

Il teorema del soul

Andrea Bianchi, Alessandro Pigati

In questo lavoro presentiamo la dimostrazione del teorema del soul, risultato dei primi anni '70 dovuto a Cheeger, Gromoll e Meyer. Si tratta di un risultato che a partire da un'ipotesi sulla curvatura sezionale di una varietà porta a una conclusione sulla struttura della varietà dal punto di vista differenziale, e in ultima analisi dal punto di vista topologico. Ove non diversamente specificato, le varietà riemanniane considerate saranno sempre connesse e complete.

Per enunciare il teorema occorre la seguente definizione:

DEFINIZIONE 1. *Data una varietà riemanniana completa (M, g) , un sottoinsieme $A \subseteq M$ si dice convesso se per ogni coppia di punti $p, q \in A$ e per ogni geodetica minimizzante in M che ha p e q come estremi, si ha che il supporto della geodetica è contenuto in A .*

Un insieme $A \subseteq M$ si dice totalmente convesso se la stessa proprietà vale per ogni coppia di punti $p, q \in A$ e per ogni geodetica in M , anche non minimizzante, che ha p e q come estremi.

Possiamo ora enunciare il teorema del soul

TEOREMA 1 (Soul). *Sia (M, g) una varietà riemanniana completa e non compatta con curvatura sezionale $\sec \geq 0$. Allora esiste una sottovarietà chiusa e totalmente convessa $S \subset M$, detta soul, per cui M è diffeomorfa al fibrato normale di S in M (si intende che il diffeomorfismo fa corrispondere nel modo canonico $S \subset M$ con la 0-sezione del fibrato normale $N_M(S) \rightarrow S$).*

Se poi $\sec > 0$, allora la conclusione è più forte: S è un punto e pertanto M è diffeomorfa a \mathbb{R}^n , ove n è la dimensione di M .

La dimostrazione del teorema segue la seguente strategia:

- si individua un opportuno soul S ;
- si considera la funzione f data dalla distanza dal soul: questa funzione non sarà in generale liscia, ma in un opportuno senso sarà priva di punti critici;
- in maniera analoga alla dimostrazione del lemma di Morse, si costruirà un campo vettoriale X liscio e mai nullo su $M \setminus S$ che giochi il ruolo di ∇f : la proprietà cruciale sarà che f crescerà in modo controllato lungo le curve integrali di X ;
- Tramite il flusso di X sarà possibile isotopare M in un intorno tubolare di S , il quale è diffeomorfo al fibrato normale $N_M(S)$.

Osserviamo subito che, se il teorema è vero, allora M ha il tipo di omotopia del suo soul: in particolare l'omologia di M a coefficienti in \mathbb{Z}_2 determina la dimensione del soul S (in quanto la classe fondamentale di S è la classe di omologia a coefficienti in \mathbb{Z}_2 di M di dimensione più alta). In più l'immersione di S in M è un'equivalenza omotopica, visto che M si retrae per deformazione sul soul S : questo comporta che i soul di M sono tutti omotopicamente equivalenti dentro M .

Il primo passo importante nella dimostrazione del teorema sarà generalizzare il lemma di Morse. Al posto di una funzione di Morse liscia useremo una opportuna *funzione distanza*.

DEFINIZIONE 2. Fissato un compatto K in una varietà riemanniana M , definiamo la funzione distanza da K $f = f_K: M \rightarrow [0, +\infty)$ ponendo per $x \in M$

$$f(x) = d(x, K) = \min \{ d(x, p) \mid p \in K \}$$

ove $d(\cdot, \cdot)$ è la distanza indotta dalla metrica riemanniana in M .

Per completezza di M e compatezza di K la funzione f risulta propria; è inoltre facile verificare che f è 1-lipschitziana, pertanto f sarà differenziabile quasi ovunque, per il teorema di Rademacher.

Nei punti in cui f è differenziabile il gradiente ∇f deve essere un vettore unitario. Per verificarlo, sia x un tale punto. Da un lato per 1-Lipschitzianità $|\nabla f(x)| \leq 1$. Dall'altro consideriamo una curva di minima lunghezza che congiunga K con x : essa sarà una geodetica parametrizzata per lunghezza d'arco $\gamma: [0, f(x)] \rightarrow M$ con $\gamma(0) \in K$ e $\gamma(f(x)) = x$, e lungo γ la funzione f crescerà linearmente con velocità 1. Pertanto

$$1 = \frac{d}{dt} f(\gamma(t))|_{t=0} = \langle \nabla f(x), \dot{\gamma}(0) \rangle \leq |\nabla f(x)| \cdot |\dot{\gamma}(0)| = |\nabla f(x)|$$

Notiamo che nell'ultima disuguaglianza deve in realtà valere l'uguaglianza, perché già sappiamo che $|\nabla f(x)| \leq 1$; pertanto la geodetica γ arriva in x proprio con velocità $\nabla f(x)$.

Nei punti $x \in M$ in cui f non è differenziabile, esistono comunque una o più geodetiche di lunghezza $f(x)$ che collegano K a x . Poniamo allora, per ogni $x \in M$,

$$\Gamma(x, K) = \{ \dot{\gamma}(f(x)) \mid \gamma: [0, f(x)] \rightarrow M, \gamma \text{ è una geodetica}, \gamma(0) \in K, \gamma(f(x)) = x \}$$

Questo insieme contiene, in qualche modo, i possibili valori che ∇f potrebbe avere se esistesse: in effetti se $\nabla f(x)$ esiste allora l'insieme $\Gamma(x, K)$ contiene solo esso come elemento. Tuttavia se i vettori unitari in $\Gamma(x, K)$ sono abbastanza vicini fra loro, sarà possibile scegliere un vettore unitario medio, non necessariamente in $\Gamma(x, K)$, il quale approssimerà bene ogni vettore in $\Gamma(x, K)$ e per molti scopi sarà un buon surrogato del gradiente.

DEFINIZIONE 3. Diciamo che $x \in M$ è un punto regolare per f se l'insieme $\Gamma(x, K)$ è contenuto in una semisfera aperta della sfera unitaria in $T_x M$, ossia se esiste $v \in T_x M$ unitario per cui $\langle v, w \rangle > 0$.

Più in generale, per $0 < \alpha \leq \pi/2$, diciamo che x è α -regolare se esiste $v \in T_x M$ unitario per cui $\langle v, w \rangle > \cos \alpha$, cioè l'angolo tra v e w è minore di α . Poniamo allora

$$G_\alpha f(x) = \{ w \in T_x M \mid |w| = 1, \forall w \in \Gamma(x, K) \langle v, w \rangle > \cos \alpha \}$$

L'insieme $G_\alpha f$ contiene i vettori abbastanza vicini a ciò che potrebbe essere ∇f . Il seguente lemma garantisce l'esistenza di un campo vettoriale liscio lungo il quale f cresca in modo controllato

LEMMA 1. Siano M , K , $f = f_K$, $\Gamma(\cdot, K)$ e $G_\alpha f$ come sopra; fissiamo anche $0 < \alpha < \pi/2$. Allora

- (1) $\Gamma(x, K)$ è chiuso per ogni $x \in M$;
- (2) l'insieme dei punti α -regolari è un aperto di M ;
- (3) $G_\alpha f(x)$ è un insieme convesso della sfera unitaria di $T_x M$, per ogni $x \in M$;

- (4) se U è un aperto di M costituito da punti α -regolari, allora esiste un campo vettoriale liscio e unitario X su U con $X(x) \in G_\alpha f(x)$ per ogni $x \in U$; se X è un siffatto campo e γ è una curva integrale di X , allora per ogni $s < t$ nel dominio di γ si ha

$$f(\gamma(t)) - f(\gamma(s)) > \cos \alpha (t - s)$$

PROOF. (1) Sia $(v_i)_{i \in \mathbb{N}}$ una successione di vettori unitari in $\Gamma(x, K)$ convergente a $v \in T_x M$. Allora la successione di geodetiche $\gamma_i: [0, f(x)] \rightarrow M$ definite da $\gamma_i(t) = \exp_x((t - f(x))v_i)$ converge uniformemente alla geodetica $\gamma(t) = \exp_x((t - f(x))v)$. Poiché $\gamma_i(0) \in K$ per ipotesi sui v_i e K è chiuso, anche $\gamma(0) \in K$, dunque $v \in \Gamma(x, K)$. In modo del tutto analogo, e sfruttando la continuità di $\exp: TM \rightarrow M$ e di f , si mostra che l'insieme $\Gamma(K) \subset TM$ definito come l'unione di tutti i $\Gamma(x, K)$ al variare di $x \in M$ è chiuso.

