Notes on locally CAT(1) spaces.

B. H. Bowditch
Department of Mathematical Sciences, University of Aberdeen,
Edward Wright Building, Dunbar Street, Aberdeen AB9 2TY, U.K.

0. Introduction.

In this paper, we aim to describe a few of the basic results concerning locally compact path-metric spaces satisfying the "CAT(1)" property locally. These spaces generalise the idea of a riemannian manifold having sectional curvature everywhere at most 1.

Attempts to formulate curvature as a purely metric property go back to the work of Aleksandrov, Toponogov and Busemann. Thus, one treats one of the comparison theorems of Riemannian geometry as axiomatic. Since curvature is a local property, we will want a formulation that allows us to pass readily from local to global.

In [Gr], Gromov introduced the term "CAT(χ)" for a comparison axiom intended to capture the idea of a space having curvature everywhere $\leq \chi$. This is defined in the context of "geodesic spaces" in which every pair of points a joined by a geodesic—a length-minimising rectifiable path. It is a metric condition on triangles formed from three geodesics edges, as we describe at the end of this chapter. In this paper, we shall restrict attention to complete locally compact path-metric spaces. These are necessarily geodesic spaces (Lemma 2.2).

We may speak of a space being "locally" or "globally" CAT(χ). After scaling the metric, we reduce to three qualitatively distinct cases, namely $\chi = -1, 0, 1$. If $\chi \leq 0$, a locally CAT(χ) space will be globally CAT(χ) if and only if it is simply-connected (and hence contractible). See, for example [P]. To make an analogous statement in the case $\chi = 1$, we should replace simple-connectedness (i.e. path-connectedness of the space of loops) by a related condition obtained by restricting attention to those loops which are rectifiable, and of length strictly less than 2π. Thus this new condition demands instead that the subspace of such loops be path-connected. (We explore this matter in Chapter 3.)

It seems that the cases $\chi = -1, 0, 1$, become progressively more difficult to deal with from a synthetic point of view. Much of the geometry

of strictly negatively curved spaces (corresponding to $\chi = -1$) can even be formulated combinatorially, as can be seen from Gromov's hyperbolicity criterion [Gr]. In the case of non-positive curvature, the appropriate combinatorial formulation has yet to be settled on. In the context of combinatorial group theory, the notions of combability and automaticity (and the numerous variations thereof) are candidates [E]. However a great deal can be done by synthetic means. Much of the basic theory, as developed in [BaGS] for example, can be carried out for CAT(0) spaces. We also have the weaker notion of convexity of the distance function, as formulated by Busemann [Bus]. See also [P] for an account of this. Unfortunately, positively curved spaces prove less amenable to synthetic argument, and much of the development seems confined to the Riemannian category. However we can get some mileage out of the CAT(1) axiom as we shall describe. Note that in all these cases, the CAT(χ) axiom is intended to place only an upper bound on curvature. The curvature is thus allowed to be arbitrarily, or indeed "infinitely" negative. It is not clear what is a useful way to formulate lower curvature bounds, or how to say that a metric space is "non-negatively curved".

As examples of locally CAT(χ) spaces (besides riemannian manifolds) we can consider geometric simplicial complexes obtained by gluing together simplices of constant curvature χ. In such a complex, the link of each simplex has, itself, the structure of a geometric complex built out of spherical simplices (the case $\chi = 1$). It turns out that the original complex will be locally CAT(χ) if and only if the link of each vertex is globally CAT(1). (To be formally consistent, we should say that each connected component of such a link is globally CAT(1).) By induction on dimension, one can see that this is equivalent to saying that the link of each simplex should contain no closed geodesic of length strictly less than 2π. This latter observation uses the fact that a compact locally CAT(1) space is globally CAT(1) if and only if it contains no closed geodesic of length less than 2π. For a discussion of polyhedral complexes, see [Ba] or [Br].

Another context in which the CAT(1) property arises naturally concerns the realisation of 3-dimensional hyperbolic polyhedra. Given a compact convex polyhedron in hyperbolic 3-space, we can construct its dual in the de-Sitter sphere [HR]. Intrinsically, this dual is a topological 2-sphere with a singular spherical metric, i.e. it has constant curvature 1 away from a finite number of cone points. Each of the cone points has cone angle $\geq 2\pi$. The metric is thus locally CAT(1). Hodgson and Rivin [HR] characterise the metrics that occur in this way as precisely those which contain no closed geodesic of length $\leq 2\pi$. This is slightly stronger than globally CAT(1) — a globally CAT(1) space might contain a closed geodesic of length precisely 2π. It turns out that (up to isometry) there is a bijective correspondence between such metrics and compact hyperbolic polyhedra.

A natural quantity associated to a compact locally CAT(1) space is the "systole" [CD]. This may be defined as the length of the shortest closed geodesic, provided this is $\leq 2\pi$. Otherwise, we set the systole to equal 2π. There are various equivalent ways of defining this quantity as we shall describe in Chapters 2 and 3. One can also give a definition for non-compact spaces. In all cases the space will be globally CAT(1) if and only if the systole equals 2π.

In Chapter 3, we introduce he notion of a "shrinkable loop" in a locally CAT(1) space. Briefly, a rectifiable loop of length less than 2π is said to be *shrinkable* if it can be (freely) homotoped to a point (constant loop) passing only through other rectifiable loops of length less than 2π. It turns out that if this is possible, we can always choose the homotopy so that the lengths of the intermediate loops tend monotonically to 0 (Theorem 3.1.5). As mentioned earlier, a locally CAT(1) space is globally CAT(1) if and only if every loop of length less than 2π is shrinkable. (This follows from Theorem 3.1.2 and Corollary 3.1.7.) In Section 3.1, we list various other results relating to shrinkability. Most of Chapter 3 will be devoted to proving these results. (Sections 3.5 and 3.7 are digressions from the rest of the paper.) The main technique used in these results is that of "Birkhoff curve shortening", which we describe in Section 3.3.

In Chapter 4, we mention two area inequalities from 2-dimensional riemannian geometry, to which the notion of shrinkability is relevant. Firstly it gives a convenient way of describing a dichotomy arising out of the spherical isoperimetric inequality. Secondly, it gives an hypothesis under which we can prove an area comparison theorem for triangles.

Most of the material for this paper war prepared at the University of Melbourne, under an A.R.C. fellowship. It was completed at the University of Aberdeen under an S.E.R.C. research assistantship.

Definitions.

We give the definitions of the CAT(χ) property, and Busemann's convexity condition. We shall look at the CAT(1) property more carefully in Chapter 2.

Suppose (X, d) is a complete locally-compact path-metric space. By a *geodesic* in X, we mean a globally length-minising rectifiable path between two given points. We assume geodesics to be parameterised proportionately to arc-length.

Suppose that $\alpha_1, \alpha_2, \alpha_3 : [0,1] \longrightarrow X$ are three geodesics forming a triangle, i.e. $\alpha_i(1) = \alpha_{i+1}(0)$ where $3+1=1$. Suppose that $\sum_{i=1}^{3} d(\alpha_i(0), \alpha_i(1)) < 2\pi/\sqrt{\chi}$, where this condition is deemed vacuous if $\chi \leq 0$. We may construct a *comparison triangle* in the 2-dimensional model space (S_χ, d_χ) of constant curvature χ, consisting of three geodesic segments $\alpha_1', \alpha_2', \alpha_3' : [0,1] \longrightarrow S_\chi$, with $d_\chi(\alpha_i'(0), \alpha_i'(1)) = d(\alpha_i(0), \alpha_i(1))$ for $i = 1,2,3$. This comparison triangle is well defined up to isometry in S_χ. We say that X is "(globally) CAT(χ)" if for all such triangles $(\alpha_1, \alpha_2, \alpha_3)$, for all $t, u \in [0,1]$ and for all $i, j \in \{1,2,3\}$, we have $d(\alpha_i(t), \alpha_j(u)) \leq d_\chi(\alpha_i'(t), \alpha_j'(u))$. We say that X is "locally CAT(χ)" if every point has a neighbourhood which is CAT(χ).

Note that the CAT(0) property implies the following: If $\alpha, \beta : [0,1] \longrightarrow X$ are geodesics, then the map $[(t,u) \mapsto d(\alpha(t), \beta(u))] : [0,1]^2 \longrightarrow [0, \infty)$ is convex. This property, taken as an axiom, is (equivalent to) Busemann's condition for non-positive curvature. It may be alternatively expressed in terms of bisecting edges of triangles [Bus].

1. Spherical Geometry.

In this chapter, we describe a few of the basic facts of spherical geometry relevant to later sections. We shall not write out detailed proofs here. Much of it can be dismissed as "spherical trigonometry", though most of the results can be deduced by synthetic argument without resort to computation. In particular, Propositions 1.1 and 1.2 can be thought of intuitively in terms of mechanical linkages.

Let (S^2, d_1) be the unit sphere with the intrinsic path-metric d_1. Thus, two points $x, y \in S^2$ are antipodal if and only if $d_1(x,y) = \pi$; otherwise $d_1(x,y) < \pi$. In the latter case, x and y may be joined by a unique geodesic segment $[x,y] \subseteq S^2$. If $x \in S^2$, and $y, z \in S^2 \setminus \{x\}$, neither antipodal to x, then we write $y\hat{x}z \in [0,\pi]$ for the angle between $[x,y]$ and $[x,z]$ at x. We say that a closed set $Q \subseteq S^2$ is *convex* if $[x,y] \subseteq Q$ for any pair of non-antipodal points $x, y \in Q$. We say that Q is *strictly convex* if it is convex and contains no pair of antipodal points. In the latter case, Q lies inside some open hemisphere of S^2. Moreover, if it has non-empty interior, then it is topologically a disc, bounded by a rectifiable Jordan curve, ∂Q, of length $< 2\pi$.

We shall say that a closed subset $Q \subseteq S^2$ is *small* if it is contained in some open hemisphere, and *large* if its interior contains some closed hemisphere.

Any Jordan curve $J \subseteq S^2$ bounds two (closed) topological discs. If J is rectifiable and of length $< 2\pi$, then one of these is large, and the other small. (This follows from Proposition 1.3 below.)

Here, we are only interested in polygonal curves.

Definition : A *(non-degenerate) polygon* (or *n-gon*), P, in S^2 is a cyclically ordered set of (n) points $(x_1, x_2, \ldots, x_n)$ of X such that $0 < d(x_i, x_{i+1}) < \pi$ for all i, and such that $\Gamma(P) = \bigcup_{i=1}^{n}[x_i, x_{i+1}]$ is a Jordan curve.

We are using the convention that $n + 1 = 1$. We write

$$\operatorname{perim}(P) = \operatorname{length}\Gamma(P) = \sum_{i=1}^{n} d(x_i, x_{i+1})$$

for the *perimeter* of P.

Suppose that $\operatorname{perim}(P) < 2\pi$. Let $R(P)$ be the small disc bounded by P. We refer to $R(P)$ as a *small polygonal region*. For each i, we write $\angle(P, x_i)$ for the interior angle of $R(P)$ at x_i. Thus $R(P)$ is convex if and only if $\angle(P, x_i) \leq \pi$ for all i (so that $\angle(P, x_i) = x_{i-1}\hat{x}_i x_{i+1}$). We also refer to P being "convex" in this case. Note that all triangles (3-gons) are convex.

Proposition 1.1 : *Suppose $P = (x_1, \ldots, x_n)$ and $P' = (x'_1, \ldots, x'_n)$ are convex n-gons. Suppose $d_1(x_i, x_{i+1}) = d_1(x'_i, x'_{i+1})$ for $i \in \{1, \ldots n-1\}$, and that $\angle(P, x_i) \leq \angle(P', x'_i)$ for $i \in \{2, \ldots, n-1\}$. Then $d_1(x_1, x_n) \leq d_1(x'_1, x'_n)$.* ◇

The proof, in general, is somewhat involved (see, for example, [S]). We are principally interested in the case of triangles, for which it is elementary. Thus, if we "open out" an angle of a triangle, we increase the length of the opposite side. The case of quadrilaterals ($n = 4$) is also needed in the proof of Proposition 1.2.

