

HYPERBOLIC SPACES ARE OF STRICTLY NEGATIVE TYPE

P. G. HJORTH, S. L. KOKKENDORFF, AND S. MARKVORSEN

(Communicated by Jozef Dodziuk)

ABSTRACT. We study finite metric spaces with elements picked from, and distances consistent with, ambient Riemannian manifolds. The concepts of negative type and strictly negative type are reviewed, and the conjecture that hyperbolic spaces are of strictly negative type is settled, in the affirmative. The technique of the proof is subsequently applied to show that every compact manifold of negative type must have trivial fundamental group, and to obtain a necessary criterion for product manifolds to be of negative type.

1. INTRODUCTION

A finite metric space (X, d) is a metric space with finitely many elements. Such a space is completely characterized by its metric matrix, $\mathbf{D} = \{d_{ij}\}_{1 \leq i, j \leq n}$, which is symmetric, has nonnegative elements, zeros along the diagonal, and whose elements satisfy $n(n-1)(n-2)/6$ triangle inequalities. For a classical exposition of distance geometry, including that of finite metric spaces, we refer the reader to the book [1].

Suppose we know that the n elements of X happen to be points picked from an ambient (not necessarily finite) metric space, in such a way that the metric matrix is inherited from the metric of the ambient space. It is interesting to study whether geometric properties of the ambient space may reveal themselves as purely matrix properties of \mathbf{D} .

One main result of this paper, Theorem 3.4, provides a sufficient condition for finite metric spaces embedded in a Riemannian manifold to carry the important matrix property known as strictly negative type (see the definitions below).

Not all embedded finite metric spaces are of strictly negative type. It is a property sufficient to ensure uniqueness of the so-called ∞ -extender, a constellation of weighted points realizing the transfinite diameter of the finite metric space. For a discussion of these concepts see [4].

An infinite metric space, e.g. a Riemannian manifold, is said to be of (strictly) negative type if all finite metric spaces with elements picked from the manifold and distances inherited from the manifold are of (strictly) negative type. The Euclidean spaces \mathbb{R}^m are known to be of strictly negative type. The spheres \mathbb{S}^m are not (a discussion of \mathbb{S}^m in this connection is given below). It has been conjectured ([4]) that the hyperbolic spaces $\mathbb{H}_{\mathbb{R}}^m$ are of strictly negative type. This conjecture is settled, in the affirmative, by Corollary 4.2 of section 4.

Received by the editors May 19, 2000.

2000 *Mathematics Subject Classification*. Primary 51K99, 53C20.

Key words and phrases. Metric spaces, negative type, strictly negative type, length spaces, fundamental group.

In the final section, section 5, we obtain a connection between simply connect-
edness and negative type for compact manifolds, as well as a necessary criterion for
a product manifold to be of negative type.

2. DEFINITIONS

Let (X, d) be a finite metric space with n points,¹ and let $\mathbf{D} = \{d_{ij}\}_{1 \leq i, j \leq n}$ be
the distance matrix of (X, d) . The definitions concern properties of the quadratic
form, $x\mathbf{D}x^T = \sum_{i,j} x_i x_j d_{ij}$, associated to \mathbf{D} .

Definition 2.1. (X, d) is said to be of *negative type* if $x\mathbf{D}x^T \leq 0$ for all x in the
hyperplane $\Pi_0^n = \{x \in \mathbb{R}^n \mid \sum_{i=1}^n x_i = 0\}$, and (X, d) is of *strictly negative type* if
the inequality is strict for $x \in \Pi_0^n \setminus \{0\}$.

Definition 2.2. (X, d) is called *hypermetric* if $x\mathbf{D}x^T \leq 0$ for all x in the discrete
hyperplane set $\{x \in \mathbb{Z}^n \mid \sum_{i=1}^n x_i = 1\}$.

Remark 2.3. It is not difficult to see that hypermetricity implies negative type; cf.
[5]. In [4] it was shown that regularity of the distance matrix together with hyper-
metricity implied strictly negative type. Thus, for hypermetric spaces, regularity is
equivalent to being of strictly negative type.