- (2) Supponiamo per assurdo di poter scrivere un punto α -regolare x come limite di punti non α -regolari $(x_i)_{i \in \mathbb{N}}$. Sia $v \in G_\alpha f(x)$ e sia $v_i \in T_{x_i} M$ una successione di vettori unitari convergenti, in TM , a v . Per ogni i esiste un $w_i \in \Gamma(x_i, K)$ per cui $\langle v_i, w_i \rangle \leq \cos \alpha$, visto che x_i non è α -regolare. Se U è un intorno compatto di x contenente tutti gli x_i , allora i w_i si trovano tutti nel tangente unitario di U , che è un compatto di TM ; dunque a meno di prendere una sottosuccessione degli indici, $w_i \rightarrow w$ per un opportuno $w \in T_x M$. Per chiusura di $\Gamma(K)$ si ha $w \in \Gamma(x, K)$ e per continuità della metrica si ha $\langle v, w \rangle = \lim_{i \rightarrow \infty} \langle v_i, w_i \rangle \leq \cos \alpha$, contraddicendo l'ipotesi $v \in G_\alpha f(x)$.
- (3) La sfera unitaria di $T_x M$ si intende dotata della metrica indotta dall'inclusione nello spazio (euclideo) $T_x M$. Notiamo che per ogni $v \in \Gamma(x, K)$ l'insieme $C_\alpha(v) = \{w \in T_x M \mid |w| = 1, \langle v, w \rangle > \cos \alpha\}$ è un convesso della sfera unitaria di $T_x M$. Non resta che notare che $G_\alpha f(x)$ è l'intersezione di tutti questi convessi al variare di $v \in \Gamma(x, K)$.
- (4) Intanto notiamo che l'insieme $G_\alpha f \subset TM$ dato dall'unione dei $G_\alpha f(x)$ al variare di $x \in M$ è aperto nel fibrato tangente unitario. Infatti se per assurdo $w \in G_\alpha(x)$ fosse il limite della successione di vettori unitari $w_i \in T_{x_i} M \setminus G_\alpha f(x_i)$, potremmo individuare una successione di vettori unitari $v_i \in \Gamma(x_i, K)$ per cui $\langle v_i, w_i \rangle \leq \cos \alpha$. A meno di sottosuccessioni $w_i \rightarrow w$ per un opportuno $w \in \Gamma(x, K)$ e passando al limite la disuguaglianza si avrebbe $\langle v, w \rangle \leq \cos \alpha$, contraddicendo il fatto che $w \in G_\alpha f(x)$.

Per ogni $p \in U$ possiamo allora scegliere un vettore $v_p \in G_\alpha f(p)$ ed estenderlo in un intorno U_p di p a un campo liscio di vettori unitari: per quanto appena detto questa sezione locale del fibrato tangente unitario assumerà valori in $G_\alpha f$ in un opportuno intorno di p , $U'_p \subseteq U_p$. Gli aperti U'_p formano un ricoprimento di U , e possiamo estrarne un raffinamento localmente finito $(U_i)_{i \in \mathbb{N}}$, con la proprietà che sull'aperto U_i è definito un campo unitario v_i a valori in $G_\alpha f$.

Scegliamo una partizione dell'unità liscia $(\lambda_i)_{i \in \mathbb{N}}$ subordinata al ricoprimento (U_i) e definiamo il campo vettoriale V tramite la formula

$$V(x) = \sum_{i \in \mathbb{N}} \lambda_i v_i$$

Il campo V è liscio in quanto localmente la serie utilizzata per definirlo ha finiti addendi non nulli; inoltre $V(x)$ non si annulla per alcun $x \in M$, in quanto scelto un qualsiasi $v \in \Gamma(x, K)$ si ha $\langle v, V \rangle = \sum_{i \in \mathbb{N}} \lambda_i \langle v, v_i \rangle = \sum_{i \in \mathbb{N}} \lambda_i \cos \alpha = \cos \alpha \geq 0$. Possiamo allora normalizzare V e ottenere un campo unitario X che soddisfa $X(x) \in G_\alpha f(x)$ grazie alla proprietà di convessità di $G_\alpha f(x)$.

Resta ora da verificare la proprietà di crescita controllata di $f \circ \gamma$ per una curva integrale $\gamma: I \rightarrow U$ del campo X . Poiché f e γ sono 1-lipschitziane anche $f \circ \gamma: I \rightarrow [0, +\infty)$ è lipschitziana, dunque $f \circ \gamma$ è quasi ovunque derivabile e $f \circ \gamma$ coincide a meno di una costante con l'integrale della sua derivata. Basterà allora dimostrare che nei tempi $t_0 \in I$ in cui $f \circ \gamma$ è derivabile, si ha $\frac{d}{dt}(f \circ \gamma)(t)|_{t=t_0} > \cos \alpha$.

Sia $x_0 = \gamma(t_0)$, sia $l = f(x_0)$ e sia $\bar{\gamma}: I \times [0, l] \rightarrow M$ una variazione liscia di γ che soddisfi le seguenti proprietà:

- $\bar{\gamma}(\cdot, l) = \gamma$ come curva da I in M ;
- $\bar{\gamma}(t_0, \cdot): [0, l] \rightarrow M$ è un segmento geodetico minimizzante da K a x_0 ;
- $\bar{\gamma}(\cdot, 0): I \rightarrow M$ è una curva costante su un punto $p \in K$

Indicheremo con $t \in I$ la prima coordinata e con $s \in [0, l]$ la seconda. Per ogni $t \in I$ si ha

$$\begin{aligned} \frac{1}{2}(f \circ \bar{\gamma})^2(t, l) &\leq \frac{1}{2} \int_0^l \left| \frac{\partial \bar{\gamma}}{\partial s}(t, s) \right|^2 ds \\ &\leq \frac{l}{2} \int_0^l \left| \frac{\partial \bar{\gamma}}{\partial s}(t, s) \right|^2 ds \\ &= lE(\bar{\gamma}(t, \cdot)) \\ &= lE(\bar{\gamma}_t) \end{aligned}$$

ove la prima disuguaglianza segue dalla 1-lipschitzianità di f e dal fatto che $f(\bar{\gamma}(t, 0)) = 0$ e la seconda è la disuguaglianza di Cauchy-Schwarz. Per $t = t_0$ si ha l'uguaglianza in quanto $\bar{\gamma}(t_0, \cdot)$ è una geodetica minimizzante e parametrizzata per lunghezza d'arco. Le due funzioni $\frac{1}{2}(f \circ \bar{\gamma}(t, l))^2$ e $lE(\bar{\gamma}_t)$ definite per $t \in I$ e a valori in $[0, +\infty)$ sono derivabili in $t = t_0$. Poiché una delle due è sempre minore o uguale dell'altra e in t_0 vale l'uguaglianza, necessariamente le due derivate

in t_0 devono coincidere. Pertanto

$$\begin{aligned}
l \frac{d}{dt}(f \circ \gamma)(t_0) &= \frac{d}{dt} \frac{1}{2}(f \circ \gamma(t, l))^2 \Big|_{t=t_0} \\
&= \frac{d}{dt} l E(\bar{\gamma}_t) \Big|_{t=t_0} \\
&= l \int_0^l \left\langle \bar{\nabla}_{\frac{\partial}{\partial t}} \frac{\partial \bar{\gamma}}{\partial s}, \frac{\partial \bar{\gamma}}{\partial s}(t, s) \right\rangle ds \Big|_{t=t_0} \\
&= l \int_0^l \left\langle \bar{\nabla}_{\frac{\partial}{\partial s}} \frac{\partial \bar{\gamma}}{\partial t}, \frac{\partial \bar{\gamma}}{\partial s}(t, s) \right\rangle ds \Big|_{t=t_0} \\
&= l \int_0^l \frac{\partial}{\partial s} \left\langle \frac{\partial \bar{\gamma}}{\partial t}, \frac{\partial \bar{\gamma}}{\partial s}(t, s) \right\rangle ds \Big|_{t=t_0} \\
&= l \left\langle \frac{\partial \bar{\gamma}}{\partial t}, \frac{\partial \bar{\gamma}}{\partial s}(t_0, s) \right\rangle \Big|_{s=0}^{s=l} \\
&= l \left\langle \frac{\partial}{\partial t} \gamma(t_0), \frac{\partial}{\partial s} \bar{\gamma}(t_0, l) \right\rangle \\
&> l \cos \alpha
\end{aligned}$$

La prima uguaglianza segue da $f(\gamma(t_0)) = l$. La seconda segue da quanto visto sopra. Dalla terza uguaglianza in poi sfruttiamo il pullback del fibrato TM tramite $\bar{\gamma}$ a un fibrato $\bar{\gamma}^*TM$ su $I \times [0, l]$. Oltre al fibrato, è possibile effettuare il pullback anche della connessione di Levi-Civita di M a una connessione $\bar{\nabla}$ compatibile con il pullback della metrica e con una proprietà analoga alla simmetria: dati due campi $X, Y \in T(I \times [0, l])$ vale l'uguaglianza

$$\bar{\nabla}_X(\bar{\gamma}_*Y) - \bar{\nabla}_Y(\bar{\gamma}_*X) - \bar{\gamma}_*[X, Y] = 0$$

ove la parentesi di Lie è relativa alla varietà $I \times [0, l]$ e la mappa $\bar{\gamma}_*$ manda $T(I \times [0, l])$ in $\bar{\gamma}^*TM$. La terza uguaglianza segue immediatamente. La quarta sfrutta il fatto che i campi $\frac{\partial}{\partial s}$ e $\frac{\partial}{\partial t}$ a valori in $T(I \times [0, l])$ commutano. La quinta uguaglianza segue dal fatto che $\bar{\gamma}(t_0, \cdot)$ è una geodetica, per cui $\bar{\nabla}_{\frac{\partial}{\partial s}} \frac{\partial \bar{\gamma}}{\partial s}(t_0, s) = 0$. La sesta uguaglianza è immediata. La settima segue dall'ipotesi $\bar{\gamma}(t, 0)$ costante in t e $\bar{\gamma}(t, l) = \gamma(t)$. La disuguaglianza finale segue dal fatto che $\frac{\partial}{\partial t} \gamma(t_0) = X(x_0) \in G_\alpha f(x_0)$, mentre $\frac{\partial \bar{\gamma}}{\partial s}(t_0, l) \in \Gamma(x_0, K)$. Basta ora dividere per l e si ha la tesi. \square