We shall also need to consider non-convex quadrilaterals, which we refer to as "darts".

Definition : A *dart*, P, is a quadrilateral (x, z, y, w) such that $\operatorname{perim}(P) < 2\pi$ and $\angle(P, y) \geq \pi$.

Given such a dart, we write ρ_P for the intrinsic path-metric on $R(P)$. Note that $R(P)$ is the union of two triangular regions $R(x, y, z)$ and $R(x, y, w)$. We say that the darts $P = (x, z, y, w)$ and $P' = (x', z', y', w')$ are *equivalent* if $d_1(x, z) = d_1(x', z')$, $d_1(z, y) = d_1(z', y')$, $d_1(y, w) = d_1(y', w')$ and $d_1(w, x) = d_1(w', x')$. Thus there is a natural map $f : \Gamma(P) \longrightarrow \Gamma(P')$ obtained by sending each geodesic segment of $\Gamma(P)$ isometrically onto the corresponding segment of $\Gamma(P')$. Note that in a given equivalence class, there is a 1-parameter family of darts, up to isometry in S^2. This can be thought of intuitively in terms of flexing a mechanical linkage.

Proposition 1.2 : *Suppose $P = (x, z, y, w)$ and $P' = (x', z', y', w')$ are equivalent darts. Then, the following are equivalent:*

(1) $\angle(P, x) \geq \angle(P', x')$

(2) $\angle(P, y) \leq \angle(P', y')$

(3) $\angle(P, z) \leq \angle(P', z')$ *(and/or* $\angle(P, w) \leq \angle(P', w')$.*)*

(4) $d_1(x, y) \geq d_1(x', y')$

(5) $d_1(z, w) \geq d_1(z', w')$.

(6) The natural map $f : (\Gamma(P), \rho_P) \longrightarrow (\Gamma(P'), \rho_{P'})$ is distance non-increasing.

(7) $\operatorname{area}(P) \geq \operatorname{area}(P')$. ◇

The proof is left as an exercise. It uses Proposition 1.1.

Of particular interest, is the extreme case where $\angle(P, y) = \pi$. There is precisely one such "triangular" dart in each equivalence class.

We have already observed that all triangles are convex. It will often be convenient to allow for "degenerate" triangles $T = (x, y, z)$, where the only condition is that no pair of points of $\{x, y, z\}$ are antipodal. Thus $\operatorname{perim}(T) < 2\pi$.

We shall need:

Proposition 1.3 : *For all $\epsilon > 0$, there exists $\eta > 0$ such that if $T = (x,y,z)$ is a (possibly degenerate) triangle, with $d(y,x) + d(x,z) \leq \pi - \epsilon$, then $d_1(a,x) \leq \frac{\pi}{2} - \eta$, where a is the midpoint of $[y,z]$.* ◇

As a corollary, referred to earlier, suppose that $J \subseteq S^2$ is a Jordan curve length $J < 2\pi$. Choose $y, z \in J$ so as to divide J into two subarcs, each of length less than π, and let a be the midpoint of $[x,y]$. Then, J lies in the open hemisphere centred on a. Thus, J bounds one large disc and one small disc.

2. Basic Properties.

In this chapter, we develop some of the basic properties of a locally CAT(1) space, X. We shall be principally interested, for the moment, in the case where X is compact. Much of what we do in this chapter can be viewed as a combination of the accounts of Ballmann [Ba] and Charney and Davis [CD].

In [CD], the authors define "systole", $\mathrm{sys}(X)$, of a compact locally CAT(1) space, X, as the minimum length of a closed geodesic in X. We shall find it convenient here to demand that $\mathrm{sys}(X) \leq 2\pi$, i.e. we set $\mathrm{sys}(X) = 2\pi$ if there is no closed geodesic of length $< 2\pi$. The systole seems a natural quantity to associate to X, and one can give several different characterisations. We shall see that X is globally CAT(1) if and only if $\mathrm{sys}(X) = 2\pi$.

Note that, up till now, we have not said explicitly what we mean by a "closed geodesic". There are two sensible interpretations. We could take it to mean a *closed local geodesic*, i.e. a non-constant map of a circle S^1 into X which is length minimising when restricted to all sufficiently small subarcs; or else we could take it to mean a *closed global geodesic* which is globally length minimising, i.e. the (pseudo)metric on S^1 induced by the metric d is an intrinsic path-metric on S^1. Thus a global geodesic is always embedded, whereas a local geodesic need not be. The interpretation in [CD] is that of a global geodesic. It turns out that both interpretations give rise to the same notion of systole. In fact, any closed local geodesic of length equal to $\mathrm{sys}(X)$ is necessarily a global geodesic (see Corollary 2.19).

In Chapter 3, we give another description of the systole in terms of what we call "shrinkability" of loops. This formulation also makes sense

when X is not compact.

We begin with some definitions.

Let (X,d) be a complete locally compact path-metric space. The path-metric property tells us a-priori that, given any $\epsilon > 0$ and $x, y \in X$, there is a rectifiable path joining x to y of length at most $d(x,y) + \epsilon$. We shall see that the hypotheses of completeness and local compactness imply that we can always take this path to be geodesic, i.e. we can set $\epsilon = 0$.

Given $x \in X$ and $r \geq 0$, we shall write $N(x,r) = \{y \in X \mid d(x,y) \leq r\}$ for the closed metric r-ball about x.

Lemma 2.1 : *For all $x \in X$ and $r \geq 0$, the metric ball $N(x,r)$ is compact.*

Proof : Fix $x \in X$ and suppose, for contradiction, that not all closed metric balls about x are compact. Let $r = \inf\{t \geq 0 \mid N(x,t) \text{ is not compact}\}$. By local compactness, $r > 0$.

First, we claim that $N(x,r)$ is compact. To see this, let $(y_i)_{i \in \mathbf{N}}$ be any sequence of points of $N(x,r)$. Since d is a path-metric, we can find, for any $j \in \mathbf{N} \cap [1 + \frac{1}{r}, \infty)$, points y_{ij} such that $d(a, y_{ij}) \leq r - \frac{1}{j}$ and $d(y_i, y_{ij}) \leq \frac{1}{i} + \frac{1}{j}$. For all such j, the ball $N(a, r - \frac{1}{j})$ is compact. Thus, by a diagonal sequence argument, we can find a subsequence of natural numbers, $i(k)$, such that for all $j \in \mathbf{N} \cap [1 + \frac{1}{r}, \infty)$, the sequence $(y_{i(k)j})_{k \in \mathbf{N}}$ converges as as k tend to ∞. Now, given $k, k' \in \mathbf{N}$, we have $d(y_{i(k)}, y_{i(k')}) \leq d(y_{i(k)j}, y_{i(k')j}) + \frac{2}{j} + \frac{1}{i(k)} + \frac{1}{i(k')}$ which can be made arbitrarily small. Thus, the sequence $(y_{i(k)})_{k \in \mathbf{N}}$ is Cauchy, and so converges in $N(x,r)$. Thus, $N(x,r)$ is compact as claimed.

Now, by local compactness, we can find $\epsilon > 0$ and some finite set $A \subseteq N(x,r)$ such that $N(x,r) \subseteq \bigcup_{a \in A} N(a,\epsilon)$, and such that $N(a, 3\epsilon)$ is compact for all $a \in A$. Now, since d is a path-metric, we have $N(x, r+\epsilon) \subseteq \bigcup_{a \in A} N(a, 3\epsilon)$, and we deduce that $N(x,t)$ is compact for all $t \leq r + \epsilon$, contradicting the definition of r. ◇

Definition : We define a *(global) geodesic* joining two points $x, y \in X$ to be a path $\alpha : [0,1] \longrightarrow X$ with $\alpha(0) = x$ and $\alpha(1) = y$, such that $d(\alpha(t), \alpha(u)) = \mu|t - u|$ for all $t, u \in [0,1]$ and for some fixed μ (= length α). In other words, α is a globally length-minimising path parameterised proportionately to arc-length.

Lemma 2.2 : *Given any two points $x, y \in X$, there is a geodesic joining x to y.*

Proof : Let $r = \frac{1}{2}d(x,y)$. Since d is a path-metric, we can find, for any $i \in \mathbf{N}$, a point $z_i \in X$ satisfying $d(x,z_i) \leq r + \frac{1}{i}$ and $d(y,z_i) \leq r + \frac{1}{i}$. Applying Lemma 2.1, we can find a subsequence of (z_i) converging to a point $z \in X$, satisfying $d(x,z) = r$ and $d(y,z) = r$.

We can now interpolate, by induction, to obtain a map of the diadic rationals $[0,1] \cap \mathbf{Z}[\frac{1}{2}]$ such that $d(\alpha(t), \alpha(u)) = 2r|t-u|$ for all $t, u \in [0,1] \cap \mathbf{Z}[\frac{1}{2}]$, and with $\alpha(0) = x$ and $\alpha(1) = y$. We may now extend by continuity to a geodesic $\alpha : [0,1] \longrightarrow X$. ◇

Given a geodesic $\alpha : [0,1] \longrightarrow X$, we write $\mathrm{end}(\alpha) = (\alpha(0), \alpha(1))$ for the pair of *endpoints* of α.

Definition : A *triangle* in X consists of three points $x_1, x_2, x_3 \in X$ and three geodesic segments $\alpha_1, \alpha_2, \alpha_3$ such that $\mathrm{end}(\alpha_i) = (x_{i+1}, x_{i+2})$ (taking subscripts mod 3). (Figure 2a.) We write $T = (\alpha_1, \alpha_2, \alpha_3)$, or $T = (\alpha_1, \alpha_2, \alpha_3; x_1, x_2, x_3)$ if we wish to specify the endpoints. We write $\Gamma(T)$ for the closed path $\alpha_1 \cup \alpha_2 \cup \alpha_3$, and write

$$\mathrm{perim}(T) = \mathrm{length}\,\Gamma(T) = \sum_{i=1}^{3} d(x_i, x_{i+1})$$

for the *perimeter* of T.

Definition : Suppose $T = (\alpha_1, \alpha_2, \alpha_3; x_1, x_2, x_3)$ is a triangle with perim $(T) < 2\pi$. By a *(spherical) comparison triangle* for T, we mean a triangle $T' = (\alpha'_1, \alpha'_2, \alpha'_3; x'_1, x'_2, x'_3)$ in (S^2, d_1) such that $d_1(x'_i, x'_{i+1}) = d(x_i, x_{i+1})$ for $i \in \{1,2,3\}$.

Such a triangle always exists, and is unique up to isometry in S^2. It may happen that T' is degenerate, i.e. $\Gamma(T')$ need not be a Jordan curve.

Definition : Suppose $\mathrm{perim}(T) < 2\pi$. We say that T is "CAT(1)" if $d(\alpha_i(t), \alpha_j(u)) \leq d_1(\alpha'_i(t), \alpha'_j(u))$ for all $t, u \in [0,1]$ and $i, j \in \{1,2,3\}$.

(When dealing with triangles we shall sometimes be careless in distinguishing a path α from the reverse path $[t \mapsto \alpha(1-t)]$. In other words

we shall not always insist that the edges are consistantly oriented around a triangle.)