Definition 2.4. Two points p, q form an *antipodal pair* in the metric space (X, d)
if $d(p, q)$ is the diameter of (X, d) and $d(p, q) = d(p, r) + d(r, q)$ for all $r \in X$.

Lemma 2.5. *If the metric space (X, d) contains 4 points, p_1, p_2, p_3, p_4 , satisfying
 $d_{13} = d_{24}$ and $d_{12} + d_{23} = d_{13}, d_{14} + d_{34} = d_{13}$, then it is not of strictly negative
type.*

Proof. Choose weights $x_1 = x_3 = 1$ and $x_2 = x_4 = -1$ and $x_i = 0$ for $i > 4$ if there
are additional points. Then $x = (x_1, x_2, x_3, x_4, 0, \dots) \in \Pi_0$ and it is easily checked
that $\sum_{i,j} x_i x_j d_{ij} = 0$. \square

The conditions of the lemma will be satisfied if p_1, p_3 and p_2, p_4 are two pairs of
antipodal points.

Remark 2.6. It is known that all finite subspaces of Euclidean space \mathbb{R}^m are hy-
permetric and regular and hence of strictly negative type. This is also true for
trees; see [4]. \mathbb{S}^m is hypermetric and hence of negative type. In [4] the subspaces
of \mathbb{S}^m of strictly negative type are identified as those containing at most one pair
of antipodal points.

3. THE MAIN RESULT

The Theorem below will be stated in the class of metric spaces called *length
spaces*; cf. [2]. Later we shall apply the result to important special cases from Rie-
mannian geometry, including the hyperbolic spaces. Let us first recall the definition
of a *geodesic* in a length space (M, d) :

Definition 3.1. A curve $\gamma : I \rightarrow M$, where $I \subset \mathbb{R}$ is an interval, is called a *geodesic*
if each $t \in I$ has a neighbourhood $J \subset I$ such that

$$d(\gamma(s), \gamma(t)) = |s - t|, \forall t, s \in J.$$

¹Spaces with less than 5 points have been completely characterized with respect to strictly
negative type; see [4].

Any two points in a locally compact, complete length space, e.g. a complete Riemannian manifold, can be joined by a minimizing geodesic; cf. [2].

Definition 3.2. Let $X = \{p_1, p_2, \dots, p_n\}$ be a finite subset of a length space (M, d) , and let $x = (x_1, \dots, x_n) \in \mathbb{R}^n$. The pair (X, x) will be called Σ -regular if $x = 0$ or there exist an $i \in \{1, \dots, n\}$ and a geodesic $\gamma : I \rightarrow M$, such that:

1. $x_i \neq 0, \gamma(0) = p_i$.
2. $\sum_{j \neq i} x_j d(p_j, \gamma(t))$ is C^1 as a function of t in a neighbourhood of $0 \in I$.

A pair (X, x) which is not Σ -regular will be called Σ -critical.

Remark 3.3. Note that if the pair (X, x) is Σ -critical, then at least two of the weights $x_i, x_j, i \neq j$, must be nonzero.

When M is a manifold we can consider differentiability of the distance function in the usual sense. If M is a complete Riemannian manifold equipped with the Riemannian distance function and a pair (X, x) is Σ -critical, then for each i with $x_i \neq 0$ differentiability of at least one of the distances d_{ij} must break down at p_i . It follows that p_i must be contained in the *cut locus* of at least one of the other points p_j ; cf. [2].

We want to characterize maxima of $\sum_{i,j} x_i x_j d_{ij}$ considered as a function of both finite subspaces $X \subset M$ and corresponding weight vectors $x \in \mathbb{R}^{|X|}$. In addition to the issue of negative type when $x \in \Pi_0^{|X|} \subset \mathbb{R}^{|X|}$, there is another interesting situation which can be treated simultaneously. This is the question of maximal *extender pairs*, where the weights are picked from the set

$$\Delta^{|X|} = \{x \in \mathbb{R}^{|X|} \mid \sum_i x_i = 1, x_i \geq 0 \forall i\}.$$

Theorem 3.4. Let (M, d) be a length space, and let $X = \{p_1, p_2, \dots, p_n\} \subset M$ be a finite subspace. In each of the following cases the pair (X, x) must be Σ -critical:

- (M, d) is of negative type and $\sum_{i,j} x_i x_j d_{ij} = 0$ with $x \in \Pi_0^n \setminus \{0\}$.
- (X, x) is a maximum of $\sum_{i,j} x_i x_j d_{ij}$ considered as a function of finite subsets $X \subset M$ and weight vectors in $\Delta^{|X|}$.