Possiamo ora dimostrare un lemma di deformazione simile al lemma di Morse

LEMMA 2 (Deformazione). *Siano $M, K, f = f_K$ come sopra, fissiamo $0 < a < b$ e supponiamo che per qualche $\alpha < \pi/2$ tutti i punti $x \in f^{-1}([a, b])$ siano α -regolari. Allora $f^{-1}([0, b])$ e $f^{-1}([0, a])$ sono omeomorfi e il primo si retrae per deformazione sul secondo.*

PROOF. Poiché la funzione f è propria e per il lemma 1 l'insieme dei punti α -regolari è aperto, dunque esiste $\varepsilon > 0$ per cui $f^{-1}([a - \varepsilon, b])$ è costituito solo di punti α -regolari. Sempre in virtù del lemma 1 possiamo individuare un campo unitario liscio X su un intorno di $f^{-1}([a - \varepsilon, b])$ a valori in $G_\alpha f$. Scegliamo una funzione liscia $\lambda: [0, b] \rightarrow [0, 1]$ con $\lambda(t) = 0$ per $t \leq a - 2\varepsilon/3$ e $\lambda(t) = 1$ per $t \geq a - \varepsilon/3$. Il campo $Y(x) = \lambda(f(x))X(x)$ si estende allora a un campo liscio su tutto $f^{-1}([0, b])$, ponendolo nullo ove non era già definito. Per il lemma 1 la funzione f cresce lungo le linee integrali di Y con velocità maggiore di $\cos \alpha$ almeno finché ci si trova in $f^{-1}([a - \varepsilon/3, b])$, ove Y coincide con il vecchio X .

Se chiamiamo F^t il flusso del campo $-Y$ al tempo t , è facile verificare, sempre grazie al lemma 1, che esso è ben definito per ogni tempo $t \geq 0$ su $f^{-1}([0, b])$ e che per tempi maggiori di $\frac{b-a+\varepsilon/3}{\cos \alpha}$ l'immagine è interamente contenuta in $f^{-1}([0, a - \varepsilon/3])$.

Definiamo allora per $x \in f^{-1}([0, b])$

$$\tau(x) = \min \{ t \geq 0 \mid F^t(x) \in f^{-1}([0, a - \varepsilon/3]) \}.$$

Per $x \in f^{-1}([a - \varepsilon/3, b])$ definiamo le funzioni σ e β . Poniamo

$$\sigma(x) = \min \{ t \geq 0 \mid F^{-t} \text{ è definito e } \in f^{-1}(b) \};$$

anche σ sarà ben definita e limitata da $\frac{b-a+\varepsilon/3}{\cos \alpha}$; poniamo poi $\beta(x) = F^{-\sigma(x)}(x)$.

Per $x \in f^{-1}([a, b])$ definiamo infine

$$\rho(x) = \min \{ t \geq 0 \mid F^t(x) \in f^{-1}(a) \},$$

che è ancora ben definito e limitato da $\frac{b-a}{\cos \alpha}$. L'omeomorfismo $\phi: f^{-1}([0, b]) \rightarrow f^{-1}([0, a])$ sarà ottenuto seguendo le linee del campo $-Y$ per un opportuno tempo $\phi(x)$, in modo che l'insieme $f^{-1}([0, a - \varepsilon/3])$ resti fisso, mentre l'insieme $f^{-1}([a - \varepsilon/3, b])$ sia mappato omeomorficamente in $f^{-1}([a - \varepsilon/3, a])$.

Basta porre $\phi(x) = 0$ per $f(x) \leq a - \varepsilon$, e $\phi(x) = \frac{\rho(\beta(x))}{\tau(\beta(x))}\tau(x)$ per $f(x) \geq a - \varepsilon/3$. L'omeomorfismo $f^{-1}([0, b]) \rightarrow f^{-1}([0, a])$ sarà dunque $x \mapsto F^{\phi(x)}(x)$.

La retrazione invece è data seguendo il flusso di Y finché non si raggiunge l'insieme $f^{-1}([0, a])$, quindi fermandosi: in simboli $x \mapsto x$ se $f(x) \leq a$ e $x \mapsto F^{\rho(x)}(x)$ se $f(x) \geq a$, \square

Adattando la dimostrazione precedente si ottiene la seguente versione liscia del lemma di deformazione:

COROLLARIO 1. *Se K è una sottovarietà compatta e ogni punto in $M \setminus K$ è regolare per $f_K = d(\cdot, K)$ allora M è diffeomorfa al fibrato normale NK di K . In particolare, se K è un punto allora $M \cong \mathbb{R}^n$.*

A questo punto inizia la nostra ricerca di un opportuno soul dentro una varietà M con $\text{sec} \geq 0$. Visto che il soul deve essere totalmente convesso, tornerà utile studiare le proprietà generali degli insiemi totalmente convessi e trovare dei metodi per costruire siffatti insiemi.

1. Lemmi preliminari sugli insiemi totalmente convessi

Il seguente lemma dà una descrizione della struttura di un insieme totalmente convesso

LEMMA 3. *Sia A un insieme chiuso e totalmente convesso dentro una n -varietà riemanniana M .*

Allora A è la chiusura di una opportuna sottovarietà di M di dimensione k per un opportuno $0 \leq k \leq n$. Chiamiamo $\overset{\circ}{A}$ questa sottovarietà e $\partial A = A \setminus \overset{\circ}{A}$ il

bordo di A . Allora anche $\overset{\circ}{A}$ è totalmente convessa e per ogni $x \in \partial A$ esiste un vettore unitario $v \in T_x M$ che punta verso l'interno di A , ossia per ogni geodetica $\gamma: [0, l] \rightarrow A$ con $\gamma(0) = x$ e $\gamma(l) \in \overset{\circ}{A}$ si ha $\langle v, \dot{\gamma}(0) \rangle > 0$.

PROOF. Useremo un risultato sulla stima del raggio di convessità: per la precisione, esiste una funzione $\varepsilon: M \rightarrow (0, +\infty)$ per cui per ogni $p \in M$ la funzione $d(\cdot, p)$ è liscia e strettamente convessa sulla palla $B(p, \varepsilon(p))$.

Sia k il massimo intero per cui A contiene sottovarietà di dimensione k . Se $k = 0$ allora A è un punto, altrimenti dati due punti in A una qualsiasi geodetica minimale che li congiunga sarebbe una 1-varietà contenuta in A . Sia N l'unione di tutte le k -sottovarietà contenute in A . Vogliamo dimostrare che A è la chiusura di N , e che N è una k -sottovarietà totalmente convessa di M .

Intanto dimostriamo che N è una k -sottovarietà di M . Sia $p \in N$, sia $N_p \subseteq A$ una k -sottovarietà di M contenente p . Fissiamo δ abbastanza piccolo per cui: $B(p, 2\delta)$ è convesso in M (dunque basta $2\delta < \varepsilon(p)$); per ogni punto $q \in B(p, \delta)$ si abbia che il raggio di iniettività dell'esponenziale in q è almeno 3δ (qui si sfrutta la continuità del raggio di iniettività e il fatto che esso è positivo in p); $N_p \cap B(p, 2\delta)$ è una sottovarietà chiusa in $B(p, 2\delta)$. Ridefiniamo N_p come l'intersezione di N_p con $B(p, 2\delta)$.

Vogliamo dimostrare che $A \cap B(p, \delta) = N_p \cap B(p, \delta)$, per concludere che N è una k -sottovarietà di M . Sia per assurdo $q \in A \cap B(p, \delta) \setminus N_p$. Per il punto q esiste almeno un punto $q' \in N_p$ a minima distanza: infatti p dista meno di δ da q , e su $(N_p \cap \overline{B(q, \delta)}) \subseteq B(p, 2\delta)$ esiste il minimo per compattezza.

Sia q' questo minimo e colleghiamo q i segmenti geodetici minimizzanti (interni a $B(p, 2\delta)$ che è convessa) ai punti di N_p in un intorno di q' : il segmento che arriva in q' è ortogonale a N_p , per cui anche i segmenti geodetici che collegano q con un intorno di q' sono trasversi a N_p . Ne deriva che l'unione di questi segmenti è vicino a q' l'immagine di un embedding di una $k + 1$ -varietà, la quale deve essere però contenuta in A , assurdo.