The crucial lemma, which gets the whole subject going, says that if we can cut a triangle into two smaller triangles, by joining a vertex to the a point on the opposite edge, so that each of these smaller triangles is CAT(1), then the original triangle must be CAT(1). More formally:

Lemma 2.3 : *Suppose $T = (\gamma, \beta, \alpha; x, z, w)$ is a triangle in X with $\mathrm{perim}(T) < 2\pi$. Suppose $t \in [0,1]$. Let $y = \gamma(t)$, and let γ_1 and γ_2 be (linear reparameterisations of) $\gamma|[0,t]$ and $\gamma|[t,1]$ respectively. Suppose δ is a geodesic joining x to y, so that we have cut T into two triangles $T_1 = (\gamma_1, \delta.\alpha; x, y, z)$ and $T_2 = (\gamma_2, \delta.\beta; x, y, w)$ with $\mathrm{perim}(T_i) < 2\pi$. (Figure 2b.) If T_1 and T_2 are both CAT(1), then T is CAT(1).*

Proof : We can construct comparison triangles for $T'_1 = (x', y', z')$ and and $T'_2 = (x', y', w')$ for T_1 and T_2 respectively, so that the triangular regions $R(T_1)$ and $R(T_2)$ lie on opposite sides of the common edge $[x', y']$. Let P be the quadrilateral (x', z', y', z'). Thus $R(P) = R(T_1) \cup R(T_2)$. Let ρ be the intrinsic path-metric on $R(P)$. There is a natural piecewise isometric map $g : (\Gamma(P), \rho) \longrightarrow (\Gamma(T), d)$. Since T_1 and T_2 are CAT(1), it follows easily that this is distance non-increasing. In particular, we see that $\rho(z', w') = \rho(z', y') + \rho(y', w')$, and so $\angle(P, y') \geq \pi$. Thus P is a "dart" in the sense of Chapter 1.

We can "open out" the angle at y to form an equivalent dart $Q = (x'', z'', y'', w'')$ with $\angle(Q, y'') = \pi$. (Figure 2c.) Thus (x'', z'', w'') is a comparison triangle for T. Now Lemma 2.2(6) tells us that the natural map $f : (\Gamma(Q), d_1) \longrightarrow (\Gamma(P), \rho)$ is distance non-increasing, and so therefore is the composition $g \circ f : (\Gamma(Q), d_1) \longrightarrow (\Gamma(T), d)$. This is the CAT(1) property for T. ◇

A degenerate case of a triangle is a *bigon* B consisting of two geodesic arcs α, β joining the same pair of points x, y. Thus $\mathrm{perim}(B) = 2d(x,y)$. From the definition, B is CAT(1) if and only if $\alpha = \beta$.

Definition : We say that X is "r-CAT(1)" if every triangle of perimeter strictly less than $2r$ is CAT(1).

We say that X is "globally CAT(1)" if it is π-CAT(1).

We say that X is "locally CAT(1)" if every point has a compact neighbourhood which is CAT(1) in the induced path-metric.

Note that the CAT(1) property implies local convexity. (We say more about this at the end of this chapter.) In other word, in the definition of locally CAT(1), we can assume that the neighbourhoods we take are convex. Thus, the induced path-metric agrees with the metric d. This observation leads to an alternative formulation of locally CAT(1):

Lemma 2.4 : *The space X is locally CAT(1) if and only if for all compact subsets $K \subseteq X$, there is some $\epsilon > 0$ such that if T is a triangle in X with all its vertices in K, and with $\mathrm{perim}(T) < 2\epsilon$, then T is CAT(1)* ◇

In particular, a compact path-metric space is CAT(1) if and only if it is ϵ-CAT(1) for some $\epsilon > 0$.

For the rest of this chapter, we shall assume that X is compact and ϵ-CAT(1).

We write $\mathrm{geod}(X)$ for the space of all geodesics in X, with the sup-norm metric d_{sup}. Thus $d_{sup}(\alpha, \beta) = \max\{d(\alpha(t), \beta(t)) \mid t \in [0,1]\}$. Since X has finite diameter, geodesics are uniformly lipshitz, and so:

Lemma 2.5 : *The space $(\mathrm{geod}(X), d_{sup})$ is compact.* ◇

We have defined the endpoint map $\mathrm{end} : \mathrm{geod}(X) \longrightarrow X \times X$ by $\mathrm{end}(\alpha) = (\alpha(0), \alpha(1))$. Clearly this is continuous, and Lemma 2.2 tells us that it is surjective. Note that $\mathrm{length}(\alpha) = d(\alpha(0), \alpha(1))$, and so $\mathrm{length} : \mathrm{geod}(X) \longrightarrow [0, \infty)$ is also continuous. Given $\alpha \in \mathrm{geod}(X)$, we write $-\alpha$ for $[t \mapsto \alpha(1-t)] \in \mathrm{geod}(X)$.

If $x, y \in X$ with $d(x,y) < \epsilon$, then there is a unique geodesic joining x to y, which we write as $[x \rightarrow y]$.

Lemma 2.6 : *Suppose $\alpha, \beta \in \mathrm{geod}(X)$ with $\mathrm{end}(\alpha) = (z, x)$ and $\mathrm{end}(\beta) = (z, y)$. Suppose $d_{sup}(\alpha, \beta) < \epsilon$ and $d(x,z) + d(y,z) + \epsilon \leq 2\pi$. Then $T = (\beta, \alpha, [x \rightarrow y]; x, y, z)$ is CAT(1)*

Proof : We can find $0 = t_1 < t_2 < \cdots < t_k = 1$ such that $\mathrm{perim} S_i < 2\epsilon$ and $\mathrm{perim} T_i < 2\epsilon$ for all i, where S_i and T_i are respectively the triangles with vertices $(\alpha(t_i), \alpha(t_{i+1}), \beta(t_i))$ and $(\beta(t_i), \beta(t_{i+1}), \alpha(t_{i+1}))$. (Figure 2d.) We can now apply Lemma 2.3 inductively. ◇

Corollary 2.7 : *If $\alpha, \beta \in \mathrm{geod}(X)$ with $\mathrm{length}\,\alpha = \mathrm{length}\,\beta \leq \pi$, $\mathrm{end}\,\alpha = \mathrm{end}\,\beta$ and $d_{sup}(\alpha, \beta) < \epsilon$, then $\alpha = \beta$.*

Proof : By Lemma 2.6, the bigon $\alpha \cup -\beta$ is CAT(1). ◇

Corollary 2.8 : *If $x, y \in X$ with $d(x,y) < \pi$, then $\mathrm{end}^{-1}(x,y) \subseteq \mathrm{geod}(X)$ is finite.*

Proof : By Lemma 2.5, and Corollary 2.7. In fact, $|\mathrm{end}^{-1}(x,y)|$ is bounded by the number of $(\epsilon/2)$-balls we can pack disjointly into $(\mathrm{geod}(X), d_{sup})$. ◇

Given $r > 0$, write

$$P(r) = \{(x,y) \in X \times X \mid d(x,y) < r\}.$$

Given $(x,y) \in P(\pi)$, write $n(x,y) = |\mathrm{end}^{-1}(x,y)|$. Thus, by Lemma 2.2, $n(x,y) \geq 1$ for all (x,y)

Lemma 2.9 : *The map $n : P(\pi) \longrightarrow \mathbf{N}$ is upper-semicontinuous.*

Proof : Suppose (x_i, y_i) is a sequence converging to $(x,y) \in P(\pi)$, and that $n(x_i, y_i) \geq m$ for all $i \in \mathbf{N}$. For each i, choose $\{\alpha_{i1}, \ldots, \alpha_{im}\}$ to be m distinct geodesics with $\mathrm{end}(\alpha_{ij}) = (x_i, y_i)$. Since $\mathrm{geod}(X)$ is compact, passing to a subsequence we have $\alpha_{ij} \rightarrow \alpha_j \in \mathrm{geod}(X)$ for each j Clearly $\mathrm{end}(\alpha_j) = (x,y)$. Since, for all i, $d_{sup}(\alpha_{ij}, \alpha_{ik}) \geq \epsilon$ if $j \neq k$, we see that the α_j are all distinct. Thus $n(x,y) \geq m$. ◇

Now let $P_0(\pi) = \{(x,y) \in P(\pi) \mid n(x,y) \geq 2\}$. By Lemma 2.9, this is a closed subset of $P(\pi)$. If $P_0 \neq \emptyset$, set $l = \min\{d(x,y) \mid (x,y) \in P(\pi)\}$. Otherwise, set $l = \pi$. (Thus, l is the "injectivity radius" of X. We shall see that it is equal to half the systole.)

Given $(x,y) \in P(l)$, there is a unique geodesic $[x \rightarrow y] \in \mathrm{geod}(X)$ with $\mathrm{end}\,[x \rightarrow y] = (x,y)$.

Lemma 2.10 : *The map $[(x,y) \mapsto [x \rightarrow y]] : P(l) \longrightarrow \mathrm{geod}(X)$ is continuous.*

Proof : We know that end : $\text{end}^{-1}P(l) \longrightarrow P(l)$ is a continuous bijection. We need to know that its inverse is continuous. This follows from the observations that $P(t)$ is open in $P(l)$ for $t < l$, that $P(l) = \bigcup_{t<l} P(t)$, that $P(l)$ is hausdorff and that $\text{end}^{-1}P(t)$ is relatively compact in $\text{end}^{-1}P(l)$ for all $t < l$. Thus $\text{end}^{-1}|P(t)$ is continuous for all $t < l$. ◇

Proposition 2.11 : *X is l-CAT(1).*

Proof : Suppose $T = (\alpha, \beta, \gamma; x, y, z)$ has $\text{perim}(T) < 2l$. Thus $d(z, \gamma(t)) < l$ for all $t \in [0,1]$, so we can set $\delta_t = [z \to \gamma(t)]$. By Lemma 2.10, the map $[t \mapsto \delta_t] : [0,1] \longrightarrow \text{geod}(X)$ is continuous. Thus, we can find $0 = t_1 < t_2 < \cdots < t_k = 1$ with $d_{sup}(\delta_{t_i}, \delta_{t_{i+1}}) < \epsilon$ for all $i \in \{0, 1, \ldots, k-1\}$. (Figure 2e.) By Lemma 2.6, each triangle $T_i = (\delta_{t_i}, \delta_{t_{i+1}}, [\gamma(t_i) \to \gamma(t_{i+1})])$ is CAT(1). By Lemma 2.3, T is CAT(1). ◇

We want to relate the "injectivity radius", l, to the "systole" of X, i.e. the length of the shortest closed geodesic.

Suppose σ is a path-metric on the circle S^1. Thus, up to homeomorphism of S^1, σ is a multiple of the standard path-metric on S^1 as the unit circle in $\mathbf{R}^2$. Let $r = \frac{1}{2}\text{length}(S^1, \sigma) = \text{diam}(S^1, \sigma)$. We say that the points $t, t' \in S^1$ are *antipodal* if $\sigma(t, t') = r$.

Definition : A *closed (global) geodesic* (of length $2r$) is a map $\gamma : S^1 \longrightarrow X$ such that $d(\gamma(t), \gamma(u)) = \sigma(t, u)$ for all $t, u \in S^1$, for some path-metric σ on S^1.