Proof. The proof goes by contradiction. Let $X = \{p_1, p_2, \dots, p_n\}$ and $x \in \mathbb{R}^n$ be given, and assume that one of the hypotheses above is satisfied with the pair (X, x) being Σ -regular. We have $x \neq 0$ and may assume that the index from Definition 3.2 is $i = 1$.

Introduce an additional point p_0 and consider the vector $(\frac{x_1}{2}, \frac{x_1}{2}, x_2, x_3, \dots, x_n) \in \mathbb{R}^{n+1}$. This new vector will be in one of the relevant subsets of weight vectors, i.e. Π_0^{n+1} or Δ^{n+1} . Let the quadratic form \tilde{U} associated to the augmented space $\tilde{X} = \{p_0, p_1, \dots, p_n\}$ be given by

$$\begin{aligned} \tilde{U} &= \sum_{0 \leq i, j \leq n} x_i x_j d_{ij} = \underbrace{\frac{2}{4} x_1^2 d_{01}}_{p_0 \leftrightarrow p_1} + 2 \underbrace{\sum_{j>1} \frac{1}{2} x_1 x_j d_{1j}}_{p_1 \leftrightarrow p_2, \dots, p_n} + 2 \underbrace{\sum_{j>1} \frac{1}{2} x_1 x_j d_{0j}}_{p_0 \leftrightarrow p_2, \dots, p_n} + \underbrace{\sum_{i,j>1} x_i x_j d_{ij}}_{\text{other interactions}} \\ &= \frac{1}{2} x_1^2 d_{01} + \sum_{i,j \geq 1} x_i x_j d_{ij} + \sum_{j>1} x_1 x_j (d_{0j} - d_{1j}) = \frac{1}{2} x_1^2 d_{01} + \sum_{j>1} x_1 x_j (d_{0j} - d_{1j}). \end{aligned}$$

We may consider the above as a function $\tilde{U} : M \rightarrow \mathbb{R}$ of the position of p_0 . Let a geodesic $\gamma : I \rightarrow M$ with $\gamma(0) = p_1$ as in Definition 3.2 be given, and let p_0 move

along γ , i.e. consider the function $\phi = \tilde{U} \circ \gamma$. For $t = 0$ we have $p_0 = \gamma(0) = p_1$ and we get the original space X back, $\phi(0) = U_0$. Hence by hypothesis, $t = 0$ is a maximum of ϕ .

As a function of t , the first term in the above sum becomes $\frac{1}{2}x_1^2 d(p_1, \gamma(t)) = \frac{1}{2}x_1^2|t|$, at least in a neighbourhood of 0. Since this term has a minimum at $t = 0$, the second term $x_1 \sum_{j>1} x_j(d(p_j, \gamma(t)) - d_{1j})$ must have a maximum. But the pair (X, x) was assumed Σ -regular, so this sum is C^1 as a function of $t \in I$, and hence $t = 0$ is a critical point, $\frac{d}{dt}|_{t=0} x_1 \sum_{j>1} x_j(d(p_j, \gamma(t)) - d_{1j}) = 0$.

By elementary Taylor theory the second term is $0 + \epsilon(x)|x|$. It follows that the first term will dominate for $t \rightarrow 0$, and hence that $\tilde{U} > U_0$ for p_0 “close” to p_1 along γ . But this contradicts the assumption that U_0 was a maximum. \square

Corollary 3.5. *A length space (M, d) which is of negative type and does not admit Σ -critical pairs is also of strictly negative type.*

By the above we see immediately that \mathbb{R}^m is of strictly negative type.