Vogliamo ora dimostrare che se $\gamma: [0, l] \rightarrow A$ è una geodetica con $\gamma(0) \in N$, allora $\gamma: [0, l] \rightarrow N$: da qui seguirebbe sia che N è denso in A (ogni punto di A può essere collegato con una geodetica a un punto di N), sia che N è totalmente convesso. Poiché il supporto di γ è compatto, il raggio di iniettività relativo alla varietà M è positivamente limitato dal basso lungo γ . È possibile allora individuare dei tempi $0 = t_1 < t_2 < \dots < t_h = l$ per cui γ non congiunge i punti $\gamma(t_i)$ e $\gamma(t_{i+2})$: se riusciamo a dimostrare la tesi per

una geodetica che non coniuga gli estremi, avremmo induttivamente che γ manda in N gli intervalli $[0, t_2)$, $[t_1, t_3)$, e così via fino all'ultimo.

Supponiamo dunque che $\gamma(0)$ e $\gamma(l)$ non siano coniugati lungo γ . Consideriamo una $k - 1$ -sottovarietà S contenuta in N e trasversa a γ nel punto $\gamma(0)$. La mappa $\exp_{\gamma(0)}$ è una carta intorno a $\gamma(0)$, per cui la geodetica γ sarà affiancata da altrettante geodetiche, trasverse a S , in un intorno di $\gamma(0)$. Queste geodetiche coprono una k -varietà, dunque un aperto di A , contenente tutta la parte interna della geodetica γ .

Vogliamo ora dimostrare che per ogni $p \in \partial A$ c'è un *vettore entrante* in A .

Intanto notiamo che N è totalmente geodetica, ossia le geodetiche di N sono anche geodetiche di M : questo segue dal fatto che N è totalmente convessa. Una conseguenza di questo fatto è che il trasporto parallelo in M lungo curve di N manda vettori tangenti a N in vettori tangenti a N . È dunque ben definito lo spazio tangente $T_x A$ anche nei punti $x \in \partial A$.

Notiamo ora che se la proposizione è vera per punti in un insieme denso di ∂A , allora è vera per ogni punto $p \in \partial A$: basta approssimare p con una successione $(p_i)_{i \in \mathbb{N}}$ di punti del bordo per cui la tesi è vera, scegliere per ognuno di essi un vettore entrante v_i , far convergere a meno di sottosuccessioni v_i a un $v \in T_p M$. A questo punto fissata una geodetica γ uscente da p e a valori in N per tempi positivi, si fissa un punto p' su di essa abbastanza vicino a p perché il raggio di iniettività dell'esponenziale di M centrato in p' sia maggiore della distanza da p , e si collega p' con i p_i tramite le geodetiche minimizzanti.

La successione dei vettori velocità in p_i di queste geodetiche converge alla velocità della geodetica γ in p . Passando al limite il prodotto scalare si ottiene $\langle \dot{\gamma}(0), v \rangle \geq 0$. Per ottenere il maggiore stretto è sufficiente notare che, grazie all'argomento della costruzione della $k - 1$ -sottovarietà trasversa a una geodetica e del relativo cono geodetico centrato in p , l'insieme dei vettori velocità iniziali di geodetiche uscenti da p e a valori in N per tempi positivi, è un aperto di $T_x A$.

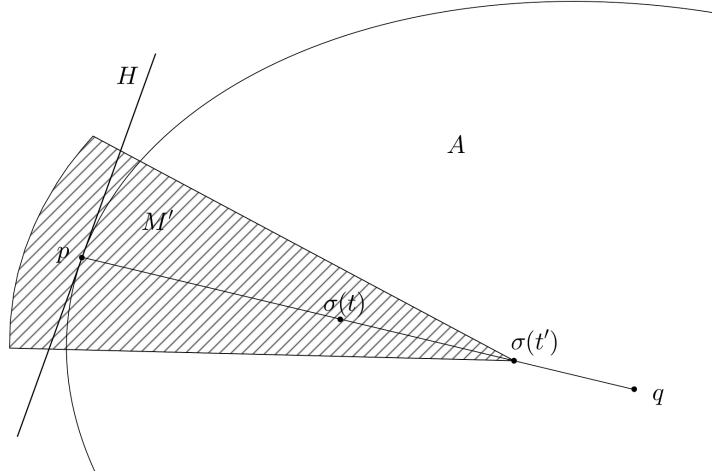
Consideriamo ora i punti p per cui esiste $q \in N$ per cui $d(q, \partial A) = d(q, p) = \varepsilon$, e supponiamo anche che q sia abbastanza vicino a p per cui il raggio di iniettività dell'esponenziale in q è maggiore di $d(q, p)$. Allora in A la palla $B_A(q, \varepsilon)$ è una varietà liscia diffeomorfa a una palla euclidea, e la geodetica minimale γ da p a q parte da p con una velocità v per cui esistono, per tempi piccoli, geodetiche a valori in $B_A(q, \varepsilon)$ salvo che per il punto di partenza p . L'insieme dei vettori velocità di queste geodetiche copre tutta la parte interna di un semispazio di $T_p A$: per la precisione si tratta di tutti i vettori unitari di $T_p A$ che hanno prodotto scalare positivo con $\dot{\gamma}(0)$: ciò segue dal lemma di Gauss.

Se per assurdo esistesse una geodetica σ per cui $\sigma(0) = p$, σ ha valori in N per tempi positivi piccoli e $\langle \dot{\sigma}(0), \dot{\gamma}(0) \rangle \leq 0$ (si tratta delle due velocità in p), allora anche per tempi negativi σ avrebbe valori in N , coincidendo con la geodetica uscente da p e velocità iniziale $-\dot{\sigma}(0)$. Di conseguenza p sarebbe contenuto all'interno di una geodetica con estremi in N , e pertanto avremmo $p \in N$, contro l'ipotesi $p \in \partial A$. \square

LEMMA 4. *Supponiamo $\text{sec} \geq 0$. Dato un A compatto e totalmente geodetico con $\partial A \neq \emptyset$, la funzione $d(\cdot, \partial A)$ è concava n.s.b. su $\text{int } A$. Se $\text{sec} > 0$ allora è strettamente concava n.s.b. lungo direzioni non radiali.*

Qui diciamo che $v \in T_p(\text{int } A) \setminus \{0\}$ è *non radiale* se esiste una geodetica σ da p a ∂A di lunghezza minima con $\langle \dot{\sigma}(0) \rangle \neq \langle v \rangle$.

PROOF. Dato $q \in \text{int } A$ sia $p \in \partial A$ tale che $d(q, p) = d(q, \partial A)$. Sia $\sigma : [0, a] \rightarrow A$ una geodetica minimale da p a q : σ è tutta contenuta in A per convessità. Scegliamo $0 < t < t' < a$ e osserviamo che $p \notin \text{Cut}_{\sigma(t')}$. Siano poi $H := \exp_p(B(0, \text{inrad}(p)) \cap \dot{\sigma}(0)^\perp)$ e M' il cono di geodetiche $\bigcup_v \{ \exp_{\sigma(t')}(\alpha v) : 0 < \alpha < t' + \epsilon \}$ con $v \in T_{\sigma(t')}$ unitario che forma con $-\dot{\sigma}(t')$ un angolo $< \epsilon$ (vedi figura). H e M' sono sottovarietà embedded. Inoltre $M' \cap \text{int } A$, essendo aperto in $\text{int } A$, è totalmente geodetico: quindi qui vale ancora $\text{sec} \geq 0$ con la metrica indotta.



Lavoriamo in M' ; poniamo $H' := H \cap M'$ e osserviamo che per ϵ piccolo essa è una sottovarietà embedded di M' . $f := d(\cdot, H')$ (la distanza è quella associata alla metrica su M') è una barriera per $d(\cdot, \partial A)$ in $\sigma(t)$: per ogni $x \in M' \cap \text{int } A$ è $f(x) \geq d(x, \partial A)$ perché una curva da x ad H' in M' interseca necessariamente ∂A ; inoltre $d(\sigma(t), \partial A) = t \geq d(\sigma(t), H')$, da cui $d(\sigma(t), \partial A) \geq d(\sigma(t), H')$. Infine $d(\cdot, H')$ è liscia vicino a $\sigma(t)$: ciò segue dal fatto che $\sigma(t)$ è interno a una geodetica minimizzante da $\sigma(t')$ ad H' e si può mostrare generalizzando gli argomenti già visti nel caso della distanza da un punto.

Preso ora $v \in T_{\sigma(t)}(\text{int } A)$, v non parallelo a $\dot{\sigma}(t)$, estendiamolo a un campo parallelo E lungo σ . Ora la prima equazione fondamentale (vedi la Proposizione 1) dà

$$\frac{d}{dt} \text{Hess } f(E, E) = -R(\dot{\sigma}, E, \dot{\sigma}, E) - \text{Hess}^2 f(E, E) \leq 0$$

e $\text{Hess } f(E, E) = 0$ in $t = 0$ ¹, quindi $\text{Hess } f(v, v) \leq 0$ e anzi $\text{Hess } f(v, v) < 0$ se $\text{sec} > 0$. Questo mostra la tesi in $\sigma(t)$.