Lemma 2.12 : *If $\gamma : (S^1, \sigma) \longrightarrow (X, d)$ is distance non-increasing, and $d(\gamma(t), \gamma(t')) = r$ whenever $t, t' \in S^1$ are antipodal, then γ is a global geodesic.*

Proof : Suppose $t, u \in S^1$. Then $d(\gamma(t), \gamma(u)) \leq \sigma(t, u)$ and $d(\gamma(t'), \gamma(u) \leq \sigma(t', u)$. But $\sigma(t, u) + \sigma(t', u) = \sigma(t, t') = r = d(\gamma(t), \gamma(t')) \leq d(\gamma(t), \gamma(u)) + d(\gamma(t'), \gamma(u))$. Thus $d(\gamma(t), \gamma(u)) = \sigma(t, u)$. ◇

Suppose $\alpha, \beta \in \text{geod}(X)$ are non-constant geodesics with $\text{end}(\alpha) = \text{end}(\beta)$. Let σ be a path-metric on S^1 of length $2r$, where $r = \text{length}\,\alpha = \text{length}\,\beta$. We divide S^1 into two subintervals I, J, each of length r, and

define a map $\gamma : S^1 \longrightarrow X$ by letting $\gamma|I$ and $\gamma|J$ be linear reparametrisations of α and $-\beta$ respectively. We write $\gamma = \alpha \cup -\beta$. Note that $\gamma : (S^1, \sigma) \longrightarrow (X, d)$ is distance non-increasing and $d(\gamma(t), \gamma(u)) = \sigma(t, u)$ if $t, u \in I$ or if $t, u \in J$.

Lemma 2.13 : *Suppose $\alpha, \beta \in \text{geod}(X)$ with $\alpha \neq \beta$ and $\text{end}\,\alpha = \text{end}\,\beta = (x, y)$. Suppose $d(x, y) = l$, and that $l < \pi$. Then $\gamma = \alpha \cup -\beta$ is a closed global geodesic.*

Proof : Suppose not. Then, by Lemma 2.12, we can find $t \in (0,1)$ such that $d(\alpha(t), \beta(1-t)) < l$. Let $\delta = [\alpha(t) \to \beta(1-t)]$. Thus δ cuts the bigon (α, β) into two triangles $T_1 = (\alpha_1, \beta_1, \delta)$ and $T_2 = (\alpha_2, \beta_2 . \delta)$ where $\alpha_1, \alpha_2, \beta_1, \beta_2$ are, respectively, linear reparameterisations of $\alpha|[0,t], \alpha|[1-t, 1], \beta|[0,t], \beta|[1-t, 1]$. (Figure 2f.) Now, $\text{perim}(T_i) < 2l$, and so by Proposition 2.11, each T_i is CAT(1). Thus, by Lemma 2.3, the bigon (α, β) is CAT(1), and so $\alpha = \beta$. ◇

Corollary 2.14 : *If $l < \pi$, then there is a closed global geodesic of length equal to $2l$.*

Proof : By Lemma 2.9, and the definition of l, we can find $x, y \in X$ with $d(x, y) = l$ and $n(x, y) \geq 2$. Thus, there are at least two distinct geodesics $\alpha, \beta \in \text{geod}(X)$ with $\text{end}\,\alpha = \text{end}\,\beta = (x, y)$. ◇

In summary, we have shown (Proposition 2.11 and Corollary 2.14) that:

Theorem 2.15 : *If X is compact and locally CAT(1), then there is some $l \in (0, \pi]$ such that X is l-CAT(1), and either $l = \pi$, or else there is a closed global geodesic of length equal to $2l$.* ◇

Now, if γ is a closed global geodesic of length $< 2\pi$, we may divide γ into three geodesic paths, $\gamma = \alpha_1 \cup \alpha_2 \cup \alpha_3$, of equal length, to form a triangle $T = (\alpha_1, \alpha_2, \alpha_3)$ with $\text{perim}(T) = \text{length}\,\gamma$. Clearly the CAT(1) property fails for T. It follows that the quantity l as described by Theorem 2.15 is uniquely determined.

Definition : We define the *systole*, $\text{sys}(X)$ of X to be equal to $2l$ where l is the quantity described by Theorem 2.15.

Thus, either $\text{sys}(X) = 2\pi$, or else it is the length of the shortest closed global geodesic. We shall see that we can replace "global" by "local" in the above statement. We thus go on to describe local geodesics.

Definition : A path $\alpha : [0,1] \longrightarrow X$ is a *local geodesic* if, for some $\eta > 0$ and $\mu \geq 0$, we have $d(\alpha(t), \alpha(u)) = \mu|t-u|$ whenever $t, u \in [0,1]$ and $|t-u| \leq \eta$.

Note that $d(\alpha(t), \alpha(u)) \leq \mu|t-u|$ for all $t, u \in [0,1]$. Also, a local geodesic is a global geodesic if and only if $\text{length}\,\alpha = d(\alpha(0), \alpha(1))$.

As for global geodesics, we shall write $\text{end}\,\alpha = (\alpha(0), \alpha(1))$.

Lemma 2.16 : *Suppose $\alpha : [0,1] \longrightarrow X$ is a local geodesic with* $\text{length}\,\alpha \leq l = \frac{1}{2}\text{sys}(X)$. *Then α is a global geodesic.*

Proof : Let $t_0 = \max\{t \in [0,1] \mid \text{length}(\alpha|[0,t]) = d(\alpha(0), \alpha(t))\}$. Certainly $t_0 > 0$. If α is not a geodesic, then $t_0 < 1$. In this case, choose $u_0, u_1 \in (0,1)$ so that $|u_1 - u_0| < \eta$ and $u_0 < t < u_1$. Thus $d(\alpha(u_0), \alpha(u_1)) = \mu(u_1 - u_0) = \text{length}(\alpha|[u_0, u_1])$. Let $\alpha_1, \alpha_2 : [0,1] \longrightarrow X$ be linear reparameterisations of $\alpha|[0, u_0]$ and $\alpha|[u_0, u_1]$ respectively. Thus $\alpha_1, \alpha_2 \in \text{geod}(X)$. (Figure 2g.) Since $u = 1 > t_0$, we have $d(\alpha(0), \alpha(u_1)) < \text{length}(\alpha|[0, u_1]) < l$. Let $\beta = [\alpha(0) \to \alpha(u_1)] \in \text{geod}(X)$ and let $T = (\alpha_1, \alpha_2, \beta)$. Thus, $\text{perim}(T) < 2l$, and so T satisfies CAT(1). Now $\text{length}\,\beta < \text{length}\,\alpha_1 + \text{length}\,\alpha_2$, from which we can deduce that $d(\alpha(0), \alpha(t)) < \text{length}(\alpha|[0,t])$ for all $t \in (u_0, u_1)$. This contradicts the assumption that $t_0 < 1$, and so α must be geodesic. ◊

Suppose $\alpha, \beta : [0,1] \longrightarrow X$ are local geodesics with $\text{end}\,\alpha = \text{end}\,\beta$. Provided α and β are not both constant, we can define $\gamma = \alpha \cup -\beta : S^1 \longrightarrow X$ as in the case of global geodesics. If σ is the induced path-metric on S^1, then $\gamma : (S^1, \sigma) \longrightarrow (X, d)$ is distance non-increasing and $\text{length}\,\gamma = \text{length}(S^1, \sigma) = \text{length}\,\alpha + \text{length}\,\beta$.

Lemma 2.17 : *Suppose $\alpha, \beta : [0,1] \longrightarrow X$ are local geodesics with $\alpha \neq \beta$, $\text{end}\,\alpha = \text{end}\,\beta$, and $\text{length}\,\alpha + \text{length}\,\beta \leq \text{sys}(X)$. Then, either $\text{length}\,\alpha = \text{length}\,\beta = \pi$, or else $\gamma = \alpha \cup -\beta$ is a closed global geodesic.*

(Note that it follows that $\text{length}\,\gamma = \text{length}\,\alpha + \text{length}\,\beta = \text{sys}(X)$.)

Proof : Let $r = \frac{1}{2}(\text{length}\,\alpha + \text{length}\,\beta) \leq \frac{1}{2}\text{sys}(X)$. If $\text{length}\,\alpha = \text{length}\,\beta$, then, by Lemma 2.16, $\alpha, \beta \in \text{geod}(X)$. Since $\alpha \neq \beta$, we must have $r = \frac{1}{2}\text{sys}(X)$. Thus, either $r = \pi$, or else, by Lemma 2.13, γ is a closed global geodesic.

Thus, we suppose that $\text{length}\,\alpha > \text{length}\,\beta$. Let $t_0 \in [0,1)$ be such that $\text{length}(\alpha|[0,t_0]) = r$. Let $x = \alpha(0) = \beta(0)$ and $y = \alpha(t_0)$, and let $\alpha_1 : [0,1] \longrightarrow X$ be a linear reparameterisation of $\alpha|[0,t_0]$. Thus, by Lemma 2.16, $\alpha_1 \in \text{geod}(X)$, and so $d(x,y) = r$. Now $\beta \cup -(\alpha|[t_0,1])$ is a path of length r joining x to y, and so may be reparameterised to give $\beta_1 \in \text{geod}(X)$ with $\text{end}\,\beta_1 = (x,y)$. Note that $\gamma = \alpha \cup -\beta = \alpha_1 \cup -\beta_1$. Since α is locally injective near t_0, we must have $\alpha_1 \neq \beta_1$. Now if $r < \pi$, then by Lemma 2.13, γ is a closed global geodesic.

Thus, we are reduced to the case where $r = \pi = \frac{1}{2}\text{sys}(X)$. We have $\alpha_1, \beta_1 \in \text{geod}(X)$, with $\text{end}\,\alpha_1 = \text{end}\,\beta_1 = (x,y)$, and we know that $\alpha_1 \cup -\beta_1$ is locally geodesic at y, and hence everywhere except possibly at x. For $t \in (0,1)$, consider the path $(\alpha|[t,1]) \cup -(\beta|[1-t,1])$. This is a local geodesic of length π, and so by Lemma 2.16, we have $d(\alpha(t), \beta(1-t)) = \pi$. In other words, we have shown that $d(\gamma(u), \gamma(u')) = \pi$ for all pairs of antipodal points $u, u' \in S^1$. Thus, by Lemma 2.12, γ is a global geodesic. ◊

Corollary 2.18 : *If α and β are distinct local geodesics with $\text{end}\,\alpha = \text{end}\,\beta$, then $\text{sys}(X) \leq \text{length}\,\alpha + \text{length}\,\beta$.* ◊

Also, in the case where β is a constant path, we obtain:

Corollary 2.19 : *If α is a non-constant local geodesic with both endpoints equal, then $\text{sys}(X) \leq \text{length}\,\alpha$. Moreover, if $\text{sys}(X) = \text{length}\,\alpha$, then α is a closed global geodesic.* ◊

Of particular interest is the case of a closed local geodesic.

Definition : A map $\gamma : S^1 \longrightarrow X$ is a *closed local geodesic* if for some path-metric σ on S^1, and some $\eta > 0$, we have $d(\gamma(t), \gamma(u)) = \sigma(t,u)$ whenever $t, u \in S^1$ with $\sigma(t,u) \leq \eta$.

Thus $\gamma : (S^1, \sigma) \longrightarrow (X, d)$ is distance non-increasing.

As a special case of Corollary 2.19, we see that any closed local geodesic γ must have length at least $\mathrm{sys}(X)$. If $\mathrm{length}\,\gamma = \mathrm{sys}(X)$, then γ is a global geodesic.

We conclude this chapter by formulating, more carefully, the local convexity property alluded to earlier. We need not assume that X is compact for this.

Suppose that X is l-CAT(1), and $r < l/2$. Suppose $x \in X$ and $y, z \in N(x,r)$. Thus $d(y,z) < l$, and so there is a well-defined triangle $T = ([x \to y], [y \to z], [z \to z])$ with vertices x, y, z. Now $\mathrm{perim}(T) < l$, and so, applying the CAT(1) property, we see that the geodesic $[y \to z]$ maps into $N(x,r)$. We express this by saying:

Lemma 2.20 : *If $r < l/2$, then $N(x,r)$ is convex.* ◇

Thus, the metric d restricted to $N(x,r)$ is a path-metric on $N(x,r)$, and so $N(x,r)$ is intrinsically a (globally) CAT(1)-space.

