)\| = \sin\theta$, which means the circular cone $C_q(\nabla\psi(q), \theta)$ is contained in $\overline{C}_q$. Then, by Proposition A.2(2), we see that $C_q^*$ is contained in $C_q(-\nabla\psi(q), \pi/2 - \theta)$, and therefore $\inf\{\angle(\gamma'(0), v) | v \in \{v_i\}_q\} \le \pi/2 - \theta$. We now apply Theorem A.3 to obtain

$$(\gamma'(0))^+(\psi) \le -\|\gamma'(0)\| \cos(\pi/2 - \theta) = -\|\gamma'(0)\| \sin\theta = -\|\nabla\psi(q)\| L[\gamma]/\delta.$$

Then by the convexity of $\psi \circ \gamma$ and the fact $\psi \circ \gamma(\delta) - \psi \circ \gamma(0) = t_{i-1} - t_i = -\delta$ we have $(\gamma'(0))^+(\psi) \ge -1$. Since $L[\gamma_{p,k}|_{[t_{i-1}, t_i]}] = L[\gamma]$,

$$L[\gamma_{p,k}|_{[t_{i-1}, t_i]}] \le \delta \|\nabla\psi(q)\|^{-1} \le (t_i - t_{i-1}) \sup\{\|\nabla\psi\|^{-1}\}.$$

Thus the corollary is proved.

**Lemma A.7.** *With $C$ and $\psi$ as above, for any $b \in [0, a_0)$ and $p \in \partial C^b$ let $\gamma_\tau: [0, \tau] \to C$ be the minimal connection from $p$ to $C^{b+\tau}$, $\tau > 0$, i.e., $d(p, \gamma_\tau(\tau)) = d(p, \partial C^{b+\tau})$. Then*

$$\lim_{\tau \to 0+} \gamma_\tau'(0) = \frac{\nabla\psi}{\|\nabla\psi\|^2}(p).$$

*Proof.* Put

$$\sigma(t) = \exp_p t \frac{\nabla\psi}{\|\nabla\psi\|}(p).$$

By the definition of $\nabla\psi$,

$$\|\nabla\psi\|(p) = \lim_{t \to 0^+} \frac{\psi \circ \sigma(t) - \psi(p)}{t}.$$

Thus $\psi \circ \sigma(t) = b + t(\|\nabla \psi\|(p) + O(t))$. If $\sigma(t) \in \partial C^{b+\tau}$ for small $t > 0$, then $\psi \circ \sigma(t) = b + \tau$. Since $\gamma_\tau$ is the minimal connection from $p$ to $\partial C^{b+\tau}$, we have $\tau\|\gamma'_\tau(0)\| = L[\gamma_\tau] \leq L[\sigma|_{0,t]}] = t$. Therefore,

$$\|\gamma'_\tau(0)\|^{-1} \geq \frac{\tau}{t} = \frac{\psi \circ \sigma(t) - b}{t} = \|\nabla \psi\|(p) + O(t).$$

On the other hand, since $\psi \circ \gamma_\tau$ is a convex function and $\psi \circ \gamma_\tau(\tau) - \psi \circ \gamma_\tau(0) = \tau$, we have

$$(\gamma'_\tau(0))^+(\psi) = \lim_{s \to 0^+} \frac{\psi \circ \gamma_\tau(s) - \psi \circ \gamma_\tau(0)}{s} \geq 1.$$

Put $\alpha(\tau) = \|\nabla \psi\|^{-1}\|\gamma'_\tau(0)\|^{-1}(\gamma'(0))^+(\psi)$. Since $\nabla \psi$ is the gradient direction, we know that $\alpha(\tau) \leq 1$ for any $\tau > 0$. Combining the two inequalities above gives

$$\begin{aligned} \alpha(\tau)\|\nabla \psi\|(p) &= \|\gamma'_\tau(0)\|^{-1}(\gamma'_\tau(0))^+(\psi) \\ &\geq \|\gamma'_\tau(0)\|^{-1} \geq \|\nabla \psi\|(p) + O(t). \end{aligned}$$

Hence as $t \to 0$ (or $\tau \to 0$) we shall have $\alpha(\tau) \to 1$ and $\|\gamma'_\tau(0)\| \to \|\nabla \psi\|^{-1}(p)$, which imply the lemma by the uniqueness of the gradient direction.