4. APPLICATIONS

Here we shall apply Theorem 3.4 to Riemannian manifolds equipped with the Riemannian distance function. We immediately have the following:

Theorem 4.1. *If M is a Riemannian manifold and $d : M \times M \setminus \{(p, p) | p \in M\} \rightarrow \mathbb{R}_+$ is C^1 , then negative type of (M, d) implies strictly negative type of (M, d) .*

Proof. Since d is C^1 away from the diagonal in $M \times M$, each pair (X, x) is clearly Σ -regular. Hence a nontrivial zero of $\sum_{i,j} x_i x_j d_{ij}$ cannot exist. \square

A case of principal interest is the hyperbolic spaces over the real numbers. By a result of G. Robertson (cf. [7]), $\mathbb{H}_{\mathbb{R}}^m$ is hypermetric and hence of negative type. There are various explicit formulae for the distance function of $\mathbb{H}_{\mathbb{R}}^m$, depending on the model,² revealing that d is smooth off the diagonal; cf. [3] [7]. This is in fact true for any so-called *Hadamard manifold*, i.e. a simply connected, complete Riemannian manifold of nonpositive curvature (such a manifold is diffeomorphic to \mathbb{R}^m by the Hadamard covering theorem; cf. [6]).

Corollary 4.2. *$\mathbb{H}_{\mathbb{R}}^m$ is of strictly negative type.*

Complex hyperbolic spaces $\mathbb{H}_{\mathbb{C}}^m$ are also known to be of negative type; cf. [3], [7]. They are symmetric, simply connected spaces with sectional curvature $-1 \leq K \leq -\frac{1}{4}$ ([6], chap. 11), and hence Hadamard manifolds diffeomorphic to \mathbb{R}^{2m} . So we also have:

Corollary 4.3. *$\mathbb{H}_{\mathbb{C}}^m$ is of strictly negative type.*

Theorem 3.4 also provides a new way to prove the result on spheres quoted previously; cf. Theorem 9.1 in [4].

Corollary 4.4. *The finite subspaces of \mathbb{S}^m of strictly negative type are exactly those subspaces which do not contain two pairs of antipodal points.*

²E.g. for the upper half space model: $\cosh d(p, q) = 1 + \frac{\|p-q\|^2}{2p_m q_m}$, where p_m, q_m are the last coordinates.

Proof. Since \mathbb{S}^m is of negative type, any pair (X, x) with $x \in \Pi_0 \setminus \{0\}$ giving a zero of $\sum_{i,j} x_i x_j d_{ij}$ must be Σ -critical. But since differentiability of the distance from a point breaks down exactly at the antipode, for each point p_i receiving a nonzero weight x_i , the antipodal point must also receive a nonzero weight. But only two points with nonzero weights will not produce a zero, hence there must be at least two pairs of antipodal points. On the other hand, observe that if X does contain two pairs of antipodal points, then by Lemma 2.5, a nontrivial zero of $\sum_{i,j} x_i x_j d_{ij}$ exist. \square

5. CLOSED GEODESICS AND NEGATIVE TYPE

The technique developed in Theorem 3.4 can also be applied in another direction to show that a large class of compact Riemannian manifolds are *not* of negative type. First some preliminary results:

Lemma 5.1. *In a compact length space which is not simply connected, there is among the closed curves which are nontrivial wrt free homotopy a curve of minimal length. This curve γ is a simply closed geodesic, and the image $\gamma(\mathbb{R})$ is as a metric subspace isometric to \mathbb{S}^1 , modulo a scale factor.*

Proof. For the existence of the minimal, simply closed geodesic, see [2], p. 215. Let $\gamma : \mathbb{R} \rightarrow M$ be this geodesic, and let $2L > 0$ be the period, i.e. $\gamma(t+2L) = \gamma(t) \forall t \in \mathbb{R}$. γ will have to minimize distance between $\gamma(t)$ and $\gamma(t+s)$ for $0 \leq s \leq L$, that is, $d(\gamma(t), \gamma(t+s)) = s$. For if there were a shorter connection ω between $\gamma(t)$ and $\gamma(t+s)$, two closed curves shorter than γ could be constructed by ω and the two arcs of γ . These two curves would have to be homotopically trivial, and hence this would also be true for γ , a contradiction. Hence the image of γ is isometric to a circle of radius $\frac{L}{\pi}$. \square

Corollary 5.2. *A compact length space of strictly negative type is simply connected.*