3. Spaces of loops.

3.1. Introduction.

Suppose X is a connected, complete, locally compact, and locally CAT(1). By a *loop* in X, we mean any continuous map of the circle S^1 into X. We write $\Omega(X)$ for the space of all loops with the sup-norm metric, d_{sup}. Thus, $d_{sup}(\alpha,\beta) = \max\{d(\alpha(t),\beta(t)) \mid t \in S^1\}$. Given $\gamma \in \Omega(X)$, we write $\mathrm{length}\,\gamma \in [0,\infty]$ for the rectifiable length of γ. Thus, the map $\mathrm{length} : \Omega(X) \longrightarrow [0,\infty]$ is lower semi-continuous. Given $r \in [0,\infty)$, we write

$$\Omega(X,r) = \{\gamma \in \Omega(X) \mid \mathrm{length}\,\gamma < r\}.$$

We say that γ is *rectifiable* if $\mathrm{length}\,\gamma < \infty$. (Note that we are not making any assumptions about the parameterisation of a rectifiable loop.)

Suppose $\alpha, \beta \in \Omega(X,r)$. We say that α and β are *r-homotopic* if they lie in the same path-connected component of $\Omega(X,r)$. Note that there is precisely one component containing all the constant loops in X. Of

particular interest is the case $r = 2\pi$. If $\alpha, \beta \in \Omega(X, 2\pi)$, we write $\alpha \sim \beta$ to mean that they are (2π)-homotopic. We write $\alpha \sim 0$, and say that α is *shrinkable* if it is (2π)-homotopic to a constant loop. We write $\alpha \not\sim 0$ for not $\alpha \sim 0$.

The property of shrinkability seems a natural one in this context. We shall show:

Theorem 3.1.1 : *Suppose $x, y \in X$, and that $\alpha_1, \alpha_2, \alpha_3 : [0,1] \longrightarrow X$ be three paths joining x to y. Let $\gamma_i = \alpha_{i+1} \cup -\alpha_{i+2} \in \Omega(X)$. Suppose* $\mathrm{length}\,\gamma_i < 2\pi$ *for all $i \in \{1,2,3\}$. If $\gamma_1 \sim 0$ and $\gamma_2 \sim 0$ then $\gamma_3 \sim 0$.*

Theorem 3.1.2 : *Suppose $T = (\alpha,\beta,\gamma)$ is a triangle in X, with* $\mathrm{perim}(T) < 2\pi$*; so that $\Gamma(T) = \alpha \cup \beta \cup \gamma \in \Omega(X, 2\pi)$. If $\Gamma(T) \sim 0$, then T is CAT(1).*

Theorem 3.1.3 : *Suppose α, β are two local geodesics joining the same pair of points, with* $\mathrm{length}\,\alpha + \mathrm{length}\,\beta < 2\pi$*; so that $\gamma = \alpha \cup -\beta \in \Omega(X, 2\pi)$. Then $\gamma \not\sim 0$.*

An immediate corollary of Theorem 3.1.3 is:

Proposition 3.1.4 : *If $\gamma \in \Omega(X, 2\pi)$ is a closed local geodesic, then $\gamma \not\sim 0$.*

The definition of r-homotopy we have given is quite weak. Note that the map $[\gamma \mapsto \mathrm{length}\,\gamma] : \Omega(X) \longrightarrow [0,\infty]$ is not upper-semicontinuous. Thus, for example, an r-homotopy $[t \mapsto \gamma_t] : [0,1] \longrightarrow \Omega(X,r)$ of two loops might takes us through loops γ_t of length arbitrarily close to r.

We define the stronger notion of *monotone homotopy*:

Definition : Suppose $\alpha, \beta \in \Omega(X)$. We say that α is monotonically homotopic to β, and write $\alpha \searrow \beta$ if there is a path $[t \mapsto \gamma_t] : [0,1] \longrightarrow \Omega(X)$, such that $[t \mapsto \mathrm{length}\,\gamma_t] : [0,1] \longrightarrow [0,\infty]$ is continuous, and $\mathrm{length}\,\gamma_t \leq \mathrm{length}\,\alpha$ for all t.

Thus if $\alpha \searrow \beta$, then $\mathrm{length}\,\beta \leq \mathrm{length}\,\alpha$. If $\mathrm{length}\,\beta < \infty$, we shall demand that $\mathrm{length}\,\gamma_t < \infty$ for all $t > 0$. Note that it is easy to lengthen any intermediate loop γ_t, for example by folding it back on itself, and so there is no loss is assuming that the map $[t \mapsto \mathrm{length}\,\gamma_t]$ is monotonically decreasing, thus justifying the terminology. Most of the monotone homotopies we construct will have this property anyway.

Clearly, $\gamma \searrow 0$ implies $\gamma \sim 0$. In fact we shall show:

Theorem 3.1.5 : *Suppose $\gamma \in \Omega(X, 2\pi)$. Then $\gamma \sim 0$ if and only if $\gamma \searrow 0$.*

In the case where X is compact, we will have in addition:

Theorem 3.1.6 : *Suppose X is compact, and $\gamma \in \Omega(X)$. Then either $\gamma \searrow 0$, or else $\gamma \searrow \alpha$ where α is a closed local geodesic.*

Putting this together with Corollary 2.19, we get:

Corollary 3.1.7 : *Suppose X is compact. If $\gamma \in \Omega(X)$ and* length $\gamma <$ sys(X), *then $\gamma \sim 0$.*

These results give us an alternative definition of the systole of X, as the minimum of 2π and the minimum length of a non-shrinkable loop in $\Omega(X, 2\pi)$. This will also serve as a definition of systole in the case where X is only locally compact (provided we replace "minimum" by "infinum"). In this case (by Theorem 3.1.3), if $r = \frac{1}{2}\text{sys}(X) > 0$, then X is r-CAT(1). In particular, X is globally CAT(1) if and only if sys$(X) = 2\pi$. This is the same as saying that $\Omega(X, 2\pi)$ is path-connected.

There is some degree of arbitrariness in the formulations of r-homotopy and monotone homotopy we have chosen. For example, it would perhaps be more natural to keep track of parameterisations by demanding that all loops and homotopies be lipshitz. It turns out that this approach would lead to essentially the same notions, as we shall observe in Section 3.5. Similarly, we could restrict attention to smooth maps in the riemannian category, or to piecewise linear maps in the context euclidean polyhedral complexes.

In order to prove the results stated in this section, we shall reduce ourselves to considering polygonal loops to which we can apply the "Birkhoff curve-shortenning process". If X is l-CAT(1), and we start with a polygonal (i.e. piecewise geodesic) loop, each of whose geodesic segments has length less than l, then we can attempt to shorten it by cyclically joining the midpoints of each segment. The Birkhoff process is the iteration of this procedure. If X is compact, then some subsequence must converge, either to a point, or to a closed local geodesic. If the length of the the original

loop is $< 2\pi$, then we will get convergence to a point if and only if this loop is shrinkable.

It seems to be an interesting question as to when the Birkhoff process converges without having to pass to a subsequence. We give a brief discussion of this in Section 3.7. It is not relevant to the rest of the paper.

In proving the results given here, it will be convenient to deal first with the case where X is l-CAT(1) for some $l > 0$ (i.e. "uniformly locally CAT(1)"). We describe the Birkhoff process in Section 3.3, and give complete proofs in Section 3.4. In Section 3.6, we describe how to deal with the case where X is not l-CAT(1). We begin, in Section 3.2, with a discussion of cartesian products of CAT(1) spaces, which provides a convenient means of overcoming a technical difficulty in Section 3.3.

3.2. Cartesian products.

Suppose (X_1, d_1) and (X_2, d_2) are complete locally compact path-metric spaces. Then so is (X, d), where $X = X_1 \times X_2$ and $d((x_1, x_2), (y_1, y_2)) = \sqrt{(d_1(x_1, y_1)^2 + d_2(x_2, y_2)^2)}$.

Proposition 3.2.1 : *If (X_1, d_1) and (X_2, d_2) are r-CAT(1), then so is (X, d).*

Proof : Let $p_i : X_1 \times X_2 \longrightarrow X_i$ be projection to X_i. Suppose $T = (\alpha, \beta, \gamma)$ is a triangle in X with perim$(T) < 2r$. Thus $T_i = (p_i \circ \alpha, p_i \circ \beta, p_i \circ \gamma)$ is a triangle in X_i. Let $T_i' = (\alpha_i', \beta_i', \gamma_i')$ be a comparison triangle in S^2 for T_i. Let $\alpha' = [t \mapsto (\alpha_1'(t), \alpha_2'(t))] : [0,1] \longrightarrow S^2 \times S^2$. Similarly define β' and γ', so that $T = (\alpha', \beta', \gamma')$ is a triangle in $S^2 \times S^2$. Let ρ be the product riemannian metric on $S^2 \times S^2$. With respect to this metric, we have perim$(T') =$ perim$(T) < 2r$, and the natural map $(\Gamma(T), d) \longrightarrow (\Gamma(T'), \rho)$ is distance non-increasing. Now $(S^2 \times S^2, \rho)$ is a riemannian manifold of curvature ≤ -1, with no geodesics of length less that 2π, and is therefore π-CAT(1). Thus, T' satisfies CAT(1), and so therefore does T. ◇

3.3. Polygonal curve shortening.

In this section, we assume that X is locally compact, and l-CAT(1) for some $l > 0$.

We are interested in polygonal loops, which we can think of formally as cyclically ordered n-tuples of points of X. We describe the Birkhoff process of shortenning such loops. We are principally interested in the question of when the Birkhoff process converges to a point.

We write $C(X) = X^n$ for the set of n-tuples $\underline{x} = (x_1, x_2, \ldots, x_n)$, which we think of as cyclically ordered so that $x_{n+1} = x_1$. Throughout this section, n will be constant, and so will not feature in our notation. We define the maps $M, L, E : C(X) \longrightarrow [0, \infty)$ by

$$M(\underline{x}) = \max\{d(x_i, x_{i+1}) \mid i = 1, \ldots, n\}$$
$$L(\underline{x}) = \sum_{i=1}^{n} d(x_i, x_{i+1})$$
$$E(\underline{x}) = \sum_{i=1}^{n} d(x_i, x_{i+1})^2.$$

Thus we may think of $M(\underline{x})$, $L(\underline{x})$ and $E(\underline{x})$, as the "mesh", "length" and "energy" of $\underline{x}$ respectively. Note that if $M(\underline{x}) \leq l$, then we may join consecutive points x_i and x_{i+1} by a unique geodesic in X, thus justifying our interpretation of $\underline{x}$ as a polygonal loop.

Lemma 3.3.1 : *For all $\underline{x} \in C(X)$, we have*

$$E(\underline{x}) \leq L(\underline{x})^2 \leq nE(\underline{x}).$$

◇

We say that $\underline{x}$ is *constant* if $\underline{x} = (x, x, \ldots, x)$ for some $x \in X$. Thus, $\underline{x}$ is constant if and only if $L(\underline{x}) = 0$, or if and only if $E(\underline{x}) = 0$.

Given $r > 0$, let

$$C(X, r) = \{\underline{x} \in C(X) \mid L(\underline{x}) < r\}.$$

Given $h > 0$, let

$$C_h(X) = \{\underline{x} \in C(X) \mid M(\underline{x}) < h\}.$$

Let $C_h(X, r) = C_h(X) \cap C(X, r)$. Each of these sets is open in $C(X)$.

We define the Birkhoff process on $C_h(X)$, where $h < l$.