We defined $\varphi_p$ as a limit of the broken geodesics whose segments consist of minimal geodesics $\gamma_\tau$ as above. Therefore, the preceding lemma suggests that $\varphi_p$ might be regarded as an integral curve of the vector field $\nabla \psi / \|\nabla \psi\|^2$ which is unfortunately not differentiable (not even continuous). Actually, a more careful observation will show an even stronger result. (The proof of the next proposition is somewhat technical and we shall omit it. The proof may be found, e.g., in [4].)

**Proposition A.8.** *For any $p \in \partial C$ let $\varphi_p: [0, a_0] \to C$ be as above. For any $t \in [0, a_0)$ and any $\tau > 0$ let $\gamma_\tau: [0, \tau] \to C$ be the minimal geodesic from $\varphi_p(t) \in \partial C^t$ to $\varphi_p(t+\tau) \in \partial C^{t+\tau}$. Then*

$$\lim_{\tau \to 0^+} \gamma'_\tau(0) = \frac{\nabla \psi}{\|\nabla \psi\|^2}(\varphi_p(t)).$$

In Theorem A.5, we have shown that $\nabla \psi$ is parallel normal along any geodesic contained in $\partial C^b$, $b \in (0, a_0)$, and we know every pseudo-soul is completely contained in $\partial C^b$ for some $b$. Therefore we conclude that almost every pseudo-soul has a parallel normal vector field, namely $\nabla \psi$. In fact, we can prove this fact without the restriction on $b \in [0, a_0]$. In

[5] it was shown that if a soul is not unique, then every soul has a parallel normal vector field along it. The only nontrivial fact which the author used in the proof is that the mixed curvature terms vanish along a soul. Since we now know every pseudo-soul also has this property, we can prove the following corollary. Moreover, using the concept of a pseudo-soul and its properties, we obtain a proof simpler than Sharafutdinov's argument.

**Corollary A.9.** [5, *Theorem* 2]. *If a pseudo-soul $S$ is not a unique soul, then $\nabla\psi$ is a parallel normal vector field along $S$.*

*Proof.* For any pseudo-soul $S$, $H(S, t)$ is a continuous isometric variation of $S$ through pseudo-souls. However, according to the proposition, for each $p \in S$ we may regard $\nabla\psi(p)$ as a tangent vector of the curve $\varphi_p(t) = H(p, t)$ at $p$. Therefore $\nabla\psi$ is a variational vector field of pseudo-souls, and is therefore a global normal Jacobi field. Hence, by Corollary 2.5, $\nabla\psi$ is a parallel normal along $S$.

## References

[1] G. D. Birkhoff, *Dynamical systems,* Colloq. Publ. **9**, Amer Math. Soc., Providence, RI, 1927.

[2] J. Cheeger & D. Ebin, *Comparison theorems in Riemannian geometry*, North-Holland, North-Holland, Amsterdam, 1975.

[3] J. Cheeger & D. Gromoll *On the structure of complete manifold of nonnegative curvature*, Ann. of Math (2) **96** (1972) 413–443.

[4] V. A. Sharafutdinov, *The Pogorelov-Klingenberg theorem for manifold homeomorphic to* $\mathbf{R}^n$, Siberian Math. J. **18**, No. 4 (1977) 915–925.

[5] ——, *Convex sets in a manifold of nonnegative curvature*, Mat. Zametki **26** (1979) 129–136.

[6] M. Strake, *A splitting theorem for open nonnegatively curved manifolds*, Manuscripta Math. **61** (1988) 315–325.

[7] G. Walschap, *A splitting theorem for* 4*-dimensional manifolds of nonnegative curvature*, Proc. Amer. Math. Soc. (to appear).

[8] J.-W. Yim, *Distance nonincreasing retraction on a complete open manifold of nonnegative sectional curvature*, Ann. Global Anal. Geom. **6** (1988) 191–206.

UNIVERSITY OF PENNSYLVANIA