Proof. If the space is not simply connected, then by Lemma 5.1, two pairs of “antipodal” points $p_1 = \gamma(0), p_3 = \gamma(L)$ and $p_2 = \gamma(\frac{L}{2}), p_4 = \gamma(\frac{3L}{2})$, satisfying the conditions in Lemma 2.5, can be found, contradicting strictly negative type. \square

Recall that two points $p = \gamma(0), q = \gamma(t)$ in a Riemannian manifold are *conjugate* along the geodesic γ if the differential of the exponential map $\exp_p : T_p M \rightarrow M$ is singular at the point $t\gamma'(0) \in T_p M$; this condition is symmetric in p and q . For details on conjugate points and geodesic variations we refer to [2] or [6].

Lemma 5.3. *Assume that (M, d) is a compact Riemannian manifold which is not simply connected, and let $\gamma : \mathbb{R} \rightarrow M$ be a geodesic with period $2L$ as in Lemma 5.1. Let $p = \gamma(0)$ and $q = \gamma(L)$ be two “antipodal” points and ω a geodesic through p orthogonal to γ , i.e. $\omega(0) = p$ and $g(\gamma'(0), \omega'(0)) = 0$, where g denotes the Riemannian metric. For each $c > 0$ we have:*

$$d(p, q) - c|s| < d(q, \omega(s)), \text{ for } s \in I_c \setminus \{0\}, \text{ where } I_c \text{ is a neighbourhood of } 0.$$

Proof. First observe that two “antipodal” points $\gamma(t)$ and $\gamma(t+L)$ cannot be conjugate along one of the two arcs of γ connecting them, since then a shorter closed

curve homotopic to γ could be constructed. Hence geodesic variations ν_1, ν_2 of the two arcs of γ connecting p to q can be constructed, satisfying:

$$\begin{aligned}\nu_1(0, t) &= \gamma(t), \quad \nu_2(0, t) = \gamma(t + L), \\ \nu_1(s, 0) &= \omega(s) = \nu_2(s, L), \\ \nu_1(s, L) &= \nu_2(s, 0) = q, \\ \text{for } (t, s) &\in [0, L] \times [0, 1].\end{aligned}$$

Now by the first variation formula (cf. [2]) we have for the length of the geodesics parametrized by s : $\frac{d}{ds}|_{s=0}l_1(s) = \frac{d}{ds}|_{s=0}l_2(s) = 0$, where $l_i(s) = \int_0^L \|\frac{\partial}{\partial t}\nu_i(s, t)\|dt$. Hence for a given $c > 0$, there exists a neighbourhood I_c of 0 such that $l_1(s), l_2(s) < L + c|s|$, for $s \in I_c \setminus \{0\}$. If for some $s \in I_c \setminus 0$ we had $d(q, \omega(s)) \leq L - c|s|$, then q and $\omega(s)$ could be connected by a geodesic λ realizing the distance. But then $l_i(s) + l(\lambda) < 2L$, and two closed curves shorter than γ could be constructed by λ and $\nu_1(s, \cdot), \nu_2(s, \cdot)$. So these two curves would have to be homotopically trivial, and then this would also be true for γ , which is not the case. \square

This completes the setting for:

Theorem 5.4. *A compact Riemannian manifold (M, d) of negative type is simply connected.*

Proof. Assume that M is not simply connected and let $\gamma : \mathbb{R} \rightarrow M$ be a geodesic with period $2L$ as in Lemma 5.1. Choose two pairs of ‘‘antipodal’’ points $p_1 = \gamma(0), p_3 = \gamma(L)$ and $p_2 = \gamma(\frac{L}{2}), p_4 = \gamma(\frac{3L}{2})$, and choose weights $x_1 = x_3 = 1, x_2 = x_4 = -1$, as in Lemma 2.5. Let ω be a geodesic through p_1 as in Lemma 5.3. Now as in the proof of Theorem 3.4 introduce a new point p_0 which moves on the geodesic ω , i.e. $p_0 = \omega(s)$, and consider the new weights $x = (\frac{1}{2}, \frac{1}{2}, -1, 1, -1)$. Now for this pair we have for the quadratic form

$$\tilde{U}(s) = \frac{1}{2}d_{01} + d_{03} - (d_{02} + d_{04}) + \text{constant terms.}$$