Suppose $x, y \in X$ with $d(x, y) < l$. Let $\text{mid}(x, y)$ be the midpoint of the unique geodesic $[x \to y]$ joining x to y. Given $\underline{x} \in C_h(X)$, we write $\Gamma(\underline{x})$ for the piecewise geodesic loop $\bigcup_{i=1}^{n} [x_i \to x_{i+1}]$, which we think of as a map $S^1 \longrightarrow X$ parameterised proportionately to arc-length on each segment (see Section 3.4). We define $f(\underline{x}) \in C_h(X)$ by

$$f(\underline{x}) = (\text{mid}(x_1, x_2), \text{mid}(x_2, x_3), \ldots, \text{mid}(x_n, x_1)).$$

Thus, by Lemma 2.10, we have:

Lemma 3.3.2 : *The map $f : C_h(X) \longrightarrow C_h(X)$ is continuous.* ◇

The following are easily verified.

Lemma 3.3.3 : *If $\underline{x} \in C_h(X)$, then $M(f(\underline{x})) \leq M(\underline{x})$, $L(f(\underline{x})) \leq L(\underline{x})$ and $E(f(\underline{x})) \leq E(\underline{x})$.* ◇

Lemma 3.3.4 : *The following are equivalent:*
(1) $f^{2n}(\underline{x}) = \underline{x}$,
(2) $E(f(\underline{x})) = E(\underline{x})$,
(3) $d(x_i, x_{i+1}) = d(x_j, x_{j+1})$ for all $i, j \in \{1, 2, \ldots, n\}$ and $\Gamma(\underline{x})$ is either constant or a closed local geodesic. ◇

The idea of the Birkhoff process is to iterate the map f, in the hope that we converge on a point or a closed local geodesic. Clearly, if f were compact, then at least some subsequence must converge. For the moment, we are interested in convergence to a point.

Suppose $x \in X$. By Lemma 3.3.3, the limits $L^\infty(\underline{x}) = \lim_{r\to\infty} L(f^r(\underline{x}))$ and $E^\infty(\underline{x}) = \lim_{r\to\infty} E(f^r(\underline{x}))$ must exist. By the inequality, Lemma 3.3.1, we have that $L^\infty(\underline{x}) = 0$ if and only if $E^\infty(\underline{x}) = 0$. We write

$$C_h^0(X) = \{\underline{x} \in C_h(X) \mid L^\infty(\underline{x}) = 0\}.$$

Suppose that $r < l/2$ and $\underline{x} = (x_1, x_2, \ldots, x_r) \in C_h(X, 2\pi)$. By Lemma 2.20, the ball $N(x_1, r)$ is compact and convex, and thus intinsically CAT(1). We see, by induction, that $\Gamma(f^k(\underline{x})) \subseteq N(x_1, r)$ for all $k \in \mathbf{N}$. Moreover, some subsequence $f^{k_i}(\underline{x})$ must converge on some $\underline{y} \in C_h(X, r)$. Clearly, $E(f(\underline{y})) = E(\underline{y})$ and $L(\underline{y}) < l$. Since X contains no closed geodesic of length $< 2\pi$, Lemma 3.3.4 tells us that $\underline{y}$ must be constant. It follows that $L^\infty(\underline{x})$ and so $\underline{x} \in C_h^0(X)$. From this we conclude:

Lemma 3.3.5 : *Suppose $x \in C_h(X)$. Then, $x \in C_h^0(X)$ if and only if for some $m \in \mathbf{N}$ we have $L(f^m(\underline{x})) < l$.* ◇

Note that if $\underline{x} \in C_h^0(X)$, we may apply the above argument with r arbitrarily small. In other words, we can find balls $N(a_i, r_i)$ with $r_i \to 0$ so that $\Gamma(f^k(\underline{x})) \subseteq N(a_i, r_i)$ for all sufficiently large k (depending on i). From this we see that the sequence $\Gamma(f^r(\underline{x}))$ must converge to a point (without passing to a subsequence). Thus:

Lemma 3.3.6 : *If $x \in C_h^0(X)$, then $f^k(\underline{x})$ converges to a constant cycle in $C_h^0(X)$.* ◇

We have no direct use for Lemma 3.3.6 in this section. We discuss the matter some more in Section 3.7.

As a corollary of Lemma 3.3.5, we have:

Lemma 3.3.7 : *$C_h^0(X)$ is open in $C_h(X)$.*

Proof : If $\underline{x} \in C_h^0(X)$, then there is some m such that $L(f^m(\underline{x})) < l/2$. Now f^m is continuous (Lemma 3.3.2), and so for all $\underline{y} \in C_h(X)$ sufficiently close to $\underline{x}$, we have $L(f^m(\underline{y})) < l$. Thus, by Lemma 3.3.5, $\underline{y} \in C_h^0(X)$. ◇

A trivial example of the Birkhoff process is obtained by starting with a regular n-gon in the euclidean plane $\mathbf{E}^2$. Thus, identifying $\mathbf{E}^2$ with the complex plane, we set $\underline{x} = (x_1, \ldots, x_n) = (r, re^{2\pi i/n}, \ldots, re^{2\pi i(n-1)/n})$ so that r is the circumradius of the polygon. The Birkhoff process shrinks the polygon homothetically (on even iterations) to its centre. (Figure 3a.) Thus $f^k(\underline{x}) = (x_1 e^{\pi ik/n}\cos(\pi k/n), \ldots, x_n e^{\pi ik/n}\cos(\pi k/n))$.

We see from Lemma 3.3.7, that $C_h^0(X, 2\pi)$ is open in $C_h(X, 2\pi)$. We aim to prove that it is also closed in $C_h(X, 2\pi)$. Clearly, it is enough to show that $C_h^0(X, r)$ is closed in $C_h(X, r)$ for all $r < 2\pi$. This would follow by a similar argument to Lemma 3.3.7, provided we can achieve some kind of uniformity in the integer m involved in the proof. Specifically, we aim to show:

Proposition 3.3.8 : *Given $0 < r_0 \leq r < 2\pi$, there is some $m \in \mathbf{N}$ such that if $\underline{x} \in C_h^0(X, r)$, then $L(f^m(\underline{x})) < r_0$.*

Closedness (Theorem 3.3.15) then follows by Lemmas 3.3.2 and 3.3.5 and Proposition 3.3.8.

The idea of the proof of Proposition 3.3.8 is to show that, at each stage, the energy must decrease by a definite amount, depending (continuously) only on the length of the curve at that stage:

Lemma 3.3.9 : *There is a continuous function $\lambda : (0, 2\pi) \longrightarrow (0, \infty)$ such that if $x \in C_h^0(X, 2\pi)$, then $E(f(\underline{x})) \leq E(\underline{x}) - \lambda(L(\underline{x}))$.*

(Here λ depends on n but not on h.) We must have $\lambda(t) \to 0$ as $t \to 0$ and as $t \to 2\pi$. However, on any closed interval $[r_0, r] \subseteq (0, 2\pi)$, the function l will be bounded below by some constant $\delta > 0$. Thus, if $\underline{x} \in C_h^0(X, r)$, then $E(\underline{x}) \leq L(\underline{x})^2 \leq r^2$, and so for $m \geq r^2/\delta$, we have that $L(\underline{x}) < r_0$. (Otherwise, applying Lemma 3.3.9 inductively, we find that $E(f^m(\underline{x})) \leq r^2 - m\delta \leq 0$.) This shows that Proposition 3.3.8 follows from Lemma 3.3.9

For the proof of Lemma 3.3.9, we need some general lemmas:

Lemma 3.3.10 : *For all $\epsilon > 0$, there exists $\eta > 0$ such that the following holds.*

Suppose (Y, ρ) is a π-CAT(1) space. If $\gamma : S^1 \longrightarrow X$ is a rectifiable loop of length at most $2(\pi - \epsilon)$, then there is some $a \in X$ such that $\gamma(S^1) \subseteq N(a, \frac{\pi}{2} - \eta)$.

Proof : Given $\epsilon > 0$, choose η as in Lemma 1.3. Choose $t, u \in S^1$ so as to divide γ into two subarcs of length at most $\pi - \epsilon$. Let $y = \gamma(t)$ and $z = \gamma(u)$. Thus $d(y, z) \leq \pi - \epsilon$. Let a be the midpoint of the geodesic $[y \to z]$. Now, if $x \in \gamma(S^1)$, then $d(y, x) + d(x, z) \leq \pi - \epsilon$, and so the triangle $T = ([x \to y], [y \to z], [z \to x])$ has $\text{perim}(T) \leq 2(\pi - \epsilon)$. Thus T is CAT(1), and applying Lemma 1.3, we have $d(a, x) \leq \frac{\pi}{2} - \eta$. ◇

Lemma 3.3.11 : *Given $\eta, \mu > 0$ with $\mu < 2\eta$, then there is some $\delta > 0$ such that the following holds.*

Suppose (Y, ρ) is π-CAT(1), and that $a, x, y, z \in Y$ satisfy $\rho(a, x) \leq \frac{\pi}{2} - \eta$, $\rho(a, y) \leq \rho(a, x)$, $\rho(a, z) \leq \rho(a, x)$ and $\rho(x, z) = \rho(x, y) = \mu$. Then, $\rho(y, z) \leq 2\mu - \delta$.

Proof : We construct a quadrilateral $Q = (a', y', x', z')$ in (S^2, d_1) so that (a', x', z') and (a', x', y') are comparison triangles for (a, x, z) and (a, x, y) lying on opposite sides of the geodesic $[a' \to x']$. Since $\text{perim}(a', x', y') < \pi$, and $d_1(a', y') \leq d_1(a', x')$, we see that $a'\hat{x}'y' < \pi/2$. Similarly, $a'\hat{x}'z' < \pi/2$, and so $\angle(Q, x') < \pi$. In fact, given that $d_1(x', y') = d_1(x', z') = \mu$, we see that $\angle(Q, x')$ is bounded away from π by an amount depending on η and μ. Thus there is some $\delta > 0$, such that the distance between y' and z' (in the induced path-metric on $R(Q)$) is at most $2\mu - \delta$. The result follows by applying CAT(1) to the triangles (a, x, y) and (a, x, z). ◇

Note that in the above lemmas, η can be assumed to depend continuously on ϵ, and δ to depend continuously on η and μ.

In the proof of Lemma 3.3.9, we need to use a particular topological cellulation of the disc D. Thus, given $m, n \in \mathbf{N}$, we define a cellulation $G = G(m, m)$ of the disc, as shown in Figure 3b, for $m = 3$ and $n = 5$. The label ij refers to the vertex $v(i, j)$ of G.

The cellulation $G(m, n)$ may be described more formally in terms of the Birkhoff process applied to a regular n-gon in the euclidean plane. Let $\underline{v} \in C(\mathbf{E}^2)$ be the cycle $\underline{v} = (v(0,1), \ldots, v(0,n))$ of vertices of a regular n-gon. For $i \in \{1, \ldots, m\}$, let $f_0^i(\underline{v}) = (v(i,1), \ldots, v(i,n)) \in C(\mathbf{E}^2)$, where $f_0 : C(\mathbf{E}^2) \longrightarrow C(\mathbf{E}^2)$ is Birkhoff curve shortening on $\mathbf{E}^2$, which we have referred to earlier. We shall write $V(i) = \{v(i,j) \mid j = 1, \ldots, m\}$. We can identify the disc D with the convex hull of $V(0)$.

For $(i, j) \in \{0, \ldots, m-1\} \times \{1, \ldots, n\}$, we write $T(i, j)$ for the triangular convex hull of $\{v(i,j), v(i+1, j-1), v(i+1, j)\}$. For $j \in \{3, \ldots, n\}$, we write $T(m, j)$ for the convex hull of $\{v(m, 1), v(m, j-1), v(m, j)\}$. We see that there is a well-defined cellulation $G(m, n)$ of the disc D, with vertex set $V = \bigcup_{i=1}^{m} V(i)$, and with the set of 2-cells equal to $\{T(i, j)\}$. For each triangular 2-cell T, we refer to its set of extreme points, $V(T)$, the *principal vertices* of T. We write $F_i(G) = V$ for the set of i-cells of G. Thus $F_0(G) = V$ and $F_2(G) = \{T(i, j)\}$. Let $K_1(G) = \bigcup F_1(G) \subseteq D$ be the 1-skeleton of G.