¹infatti, detto S l'operatore forma associato alla distanza con segno da H' , è $S(\dot{\sigma}(0)) = \nabla_{\dot{\sigma}(0)} \dot{\sigma} = 0$, da cui $\text{Hess } f(\dot{\sigma}(0), \cdot) = 0$ e per bilinearità ci riduciamo a mostrare che $\text{Hess } f(v, v) = 0$ per $v \perp \dot{\sigma}(0)$. Se γ è una curva in H' con $\gamma(0) = p$ e $\dot{\gamma}(0) = v$, vale $0 = (f \circ \gamma)''(0) = \text{Hess } f(v, v) + df(\nabla_{\dot{\gamma}(0)} \dot{\gamma}) = \text{Hess } f(v, v) + g(\dot{\sigma}(0), \nabla_{\dot{\gamma}(0)} \dot{\gamma})$ e l'ultimo termine è pari a $g(\dot{\sigma}(0), \nabla_{\dot{\gamma}(0)}^M \dot{\gamma}) = 0$: questo perché $\nabla_{\dot{\gamma}(0)}^M \dot{\gamma} \in T_p H$ dato che H in coordinate polari (in p) è un iperpiano

Torniamo a ragionare in $\text{int } A$. Poiché t è arbitrario, possiamo sceglierlo molto vicino ad a in modo che $d(\cdot, H_t)$ sia liscia vicino a q , dove $H_t := \exp_{\sigma(t)}(B_\delta \cap \dot{\sigma}(t)^\perp)$ per qualche $\delta > 0$ opportuno². Posto $f_t := d(\cdot, H_t)$, notiamo che $t + f_t$ è una barriera per $d(\cdot, \partial A)$ in q : le due funzioni coincidono in q ; se poi x è vicino a q chiamiamo $z \in \tilde{H}$ il punto a distanza minima. Vale

$$d(x, \partial A) \leq d(x, z) + d(z, \partial A) = f_t(x) + d(z, \partial A) \leq f_t(x) + f(z)$$

e basta mostrare che $f(z) \leq t$. Ma f è liscia e concava vicino a $\sigma(t)$ (perché i punti vicini a $\sigma(t)$ sono anch'essi interni a geodetiche minimizzanti da qualche punto di M' ad H'), quindi considerando la geodetica minimizzante $\gamma : [0, \bar{s}] \rightarrow \text{int } A$ da $\sigma(t)$ a z otteniamo

$$f \circ \gamma(\bar{s}) \leq f \circ \gamma(0) = f(\sigma(t)) = t,$$

dato che $f \circ \gamma$ è concava con $(f \circ \gamma)'(0) = g(\dot{\sigma}(t), \dot{\gamma}(0)) = 0$ e perciò ha massimo in 0. Ora il fatto che $\text{Hess } f_t(q) \leq 0$ (e il relativo rafforzamento se $\text{sec} > 0$) si ottengono esattamente come prima, osservando che di nuovo è $(\text{Hess } f_t)_{\sigma(t)} = 0$. \square

2. Dimostrazione del teorema del soul

Ricordiamo che una funzione continua $f : \Omega \rightarrow \mathbb{R}$ ($\Omega \subseteq M$ aperto) è concava *nel senso delle barriere (n.s.b.)* se per ogni $p \in \Omega$ e ogni $\epsilon > 0$ esiste una $f_\epsilon \in C^2(U, \mathbb{R})$ (con $U \subseteq \Omega$ intorno di p che può dipendere anche da ϵ) tale che $f_\epsilon(p) = f(p)$, $f_\epsilon \geq f$ e $\text{Hess } f_\epsilon(p) \leq \epsilon g_p$.

Diciamo che una funzione continua $f : \Omega \rightarrow \mathbb{R}$ ($\Omega \subseteq M$ aperto) è concava *nel senso delle geodetiche (n.s.g.)* se per ogni tratto di geodetica $\gamma : I \rightarrow \Omega$ (parametrizzato per lunghezza d'arco) $f \circ \gamma : I \rightarrow \mathbb{R}$ è una funzione concava.

Nell'Appendice B dimostriamo l'equivalenza tra le due nozioni, che nel seguito useremo tacitamente.

Fissiamo un qualsiasi punto $p \in M$ e consideriamo la collezione $\{\gamma_\alpha\}$ di tutti i *raggi* uscenti da p , non vuota essendo M non compatta. L'ipotesi $\text{sec} \geq 0$ implica che, per ogni α , la funzione di Busemann $b_{\gamma_\alpha} : M \rightarrow \mathbb{R}$ è concava n.s.b. e quindi anche n.s.g., grazie a questo

LEMMA 5. *Se $\text{sec} \geq 0$ e b_γ è la funzione di Busemann associata a un raggio $\gamma : [0, \infty) \rightarrow M$, vale $\text{Hess } b_\gamma \leq 0$ (n.s.b.), ovvero b_γ è concava n.s.b.*

PROOF. Sappiamo che per ogni $q \in M$ vale $\text{Hess } d_q(x) \leq \frac{g}{d_p(x)}$ su tutta M (n.s.b.). Fissato p consideriamo un asintoto $\tilde{\gamma}$ uscente da p ; ricordiamo che $b_\gamma \leq b_\gamma(p) + b_{\tilde{\gamma}}$ con uguaglianza in p , per cui $b_\gamma(p) + b_{\tilde{\gamma}}$ è una barriera per b_γ in p (non necessariamente C^2). Perciò, essendo $b_{\tilde{\gamma}}(x) \leq d(x, \tilde{\gamma}(t)) - t$, $h_t(x) := b_\gamma(p) + d(x, \tilde{\gamma}(t)) - t$ è una barriera per b_γ in p (liscia vicino a p) con $\text{Hess } h_t(p) \leq \frac{g}{d_{\tilde{\gamma}(t)}(p)} = \frac{g}{t}$. Poiché t è arbitrariamente grande arriviamo alla tesi. \square

²chiamando ν un campo normale ad H_t , basta fare in modo che $\{\exp_y(u\nu) \mid y \in H_t, |u| < \epsilon\}$ sia un intorno tubolare di H_t per ogni t vicino ad a (a quel punto, se $a - t \ll \epsilon$, q starà in questo intorno e varrà $d(q, H_t) = a - t$); ciò si ottiene facilmente tramite il teorema delle funzioni implicite

Perciò anche

$$f := \inf_{\alpha} b_{\gamma_{\alpha}},$$

è concava n.s.g., dato che un estremo inferiore di funzioni concave su un intervallo di \mathbb{R} è a sua volta concavo. Inoltre f ha valori reali ed è 1-Lipschitz, poiché tutte le $b_{\gamma_{\alpha}}$ lo sono e soddisfano $b_{\gamma_{\alpha}}(p) = 0$.

Mostriamo ora che i sopralivelli $\{f \geq a\}$ sono compatti (da cui segue in particolare che f è propria). Sia per assurdo $A := \{f \geq a\}$ non compatto; possiamo assumere $a \leq 0$, da cui $p \in A$. Prendiamo una qualsiasi successione (p_n) di punti di A tale che $d(p, p_n) \rightarrow \infty$ e consideriamo geodetiche minimali σ_n da p a p_n : queste sono contenute in A (A è totalmente geodetico per concavità di f). Una loro sottosuccessione converge a una geodetica $\gamma : [0, \infty) \rightarrow M$, che è pure minimizzante tra ogni coppia di punti che le appartengono; γ è cioè un raggio uscente da p . Ma allora $f(\gamma(t)) \leq b_{\gamma}(\gamma(t)) = -t \rightarrow -\infty$ per $t \rightarrow +\infty$, perciò per t grande $\gamma(t) \notin A$. Questo è assurdo essendo γ interamente contenuto in A (grazie al fatto che A è chiuso).

Da ciò segue che b assume un valore massimo. Un buon candidato per il soul è $C_1 := \{x : f(x) = \max f\}$. Osserviamo che C_1 è compatto e totalmente convesso, però potrebbe avere un bordo non vuoto.