For $s = 0$ we get the original pair back, so that $\tilde{U}(0) = 0$. The first term is as a function of s : $\frac{1}{2}d_{01} = \frac{1}{2}d(\omega(s), \omega(0)) = \frac{1}{2}|s|$. The third term is differentiable as a function of s , since the point p_1 and hence p_0 for s ‘‘small’’ is not in the cut locus of p_2 and p_4 , and since we move orthogonal to γ this function has derivative zero (first variation formula). All in all the first and third terms is to first order $\frac{1}{2}|s| + \epsilon(s)|s|$. But by Lemma 5.3, d_{03} cannot decrease faster than this close to $s = 0$, so that $\tilde{U}(s)$ will be positive for s ‘‘small enough’’. \square

By the above theorem, simply connectedness is a necessary condition for a compact Riemannian manifold to be of negative type. That this is certainly not sufficient is easily seen by the following:

Lemma 5.5. *Assume that (M_1, d_1) is a metric space containing 4 points as in Lemma 2.5, and let (M_2, d_2) be any Riemannian manifold. Let $f : \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}$ be a C^1 -mapping on $(\mathbb{R} \times \mathbb{R}) \setminus \{(0, 0)\}$ satisfying $f(x, 0) = x$ and $f(0, y) = y, \forall (x, y) \in \mathbb{R} \times \mathbb{R}$, such that $d_3 = f(d_1, d_2)$ is a metric on $M_1 \times M_2$. Then $(M_1 \times M_2, d_3)$ is not of negative type.*

Proof. Let $p_1, p_2, p_3, p_4 \in M_1$ be the 4 points as in Lemma 2.5 and let $q \in M_2$ be any point. Then $X = \{(p_1, q), \dots, (p_4, q)\}$ also satisfies the conditions of the

lemma wrt d_3 and gives a zero of the quadratic form. It is clear that a geodesic γ of (M_2, d_2) can be lifted to a geodesic in $(M_1 \times M_2, d_3)$. Using such a (vertical) geodesic through one of the points (p_i, q) , the technique of Theorem 3.4 can be applied directly. \square

Corollary 5.6. *A Riemannian product manifold, where S^m , $m \geq 1$, is one of the factors, is not of negative type.*

REFERENCES

1. L. M. Blumenthal: Theory and Applications of Distance Geometry, 2nd ed., Chelsea, 1970. MR **42**:3678
2. I. Chavel: Riemannian Geometry: A Modern Introduction, Cambridge University Press, 1993. MR **95j**:53001
3. J. Faraut, and K. Harzallah: Distances hilbertiennes invariantes sur un espace homogène, *Ann. Inst. Fourier* 24:171-217, 1974. MR **51**:1295
4. P.G. Hjorth, P. Lisonek, S. Markvorsen, and C. Thomassen: Finite Metric Spaces of Strictly Negative Type, *Linear Algebra and its Applications* 270:255-273, 1998. MR **98k**:15027
5. J. B. Kelly: Hypermetric Spaces and Metric Transforms, in *Inequalities II* (O. Shisha, Ed.), Academic Press, 1970, pp. 201-207.
6. B. O'Neill: Semi-Riemannian Geometry, Academic Press, 1983. MR **85f**:53002
7. G. Robertson: Crofton formulae and geodesic distance in hyperbolic spaces, *J. Lie Theory* 8:163-172, 1998. CMP 99:03

DEPARTMENT OF MATHEMATICS, TECHNICAL UNIVERSITY OF DENMARK, BUILDING 303, DK-2800 LYNGBY, DENMARK

E-mail address: P.G.Hjorth@mat.dtu.dk

DEPARTMENT OF MATHEMATICS, TECHNICAL UNIVERSITY OF DENMARK, BUILDING 303, DK-2800 LYNGBY, DENMARK

E-mail address: S.L.Kokkendorff@mat.dtu.dk

DEPARTMENT OF MATHEMATICS, TECHNICAL UNIVERSITY OF DENMARK, BUILDING 303, DK-2800 LYNGBY, DENMARK

E-mail address: S.Markvorsen@mat.dtu.dk