We now turn to the proof of Lemma 3.3.9. Suppose $\underline{x} \in C_h^0(X, 2\pi)$. There is some $m \in \mathbf{N}$ such that $L(f^m(\underline{x})) < 2l$. Let $G = G(m, n)$. For $i \in \{0, \ldots, m\}$, write $f^i(\underline{x}) = (x(i,1), \ldots, x(i,n))$. We define a map $g : F_0(G) \longrightarrow X$ by setting $g(v(i,j)) = x(i,j)$.

We want to use the map g to define a path-metric ρ on the disc D. Suppose $T \in F_2(G)$ has principal vertices $V(T) = \{v_1, v_2, v_3\}$. For each such T, we may take a spherical comparison triangle for (gv_1, gv_2, gv_3).

The idea then is to glue these spherical triangles together as dictated by the combinatorics of G, thus obtaining a singular spherical metric on D. Unfortunately, we run into the technical problem that some of the comparison triangles may be degenerate. For the moment we shall define this problem away, and worry about how to deal with the degenerate case later. Let us assume that:

(*) For all $T \in F_2(G)$, and $i \in \{1, 2, 3\}$, we have that $d(gv_i, gv_{i+2}) < d(gv_i, gv_{i+1}) + d(gv_{i+1}, gv_{i+2})$ where $V(T) = \{v_i, v_2, v_3\}$.

In other words, all the triangle inequalities are strict, so we can construct a non-degenerate comparison triangle T' for T in S^2. Let $R(T)$ be the small triangular region bounded by T (Section 1). Now all the pieces $\{R(T) \mid T \in F_2(G)\}$ fit together nicely to give a singular spherical path-metric ρ on D. The boundary ∂D is a piecewise geodesic loop with vertices $V \cap \partial D = V(0) \cup V(1)$. At each interior vertex of $V \backslash \partial D$, we have a cone singularity. If $v \in V \cap \partial D$, we write $\angle(\partial D, v)$ for the interior angle of ∂D at v. If $v \in V \backslash \partial D$, we write $\angle(D, v)$ for the cone angle at v.

We may extend $g : V = F_0(G) \longrightarrow X$ to a map of the 1-skeleton $g : K_1(G) \longrightarrow X$, by mapping each edge of $F_1(G)$ linearly to the corresponding geodesic segment in X. Thus, the boundary, ∂T, of each triangle $T \in F_2(G)$ gets mapped to a triangle of perimeter less than $h < l$. Thus, applying CAT(1), we see that $g|(\partial T, \rho) \longrightarrow (X, d)$ is distance non-increasing. Since ρ is a path-metric on D, we obtain:

Lemma 3.3.12 : *The map $g : (K_1(G), \rho) \longrightarrow (X, d)$ is distance non-increasing.* ◇

We now claim:

Lemma 3.3.13 : *(D, ρ) is π-CAT(1).*

Proof : First, we show that (D, ρ) is locally CAT(1). Secondly, we show that (D, ρ) contains no simple closed geodesic. The result then follows by Theorem 2.15.

For the first part, we need to know that each interior cone-angle, $\angle(D, v)$, for $v \in V \backslash \partial D$, is at least 2π. (See the discussion of polyhedral complexes in the introduction). This is in turn equivalent to saying that each such vertex lies in the interior of a ρ-geodesic segment. Suppose

$v \in V \backslash \partial D$, so that $v = v(i,j)$ for some $(i,j) \in \{2,\dots,m\} \times \{1,\dots,n\}$. Let $v_1 = v(i-1,j)$ and $v_2 = v(i-1,j+1)$. Thus v_1 and v_2 are adjacent to v in $K_1(G)$, and gv is the midpoint of gv_1 and gv_2. Applying Lemma 3.3.12, we find that $\rho(v_1,v_2) \geq d(gv_1,gv_2) = d(gv_1,gv) + d(gv,gv_2) \geq \rho(v_1,v) + \rho(v,v_2)$. Thus, v lies in the interior of the ρ-geodesic segment joining v_1 to v_2, and so $\angle(D,v) \geq 2\pi$ as required.

For the second part, suppose, for contradiction that $\gamma \subseteq D$ is a simple closed geodesic. Thus, γ bounds an open disc $D_0 \subseteq D$. From the construction, it is clear by induction that $V \subseteq D \backslash D_0$. It follows that γ lies inside a spherical triangular region corresponding to some $T \in F_2(G)$ which is impossible. ◇

Note that the first part of the above proof also works to show that each vertex in $V(1)$ lies in the interior of a ρ-geodesic segment. We deduce:

Lemma 3.3.14 : *If $v \in V(1)$, then $\angle(\partial D, v) \geq \pi$* ◇

From the construction, if $v \in V(0)$, then $\angle(\partial D, v) < \pi$.

We recall our objective of showing that $E(\underline{x}) - E(f(\underline{x}))$ is bounded below by some positive continuous function of $L(\underline{x})$.

Set $r = \frac{1}{2}L(\underline{x})$ and $\Delta = E(\underline{x}) - E(f(\underline{x})) \geq 0$. We can thus, without loss of generality, imagine Δ to be small. This implies that for each $i \in \{1,\dots,n\}$, $d(x_j,x_{j+1})$ is close to $2r/n$. To be more precise, set $\xi_j = d(x_j,x_{j+1}) = \rho(v(0,j),v(0,j+1))$. Then $E(\underline{x}) = \sum_{j=1}^n \xi_j^2$ and $E(f(\underline{x})) \leq \sum_{j=1}^n \left(\frac{\xi_j+\xi_{j+1}}{2}\right)^2$. Thus, $\sum_{j=1}^n (\xi_{j+1}-\xi_j)^2 \leq 4\Delta$ and so $|\xi_{j+1}-\xi_j| \leq 2\sqrt{\Delta}$ for all j. Since $\sum_{j=1}^n \xi_j = 2r$, we deduce that $|\xi_j - \frac{2r}{n}| \leq n\sqrt{\Delta}$ for all j. In particular, if $\Delta \leq r^2/n^4$, then $\xi_j \geq r/n$.

Let $\epsilon = \pi - r > 0$. Let η be the constant of Lemma 3.3.10 (given ϵ). Let $\mu = \min(r/2n, \eta)$ and let δ be the constant of Lemma 3.3.11 (given η and μ). We have that length$(\partial D, \rho) = L(\underline{x}) = 2r = 2(\pi - \epsilon)$, and so, by Lemma 3.3.10, there is some $a \in D$ such that $\rho(a,y) \leq \pi - \eta$ for all $y \in \partial D$. Let $b \in \partial D$ be a point of ∂D furthest from a. Now, ∂D is a piecewise geodesic loop with vertices $V \cap \partial D = V(0) \cup V(1)$. Since (D,ρ) is π-CAT(1), we see easily that b must be the a vertex. In fact, using Lemma 3.3.14, we must have $b \in V(0)$. In other words, $b = v(0,k+1)$ for some $k \in \{1,\dots,n\}$.

Let $v_1 = v(1,k)$ and $v_2 = v(1,k+1)$. Thus $v_1, v_2 \in V(1)$ are adjacent vertices to b. Now, $\rho(b,v_1) = \rho(b,v_2) = \xi_k/2 \geq r/2n \geq \mu$. Similarly,

$\rho(b,v_2) \geq \mu$. Let y_1 and y_2 be points on the geodesic segments $[b,v_1]$ and $[b,v_2]$ respectively, such that $\rho(b,y_1) = \rho(b,y_2) = \mu$. Now $[b,v_1] \cup [b,v_2] \subseteq \partial D$, so $\rho(a,y_i) \leq \rho(a,b) \leq \pi - \eta$. Also $\mu \leq 2\eta$, and so by Lemma 3.3.11, we have $\rho(y_1,y_2) \leq 2\mu - \delta$. We deduce that $\rho(v_1,v_2) \leq \rho(b,v_1) + \rho(b,v_2) - \delta = \frac{1}{2}(\xi_k + \xi_{k+1}) - \delta$.

If we set $\zeta_j = \rho(v(1,j),v(1,j+1)) = d(x(1,j),x(1,j+1))$, we obtain $E(f(\underline{x})) = \sum_{j=1}^n \zeta_j^2$. We know that $\zeta_j \leq \frac{1}{2}(\xi_j + \xi_{j+1})$ for all j, and that $\zeta_k \leq \frac{1}{2}(\xi_k + \xi_{k+1}) - \delta$. Thus $\Delta = E(\underline{x}) - E(f(\underline{x})) = \sum_{j=1}^n \xi_j^2 - \sum_{j=1}^n \zeta_j^2 \geq \frac{1}{4}\sum_{j=1}^n (\xi_j - \xi_{j+1})^2 + \delta(\xi_k + \xi_{k+1}) - \delta^2 \geq \frac{1}{2}\delta(\xi_k + \xi_{k+1}) \geq \frac{1}{2}\delta(2r/n) = \delta r/n$.

We derived this inequality under the assumption that $\Delta \leq r^2/n^4$. In other words, either $\Delta \geq \delta r/n$ or else $\Delta \geq r^2/n^4$. So we have shown that $\Delta \geq \lambda = \min(\delta r/n, r^2/n^4)$. We see from Lemmas 3.3.10 and 3.3.11 that δ, and hence λ, can be assumed to vary continuously in $r = L(\underline{x})$.

We have thus proven Lemma 3.3.9 under the assumption of strict triangle inequalities $(*)$.

To deal with the general case, the idea is to take a cartesian product with a small regular euclidean polygon. We described earlier the Birkhoff process applied to such a polygon. In particular, we see that it must satisfy $(*)$.

Suppose Y is a π-CAT(1) space. Then Lemma 3.2.1 tells us that $X \times Y$ is l-CAT(1). If $\underline{x} = (x_1,\dots,x_n) \in C_h(X)$ and $\underline{y} = (y_1,\dots,y_n) \in C_\eta(Y)$, we write $(\underline{x},\underline{y}) \in C_{h+\eta}(X \times Y)$ for the cycle $((x_1,y_1),\dots,(x_n,y_n))$. We see that $f(\underline{x},\underline{y}) = (f(\underline{x}), f(\underline{y}))$, where f is used to represent Birkhoff curve shortening on X, Y and $X \times Y$. In particular, if $\underline{x} \in C_h^0(X)$ and $\underline{y} \in C_\eta^0(Y)$, then $(\underline{x},\underline{y}) \in C_{h+\eta}^0(X \times Y)$.

Now take $Y = \mathbf{E}^2$ to be the euclidean plane, and let $\underline{y} \in C_\eta(\mathbf{E}^2)$ be a regular n-gon of small circumradius $\epsilon > 0$. Then $(\underline{x},\underline{y}) \in C_{h+\eta}^0(X \times \mathbf{E}^2)$ satisfies condition $(*)$. Thus, we have that

$$E(\underline{x},\underline{y}) - E(f(\underline{x}), f(\underline{y})) \geq \lambda(L(\underline{x},\underline{y})).$$

We now let $\epsilon \to 0$, and deduce in the limit that

$$E(\underline{x}) - E(f(\underline{x})) \geq \lambda(L(\underline{x})).$$

This finally concludes the proof of Lemma 3.3.9, and thus also of Proposition 3.3.8.