- Trattiamo prima il caso dell'ipotesi $\text{sec} \geq 0$. Seguiamo il seguente algoritmo: se C_1 è una sottovarietà senza bordo ci fermiamo. Altrimenti consideriamo $d(\cdot, \partial C_1)$ su $\text{int } C_1$: questa è ancora concava per il Lemma 4, quindi l'insieme C_2 dei punti che realizzano il massimo di $d(\cdot, \partial C_1)$ è a sua volta compatto e totalmente convesso. Se C_2 non ha bordo ci fermiamo, altrimenti iteriamo il procedimento. Questo algoritmo si ferma in un numero finito di passi, essendo $\dim C_{i+1} < \dim C_i$: se ciò non valesse per un certo i , $\text{int } C_{i+1}$ sarebbe aperto in C_i ; presi allora un $p \in \text{int } C_{i+1}$ e una geodetica γ da p a ∂C_i di lunghezza minima, avremmo $d(\gamma(t), \partial C_i) = d(p, \partial C_i)$ per $\gamma(t)$ vicino a p , assurdo perché lungo γ $d(\cdot, \partial C_i)$ decresce linearmente.
- Vediamo ora il caso $\text{sec} > 0$. Vogliamo far vedere che C_1 è un solo punto. Sia $\bar{a} := \max f$ e prendiamo un qualsiasi $a < \bar{a}$. Poniamo $A := \{x : f(x) \geq a\}$ e osserviamo che $f \equiv a$ su ∂A (in quanto $\partial A \subseteq \overline{M \setminus A}$, che segue ad esempio dalla proprietà dell'iperpiano di supporto). Inoltre $f(x) = a + d(x, \partial A)$ per ogni $x \in A$: sicuramente $f(x) \leq a + d(x, \partial A)$ in quanto f è 1-Lipschitz; d'altro canto per ogni $\epsilon > 0$ è $f(x) \leq b_{\gamma_{\alpha}}(x) + \epsilon$ per qualche α e, detto $\tilde{\gamma}$ un asintoto di γ_{α} uscente da x , vale

$$f(\tilde{\gamma}(t)) \leq b_{\gamma_{\alpha}}(\tilde{\gamma}(t)) = b_{\gamma_{\alpha}}(x) - t \leq f(x) + \epsilon - t.$$

Se $T := \min \{t \geq 0 : \tilde{\gamma}(t) \in \partial A\}$ abbiamo $T \geq d(x, \partial A)$, da cui

$$a = f(\tilde{\gamma}(T)) \leq f(x) + \epsilon - T \leq f(x) + \epsilon - d(x, \partial A)$$

e mandando $\epsilon \rightarrow 0$ deduciamo $f(x) \geq a + d(x, \partial A)$. Da ciò abbiamo che C_1 è l'insieme dei punti a distanza massima da ∂A (in $\text{int } A$). Se ora esistessero $x, y \in C_1$ distinti, presa una geodetica minimale σ da x a y , la funzione $h(t) := d(\sigma(t), \partial A)$ sarebbe costante (totale convessità di C_1). Ma per ogni t intermedio $\dot{\sigma}(t)$ sarebbe non radiale (perché $d(\cdot, \partial A)$ non decresce lungo σ), da cui per il

Lemma 4 $\text{Hess } d(\cdot, \partial A)(\dot{\sigma}(t), \dot{\sigma}(t)) < 0$ n.s.b. e infine $h''(t) < 0$ n.s.b., che dà $h(s) \leq h(t) + c(s-t) - \epsilon(s-t)^2$ per qualche $\epsilon > 0$ (in un intorno di t), assurdo per la costanza di h . Perciò C_1 è un solo punto.

Nel primo caso sia C_m l'ultimo insieme costruito dall'algoritmo; poniamo $m := 1$ nel secondo caso. In entrambi i casi C_m è una sottovarietà compatta senza bordo. Per concludere la dimostrazione basta verificare che sia soddisfatta l'ipotesi del lemma di deformazione (Corollario 1), cioè che ogni $x \notin C_m$ sia regolare per $d(\cdot, C_m)$. Grazie alla proprietà dell'iperpiano di supporto, basta anzi mostrare che esiste un A chiuso totalmente convesso tale che $x \in \partial A$ e $C_m \subseteq \text{int } A$: a quel punto ogni geodetica da x a C_m di lunghezza minima sarà contenuta in A e il suo vettore velocità in x formerà un angolo $< \frac{\pi}{2}$ con il vettore v dell'enunciato del Lemma 3. Essendo $C_0 := M \supset C_1 \supset \dots \supset C_m$, c'è un $0 \leq i < m$ tale che $x \in \text{int } C_i \setminus \text{int } C_{i+1}$. Se $x \in \partial C_{i+1}$ scegliamo $A := C_{i+1}$ (è $i+1 < m$) e abbiamo finito. Altrimenti $x \in \text{int } C_i \setminus C_{i+1}$ e scegliamo $A := \{y \in \text{int } C_i : d(y, \partial C_i) \geq d(x, \partial C_i)\}$ (oppure $A := \{y : f(y) \geq f(x)\}$ se $i = 0$). Chiaramente $x \in A$ e non può essere $x \in \text{int } A$, altrimenti per la concavità di $d(\cdot, \partial C_i)$ (o di f) n.s.g. sarebbe $d(\cdot, \partial C_i) \equiv d(x, \partial C_i)$ (o $f \equiv f(x)$) su A , assurdo in quanto su $C_{i+1} \subset A$ la funzione assume un valore più grande (ricordiamo che $x \notin C_{i+1}$). Questo conclude la dimostrazione.

3. Commenti ed altri risultati

Supponiamo M orientabile di dimensione n . Se il soul S è orientabile e ha codimensione 1, ovvero per quanto detto all'inizio $H_{n-1}(M) \cong \mathbb{Z}$, si può dire di più sulla struttura metrica di M . Infatti in questo caso $NS \cong S \times \mathbb{R}$ è sconnesso all'infinito: scegliamo $K := \{(x, t) \in S \times \mathbb{R} : |t| \leq 1\}$; ogni compatto $C \supseteq K$ è incluso in un compatto della forma $\{(x, t) : |t| \leq R\}$, dunque per qualsiasi $x \in S$ abbiamo $(x, \pm(R+1)) \notin C$, ma $(x, R+1)$ non si può collegare a $(x, -(R+1))$ con una curva fuori da K . Da ciò segue l'esistenza di una linea γ e, in virtù del teorema di splitting di Cheeger-Gromoll, M è isometrica a $N \times \mathbb{R}$ per qualche varietà riemanniana N di dimensione $n-1$. Ovviamente anche N è compatta in quanto $H_{n-1}(N) \cong H_{n-1}(M) \cong \mathbb{Z}$.

Chiamiamo $p' := \gamma(0)$ e γ^+ , γ^- i due raggi uscenti da p' dati dalle due metà di γ . Osserviamo che in realtà γ^+ e γ^- sono gli unici raggi uscenti da p' : infatti, identificando $M = N \times \mathbb{R}$, se $\sigma(t) = (\tau(t), \alpha t)$ è un raggio deve essere $\dot{\tau} \equiv 0$; altrimenti $|\dot{\tau}| > 0$ è costante, quindi c'è un T per cui $\int_0^T |\dot{\tau}| > \text{diam}(N)$ e, sostituendo τ con la geodetica minimizzante da $\tau(0)$ a $\tau(T)$, otteniamo una geodetica $\tilde{\sigma} : [0, T] \rightarrow M$ di lunghezza minore rispetto a $\sigma|_{[0, T]}$ (con gli stessi estremi), assurdo. Ricordando la dimostrazione del teorema di splitting, sappiamo che $b_{\gamma^-} = -b_{\gamma^+}$ e concludiamo che la funzione concava costruita per mostrare il teorema del soul è $b = \min\{b_{\gamma^+}, b_{\gamma^-}\} = -|b_{\gamma^+}|$. Ma allora $S' = C_1 = \{b_{\gamma^+} = 0\}$ è isometrico a N , dove indichiamo con S' il soul costruito a partire da p' . In conclusione M è isometrica a $S' \times \mathbb{R}$. Dal risultato che ora enunciamo segue che in realtà M è pure isometrica a $S \times \mathbb{R}$.

TEOREMA 2 (Sharafudtinov). *I soul ottenuti a partire da un punto di M sono tutti isometrici tra loro, al variare del punto scelto. Inoltre esiste una retrazione $M \rightarrow S$ che non aumenta le distanze (ovvero 1-Lipschitz nel senso degli spazi metrici).*

Vale anche questo rafforzamento del teorema del soul:

TEOREMA 3 (Perelman). *Se esiste un $p \in M$ in cui tutte le curvatures sezionali sono positive allora il soul è un punto e quindi M è diffeomorfa a \mathbb{R}^n .*

Ci si può chiedere se il teorema del soul sia ancora vero assumendo solo la positività del tensore di Ricci. Tuttavia Sha e Yang, nel 1991, hanno mostrato che la somma connessa infinita $(S^2 \times S^2) \# (S^2 \times S^2) \# \dots$ ammette una metrica completa con $\text{Ric} > 0$. Questa varietà ha secondo gruppo di omotopia isomorfo a $(\mathbb{Z}_2 \times \mathbb{Z}_2)^{\oplus \mathbb{N}}$, che non è finitamente generato. Perciò essa non può avere il tipo di omotopia di una varietà compatta: questa sarebbe pure semplicemente connessa e, per un teorema di Serre, avrebbe tutti i gruppi di omotopia finitamente generati. Ciò mostra che il teorema del soul fallisce in questo caso.

4. Appendice A: Prima equazione fondamentale e raggio di convessità

In questa parte una *funzione distanza* $r : \Omega \rightarrow \mathbb{R}$ ($\Omega \subseteq M$ aperto) è una funzione liscia con $|\nabla r| = 1$ su tutto Ω . Vogliamo mettere in relazione l'evoluzione dell'hessiano di r lungo (curve integrali di) ∂_r con la curvatura. Qui $\partial_r := \nabla r$ è il gradiente di r . Osserviamo che $\text{Hess } r(X, Y) = g(\nabla_X \partial_r, Y)$ e che (simmetria dell'hessiano) $S(X) := \nabla_X \partial_r$ è un operatore autoaggiunto, detto *operatore forma*. Notiamo anche che $g(S(\partial_r), X) = g(\partial_r, S(X)) = g(\partial_r, \nabla_X \partial_r) = \frac{1}{2}X(|\partial_r|^2) = 0$ per ogni X , da cui $\nabla_{\partial_r} \partial_r = S(\partial_r) = 0$.

PROPOSIZIONE 1 (Prima equazione fondamentale). $\nabla_{\partial_r} \text{Hess } r + \text{Hess}^2 r = -R(\partial_r, \cdot, \partial_r, \cdot)$, dove $\text{Hess}^2 r(X, Y) := g(S^2(X), Y) = g(S(X), S(Y))$.

PROOF. Intanto $(\nabla_{\partial_r} \text{Hess } r)(X, Y) = \partial_r \text{Hess } r(X, Y) - \text{Hess } r(\nabla_{\partial_r} X, Y) - \text{Hess } r(X, \nabla_{\partial_r} Y)$. Ora

$$\partial_r \text{Hess } r(X, Y) = \partial_r g(\nabla_X \partial_r, Y) = g(\nabla_{\partial_r} \nabla_X \partial_r, Y) + g(\nabla_X \partial_r, \nabla_{\partial_r} Y)$$

e $\text{Hess } r(\nabla_{\partial_r} X, Y) = g(\nabla_{\nabla_{\partial_r} X} \partial_r, Y)$, $\text{Hess } r(X, \nabla_{\partial_r} Y) = g(\nabla_X \partial_r, \nabla_{\partial_r} Y)$. Perciò

$$(\nabla_{\partial_r} \text{Hess } r)(X, Y) = g(\nabla_{\partial_r} \nabla_X \partial_r, Y) - g(\nabla_{\nabla_{\partial_r} X} \partial_r, Y)$$

$$= R(X, \partial_r, \partial_r, Y) + g(\nabla_X \nabla_{\partial_r} \partial_r, Y) - g(\nabla_{\nabla_X \partial_r} \partial_r, Y) = R(\partial_r, X, \partial_r, Y) - g(S^2(X), Y),$$

che è la tesi. Nell'ultima uguaglianza abbiamo usato il fatto che $\nabla_{\partial_r} \partial_r = 0$. \square

Mostriamo ora una stima che sarà utile in seguito:

LEMMA 6. *Sia $\lambda : (0, R) \rightarrow \mathbb{R}$ localmente assolutamente continua e tale che $\lambda'(r) + \lambda^2(r) \geq -K$ nei punti di differenziabilità. Supponiamo anche che sia $\lambda(r) = \frac{1}{r} + O(r)$. Allora $\lambda(r) \geq \frac{\text{sn}'_K(r)}{\text{sn}_K(r)}$ fintanto che $\text{sn}_K(r) > 0$. Qui sn_K è la soluzione al problema di Cauchy $x'' + kx = 0$, $x(0) = 0$, $x'(0) = 1$.*

PROOF. Sia $\lambda_K := \frac{\text{sn}'_K}{\text{sn}_K}$, definita nell'intorno destro di 0 ove $\text{sn}_K > 0$. Possiamo assumere che questo intorno includa $(0, R)$. Osserviamo che λ_K risolve $\lambda'_K + (\lambda_K)^2 = -K$. Supponiamo $\lambda^2(s) + K > 0$ nell'intervallo $(0, r]$, da cui

$$\frac{\lambda'(s)}{\lambda^2(s) + K} \geq -1 = \frac{\lambda'_K(s)}{(\lambda_K)^2(s) + K}$$

e infine $\int_0^r \frac{\lambda'(s)}{\lambda^2(s) + K} ds \geq \int_0^r \frac{\lambda'_K(s)}{(\lambda_K)^2(s) + K} ds$. Perciò $F(\lambda(r)) \geq F(\lambda_K(r))$, dove F è l'unica primitiva di $\frac{1}{t^2 + K}$ infinitesima a $+\infty$. Essendo $F' > 0$ sull'intervallo di definizione di F , deduciamo che $\lambda(r) \geq \lambda_K(r)$. Esplicitando λ_K si vede che $(\lambda_K)^2 + K = -\lambda'_K > 0$, quindi $\lambda^2 + K > 0$ su tutto $(0, R)$: se fosse $\lambda^2(R') + K = 0$, per qualche $0 < R' < R$ minimo, per quanto appena mostrato avremmo $\lambda(R' - \epsilon) \geq \lambda_K(R' - \epsilon)$ per ogni $\epsilon > 0$, dunque per continuità $\lambda(R') \geq \lambda_K(R')$ e infine $\lambda^2(R') + K \geq (\lambda_K)^2(R') + K > 0$, assurdo. Perciò la tesi vale su tutto $(0, R)$. \square

COROLLARIO 2. *Se $\text{sec} \leq K$ su $B(p, R)$ (con $R \leq \text{inrad}(p)$) allora $\text{Hess } r \geq \frac{\text{sn}'_K(r)}{\text{sn}_K(r)} g_r$ fintanto che $\text{sn}_K(r) > 0$. Qui r è la distanza da p e g_r è la restrizione della metrica al tangente di $\partial B(p, r)$, ovvero $g_r = g - dr \otimes dr$.*

PROOF. Lavoriamo in coordinate polari centrate in p . Da $r \operatorname{Hess} r + dr \otimes dr = \operatorname{Hess} \frac{r^2}{2} = g + O(r^2)$ deduciamo $\operatorname{Hess} r = \frac{g_r}{r} + O(r)$. Fissato un $\theta \in S^{n-1}$, poniamo

$$\lambda(r, \theta) := \min_{v \perp \partial_r, |v|=1} \operatorname{Hess} r(v, v),$$

dove v varia nel tangente in (r, θ) ; per quanto detto prima è $\lambda(r, \theta) = \frac{1}{r} + O(r)$. $\lambda(r, \theta)$ è localmente Lipschitz in r . Sia ora r_0 un punto di differenziabilità di $r \mapsto \lambda(r, \theta)$: vogliamo mostrare che $\partial_r \lambda(r_0, \theta) + \lambda^2(r_0, \theta) \geq -K$. Scegliamo v unitario, $v \perp \partial_r$ tale che $\operatorname{Hess} r(v, v) = \lambda(r_0, v)$ ed estendiamolo a un campo parallelo V lungo la geodetica radiale da p passante per (r_0, θ) . Sia poi $\phi(r) := (\operatorname{Hess} r)_{(r, \theta)}(V, V)$. Da $\phi(r) \geq \lambda(r, \theta)$ e $\phi(r_0) = \lambda(r_0, \theta)$ segue $\phi'(r_0) = \partial_r \lambda(r_0, \theta)$, da cui

$$\begin{aligned} \partial_r \lambda(r_0, \theta) + \lambda^2(r_0, \theta) &= \phi'(r_0) + \operatorname{Hess}^2 r(v, v) \\ &= (\nabla_{\partial_r} \operatorname{Hess} r)_{(r_0, \theta)}(V, V) + \operatorname{Hess}^2 r(v, v) = -R(v, \partial_r, v, \partial_r) \geq -K, \end{aligned}$$

dove nella prima uguaglianza abbiamo usato che, per minimalità, v è anche autovettore dell'operatore forma. Dunque la funzione $r \mapsto \lambda(r, \theta)$ soddisfa le ipotesi del Lemma 6, da cui segue la tesi (perché $\operatorname{Hess} r(\partial_r, \cdot) = 0$ e quindi $(\operatorname{Hess} r)_{(r, \theta)} \geq \lambda(r, \theta)g_r$). \square

DEFINIZIONE 4. Dato $p \in M$, il raggio di convessità $\operatorname{convrad}(p)$ è il massimo $0 < R \leq \operatorname{injrads}(p)$ tale che $r = d(\cdot, p)$ è convessa su $B(p, R)$ e tale che per ogni $x, y \in B(p, R)$ la geodetica minimale che li collega è unica ed è contenuta in $B(p, R)$.

LEMMA 7. Se $R \leq \frac{1}{2} \operatorname{injrads}(x)$ per ogni $x \in B(p, R)$ e $R \leq \frac{\pi}{2\sqrt{K}}$, con $K := \sup \{ \sec(\pi) : \pi \subseteq T_x M, x \in M \}$, allora $\operatorname{convrad}(p) \geq R$ (ignoriamo la seconda condizione se $K \leq 0$).

PROOF. Per la prima condizione r è liscia su $B(p, 2R) \setminus \{p\}$ e per ogni $x, y \in B(p, R)$ c'è un'unica geodetica minimale che li unisce (in quanto $d(x, y) < 2R \leq \operatorname{injrads}(x)$). Grazie a 2, la seconda condizione implica che $\operatorname{Hess} r \geq 0$ su $B(p, R)$ e quindi su $\overline{B(p, R)}$. Resta da mostrare che per ogni $x, y \in B(p, R)$ la geodetica minimale γ che li unisce è contenuta in $B(p, R)$. Fissato $x \in B(p, R)$ sia C_x l'insieme degli y per cui ciò vale. C_x è chiaramente aperto; verifichiamo che è anche chiuso in $B(p, R)$: dato $y \in \overline{C_x} \cap B(p, R)$, per continuità è $\gamma \subseteq \overline{B(p, R)}$. Ora la funzione $r \circ \gamma$ assume valori minori di R agli estremi ed è convessa (ricordiamo che $(r \circ \gamma)'' = \operatorname{Hess} r(\dot{\gamma}, \dot{\gamma}) \geq 0$), quindi assume valori minori di R ovunque. Ciò mostra che in realtà è $\gamma \subseteq B(p, R)$, per cui $y \in C_x$. \square

COROLLARIO 3. Su ogni sottoinsieme compatto di M $\operatorname{convrad}$ è limitato dal basso da una costante positiva.

5. Appendice B: Equivalenza delle nozioni di concavità

LEMMA 8. Per una funzione continua $f : \Omega \rightarrow \mathbb{R}$ la concavità nel senso delle barriere è equivalente a quella nel senso delle geodetiche.

PROOF. Supponiamo f concava n.s.b. e consideriamo una qualsiasi geodetica $\gamma : (a, b) \rightarrow M$. Applicando la definizione di concavità n.s.b. in $\gamma(t)$ per ogni $t \in (a, b)$ deduciamo che $F := f \circ \gamma : (a, b) \rightarrow \mathbb{R}$ è concava n.s.b. Vogliamo mostrare che F è concava

nel senso standard. A questo scopo possiamo anche supporre $F''' \leq -\epsilon < 0$, sostituendo F con $F_\epsilon := F - \frac{\epsilon}{2}x^2$ (infatti mandando poi $\epsilon \rightarrow 0$ deduciamo che anche $F = \lim_{\epsilon \rightarrow 0} F_\epsilon$ è concava).

La concavità equivale al fatto che, per ogni funzione affine $\phi : (a, b) \rightarrow \mathbb{R}$, i sopralivelli $\{F - \phi \geq \alpha\}$ sono tutti convessi. Osserviamo che anche $G := F - \phi - \alpha$ soddisfa $G''' \leq -\epsilon$ n.s.b., perciò presi $c, d \in \{G \geq 0\}$ qualsiasi con $c < d$ ci siamo ridotti a vedere che $[c, d] \subseteq \{G \geq 0\}$. Se così non fosse, la funzione G assumerebbe valori negativi in $[c, d]$: dunque $m := \min_{[c, d]} f < 0$ e c'è un $\bar{x} \in (c, d)$ tale che $G(\bar{x}) = m$. Ma per ipotesi esiste $H \in C^2(U)$ ($U \subseteq (c, d)$ intorno di x_0) con $H \geq G$, $H(\bar{x}) = G(\bar{x})$ e $H''(\bar{x}) \leq -\epsilon + \frac{\epsilon}{2} = -\frac{\epsilon}{2} < 0$. Da $H \geq G \geq m = G(\bar{x}) = H(\bar{x})$ segue che anche H ha minimo in \bar{x} , perciò essendo H concava in un intorno di \bar{x} (è $H''(\bar{x}) < 0$) deduciamo che qui $H \equiv m$. Dunque anche $G \equiv m$ in un intorno di \bar{x} e da ciò segue che $(c, d) \cap G^{-1}(m)$ è aperto e chiuso. Quindi per connessione $G \equiv m < 0$ su tutto (c, d) , assurdo per la continuità di G e il fatto che $G(c), G(d) \geq 0$.

Supponiamo ora f concava n.s.g. e fissiamo un $p \in \Omega$. Usiamo coordinate normali $(x^1, \dots, x^n) : B(p, \delta) \rightarrow B_\delta$ centrate in p (indichiamo con B_r la palla aperta di \mathbb{R}^n di centro 0 e raggio r). Scegliamo $0 < R < \frac{\delta}{2}$ in modo che $B_{R'}$ sia geodeticamente convessa per ogni $R' \leq R$ e valga $\text{inrad}(x) \geq 3R$ per ogni $x \in \overline{B_R}$. Mostriamo intanto che f è Lipschitz su $\overline{B_R}$: presi $x, y \in B_R$ chiamiamo γ la geodetica minimale di velocità unitaria che li unisce (contenuta in B_R), con $\gamma(0) = x$ e $\gamma(a) = y$ ($0 \leq a \leq 2R$). Estendiamo γ a una geodetica $\mathbb{R} \rightarrow M$; per l'ipotesi fatta su $\text{inrad}(\cdot)$ l'insieme $\{t \geq 0 : \gamma(t) \in \partial B_R\}$ è non vuoto: infatti $d(\gamma(2R), x) = 2R$ e $d(x, p) \leq R$, da cui $d(\gamma(2R), p) \geq R$ (con un abuso di notazione stiamo identificando x con il corrispondente punto di M). Dunque esiste $\bar{s} := \min\{t \geq 0 : \gamma(t) \in \partial B_R\} > a$. Analogamente esiste $\bar{u} := \min\{u \geq 0 : \gamma(u) \in \partial B_{2R}\} > \bar{t}$ e ovviamente $d(\gamma(\bar{t}), \gamma(\bar{u})) \geq R$, da cui segue $\bar{u} - \bar{t} \geq R$. Dalla concavità di $f \circ \gamma$ otteniamo

$$\frac{f(y) - f(x)}{d(x, y)} = \frac{f \circ \gamma(a) - f \circ \gamma(0)}{a} \geq \frac{f \circ \gamma(\bar{u}) - f \circ \gamma(\bar{t})}{\bar{u} - \bar{t}} \geq -\frac{2C}{R},$$

dove $C := \max_{\overline{B_{2R}}} |f|$. Scambiando i ruoli di x e y arriviamo a

$$\frac{|f(y) - f(x)|}{d(x, y)} \leq \frac{2C}{R}$$

e deduciamo la lipschitzianità di $f : B_R \rightarrow \mathbb{R}$ rispetto alla metrica riemanniana su B_R . Poiché il tensore metrico (g_{ij}) ha coefficienti limitati su $\overline{B_R}$ è anche $d(x, y) \leq C'|x - y|$, quindi $f : B_R \rightarrow \mathbb{R}$ è pure Lipschitz nel senso standard del termine, con costante di Lipschitz $L := \frac{2CC'}{R}$.

Per il teorema di Rademacher f è differenziabile quasi ovunque, dunque esiste un \bar{x} arbitrariamente vicino a 0 dove esiste $df = \alpha_i dx^i$. Poniamo $g(x) := \alpha_i x^i + \epsilon|x - \bar{x}|^2$ (per un $\epsilon > 0$ fissato) e notiamo che $\text{Hess } g > 0$ in una palla $B_{R'} \subseteq B_R$ indipendente da \bar{x}^3 . Ora affermiamo che $h := g + f(\bar{x}) - g(\bar{x}) \geq f$ su tutta $B_{R'}$: infatti la funzione

³ $\text{Hess } g = \sum_i (\alpha_i - 2\bar{x}^i) \text{Hess}(x^i) + \text{Hess}(\epsilon x^2)$ e $\text{Hess}(x^i)$ si annulla in 0 (l'annullarsi dei simboli di Christoffel nell'origine permette di calcolare l'hessiano in 0 come nella metrica piatta); essendo $|\alpha_i| \leq L$, la somma $\sum_i (\alpha_i - 2\bar{x}^i) \text{Hess}(x^i)$ è molto piccola in un intorno di 0 indipendente da \bar{x} , mentre $\text{Hess}(\epsilon x^2)(0) = 2\epsilon g_0$ e quindi $\text{Hess}(\epsilon x^2) > \epsilon g_0$ vicino a 0

$h - f = g + f(\bar{x}) - g(\bar{x}) - f$ è nulla in \bar{x} , è convessa n.s.g. e ha differenziale nullo in \bar{x} , perciò componendola con un tratto geodetico uscente da \bar{x} (incluso in $B_{R'}$) si ottiene una funzione convessa, nulla in un punto e con derivata nulla nello stesso punto; dunque questa deve essere ovunque nonnegativa e la tesi segue dal fatto che $B_{R'}$ è geodeticamente convesso. Infine scegliamo una successione (\bar{x}_k) di punti di differenziabilità per f , con $\bar{x}_k \rightarrow 0$. Le relative funzioni h_k sono equicontinue con $h_k(0) \rightarrow f(0)$, quindi (a meno di sottosuccessioni) convergono uniformemente a una $\tilde{h} : B_{R'} \rightarrow \mathbb{R}$ con $\tilde{h}(0) = f(0)$ e $\tilde{h} \geq f$ su $B_{R'}$. Osserviamo che $\tilde{h}(x) = f(0) + \tilde{\alpha}_i x^i + \epsilon|x|^2$ per opportuni $\tilde{\alpha}_i \in \mathbb{R}$ (le funzioni affini $h_k(x) - \epsilon|x|^2$ devono convergere a una funzione affine). Perciò \tilde{h} è la barriera cercata: è liscia, soddisfa $\tilde{h} \geq f$ vicino a 0 con uguaglianza in 0 e $\text{Hess } \tilde{h}(0) = 2\epsilon g_0$. \square