Theorem 3.3.15 : *Suppose X is locally compact and l-CAT(1) for some $l > 0$. If $h < l$, then $C_h^0(X, 2\pi)$ is open and closed in $C_h(X, 2\pi)$.*

Proof : By definition, $C_h^0(X, 2\pi) = C_h^0(X) \cap C_h(X, 2\pi)$, so Lemma 3.3.7 tells us that $C_h^0(X, 2\pi)$ is open in $C_h(X, 2\pi)$.

Suppose, then, that the sequence $\underline{x}_i \in C_h^0(X, 2\pi)$ converges to some $\underline{x} \in C_h(X, 2\pi)$. Since $L : C(X) \longrightarrow [0, \infty)$ is continuous, we can assume that $\underline{x}_i \in C_h^0(X, r)$ for some fixed $r < 2\pi$. By Proposition 3.3.8, there is some $m \in \mathbf{N}$ such that $L(f^m(\underline{x}_i)) \leq l/2$ for all i. Since f and L are continuous, it follows that $L(f^m(\underline{x})) \leq l/2 < l$, and so, by Lemma 3.3.5, $x \in C_h^0(X, 2\pi)$. This shows that $C_h^0(X, 2\pi)$ is closed in $C_h(X, 2\pi)$. ◇

3.4. Shrinkable loops.

As in the previous section, we suppose that X is locally compact and l-CAT(1). Under this assumption, we shall prove the results 3.1.1–3.1.5 described in Section 3.1. (We shall describe how to deal with general locally CAT(1) spaces in Section 3.6.) In the case where X is compact, we deduce 3.1.6 and 3.1.7.

In Section 3.1, we defined $\Omega(X)$ as the space of loops $S^1 \longrightarrow X$, and $\Omega(X, r)$ as the subspace of rectifiable loops of length strictly less than r. We defined the equivalence relation $\sim$ of (2π)-homotopy on $\Omega(X, 2\pi)$, and the transitive relation $\searrow$ of monotone homotopy on $\Omega(X)$. In these definitions, we have made only topological hypotheses. In Section 3.5, we show that we can restrict attention to lipshitz maps and homotopies to the same effect.

Note that the definition of monotone homotopy also makes sense for paths with fixed endpoints. Thus, if $\alpha, \beta : [0, 1] \longrightarrow X$ both join x to y in X, then we write $\alpha \searrow \beta$ if there is a homotopy $[t \to \gamma_t]$ with $\gamma_0 = \alpha$, $\gamma_1 = \beta$ and $\gamma_t(0) = x$ and $\gamma_t(1) = y$ for all t and such that $[t \mapsto \text{length}\, \gamma_t]$ is continuous and monotonically decreasing.

Lemma 3.4.1 : *Suppose $x, y, z \in X$ with $d(x, y) < l$, $d(y, z) < l$ and $d(x, z) < l$. Then $[x \to y] \cup [y \to z] \searrow [x \to z]$.*

Proof : Let $\gamma_t = [x \to \beta(t)] \cup (\beta|[t, 1])$. ◇

Suppose $\underline{x} = (x_1, \ldots, x_n) \in C_h(X)$, where $h < l$ (Section 3.3). We write $\Gamma(\underline{x})$ for the piecewise geodesic loop obtained by joining together the segments $[x_i \to x_{i+1}]$ for $i = 1, \ldots, n$. To be more precise, we divide S^1 into n equal segments I_i (with respect to the standard path-metric on S^1), and map I_i linearly onto the image of $[x_i \to x_{i+1}]$. In this way $\Gamma : C_h(X) \longrightarrow \Omega(X)$ is continuous.

We draw the following corollaries to Lemma 3.4.1.

Lemma 3.4.2 : *Suppose $\gamma \in \Omega(X)$. Suppose $t_1, \ldots, t_n \in S^1$ divide S^1 into n segments $J_1, \ldots, J_n$ such that* $\text{length}(\gamma|J_i) < l$ *for all $i \in \{1, \ldots, n\}$. Then $\gamma \searrow \Gamma(\gamma(t_1), \ldots, \gamma(t_n))$.*

Proof : Choose any homeomorphism $\phi : \mathbf{R}/n\mathbf{Z} \longrightarrow S^1$ such that $\phi(i) = t_i$. Given $r \in \mathbf{N}$, let $\underline{x}_r = (\gamma \circ \phi(\frac{1}{2^r}), \gamma \circ \phi(\frac{2}{2^r}), \ldots, \gamma \circ \phi(\frac{2^r n}{2^r}))$. Thus $\underline{x}_r$ is a cycle of $2^r m$ points of X, with $M(\underline{x}_r) < l$ and $M(\underline{x}_r) \to 0$ as $r \longrightarrow \infty$. Applying Lemma 3.4.1, we see that $\Gamma(\underline{x}_{r+1}) \searrow \Gamma(\underline{x}_r)$ for all r. Also, since $\phi([0, n] \cap \mathbf{Z}[\frac{1}{2}])$ is dense in S^1, we see that $\Gamma(\underline{x}_r)$ tends to γ in $\Omega(X)$ as $r \to \infty$. We thus split the interval $[0, 1]$ into subintervals $[\frac{1}{2^{r+1}}, \frac{1}{2^r}]$ and string together these monotone homotopies so as to obtain a monotone homotopy $[t \mapsto \gamma_t]$ with $\gamma_{1/2^r} = \Gamma(\underline{x}_r)$ and $\gamma_0 = \gamma$. We conclude that $\gamma \searrow \Gamma(\underline{x}_0) = \Gamma(\gamma(t_1), \ldots, \gamma(t_n))$. ◇

Let $f : C_h(X) \longrightarrow C_h(X)$ be the Birkhoff curve shortening map. The following is immediate from Lemma 3.4.1.

Lemma 3.4.3 : *If $\underline{x} \in C_h(X)$, then $\Gamma(\underline{x}) \searrow \Gamma(f(\underline{x}))$.* ◇

Lemma 3.4.4 : *If $\underline{x} \in C_h(X)$ with $L(\underline{x}) < 2l$, then $\Gamma(\underline{x}) \searrow 0$.*

Proof : We have $d(x_0, x_i) < l$ for all i. By Lemma 3.4.1, $\Gamma(x_1, \ldots, x_{i+1}) \searrow \Gamma(x_1, \ldots, x_i)$ for all $i \in \{1, \ldots, n-1\}$. Thus, by induction, $\Gamma(\underline{x}) \searrow 0$ ◇

Lemma 3.4.5 : *If $\underline{x} \in C_h^0(X)$, then $\Gamma(\underline{x}) \searrow 0$.*

Proof : By the definition of $C_h^0(X)$, there is some $m \in \mathbf{N}$ such that $L(f^m(\underline{x})) < 2l$. By Lemmas 3.4.3 and 3.4.4, we have $\Gamma(\underline{x}) \searrow \Gamma(f^m(\underline{x})) \searrow 0$. ◇

We next turn to homotopies. We want to relate r-homotopies in $\Omega(X, r)$ to connectedness in $C_h(X, r)$. We are really only interested in the case $r = 2\pi$. Let σ be the standard path-metric S^1.

Lemma 3.4.6 : *Suppose $\alpha, \beta \in \Omega(X, 2\pi)$ and $\alpha \sim \beta$. Then, there is some $\delta > 0$ such that if $u_1, \ldots, u_n \in S^1$ divide S^1 into segments of σ-length $\leq \delta$, then $(\alpha(u_1), \ldots, \alpha(u_n))$ and $(\beta(u_1), \ldots, \beta(u_n))$ are connected by a path in $C_h(X, 2\pi)$.*

Proof : Let $[t \mapsto \gamma_t] : [0,1] \longrightarrow \Omega(X, 2\pi)$ be the homotopy joining α to β. By compactness, there is some $\delta > 0$, such that if $u, u' \in S^1$ with $\sigma(u, u') < \delta$, then $d(\gamma_t(u), \gamma_t(u')) < h$ for all $t \in [0,1]$. Thus if $u_1, \ldots, u_n$ cut S^1 into segments of length at most δ, we have that $\underline{x}(t) = (\gamma_t(u_1), \ldots, \gamma_t(u_n)) \in C_h(X)$. Also $L(\underline{x}(t)) \leq \text{length}(\gamma_t) < 2\pi$. Thus $[t \mapsto \underline{x}(t)]$ gives the desired path in $C_h(X, 2\pi)$. ◊

Lemma 3.4.7 : *Suppose $\gamma \in \Omega(X, 2\pi)$ and $\gamma \sim 0$. Then there is some $\delta > 0$ such that if $u_1, \ldots, u_n \in S^1$ divide S^1 into segments of σ-length at most δ, then $(\gamma(u_1), \ldots, \gamma(u_n)) \in C_h^0(X, 2\pi)$.*

Proof : Apply Lemma 3.4.6, with $\alpha = \gamma$ and β a constant path. We see that $\underline{x} = (\gamma(u_1), \ldots, \gamma(u_n))$ is connected to a constant cycle by a path in $C_h(X, 2\pi)$. Now all constant paths lie in $C_h^0(X, 2\pi)$, and $C_h^0(X, 2\pi)$ is open and closed in $C_h(X, 2\pi)$ (Theorem 3.3.15). Thus $\underline{x} \in C_h^0(X, 2\pi)$. ◊

We can now deduce theorem 3.1.5, namely if $\gamma \in \Omega(X, 2\pi)$ and $\gamma \sim 0$, then $\gamma \searrow 0$.

Proof of Theorem 3.1.5 : Suppose $\gamma \in \Omega(X, 2\pi)$ and $\gamma \sim 0$. Let δ be as given by Lemma 3.4.7. Let $u_1, \ldots, u_n \in S^1$ divide S^1 into segments J_i such that each has σ-length at most δ, and such that $\text{length}(\gamma|J_i) \leq h$ for all i. Let $\underline{x} = (\gamma(u_1), \ldots, \gamma(u_n))$ By Lemma 3.4.7, $\underline{x} \in C_h^0(X, 2\pi)$. By Lemmas 3.4.2 and 3.4.5, we have $\gamma \searrow \Gamma(\underline{x}) \searrow 0$ ◊

We can also give a direct proof of Proposition 3.1.4, namely if $\gamma \in \Omega(X, 2\pi)$ is a closed local geodesic, then $\gamma \not\sim 0$.

Proof of Proposition 3.1.4 : Suppose $\gamma \in \Omega(X, 2\pi)$ is a closed local geodesic and that $\gamma \sim 0$. Let δ be as given by Lemma 3.4.7. We can find $u_1, \ldots, u_n \in S^1$ dividing S^1 that the σ-lengths of all the the J_i are equal and less than δ, and such that $\text{length}(\gamma|J_i) \leq h$. Let $\underline{x} = (\gamma(u_1), \ldots, \gamma(u_n))$.

We have $\underline{x} \in C_h^0(X, 2\pi)$ and $\Gamma(\underline{x}) = \gamma$. By Lemma 3.3.4, we see that $f^{2n}(\underline{x}) = \underline{x}$, and so $\underline{x} \notin C_h^0(X, 2\pi)$. We deduce therefore that $\gamma \not\sim 0$. ◊

In the case where X is compact, we can also deduce Theorem 3.1.6, namely if $\gamma \in \Omega(X)$, then either $\gamma \searrow 0$, or else $\gamma \searrow \alpha$ where α is a closed local geodesic.

Proof of Theorem 3.1.6 : Suppose X is compact, and $\gamma \in \Omega(X)$. Let $u_1, \ldots, u_n$ divide S^1 into segments J_i such that $\text{length}(\gamma|J_i) \leq h$ for all i. Let $\underline{x} = (\gamma(u_1), \ldots, \gamma(u_n)) \in C_h(X)$. By Lemma 3.4.2, $\gamma \searrow \Gamma(\underline{x})$. If $\underline{x} \in C_h^0(X)$, then by Lemma 3.4.5, we have $\Gamma(\underline{x}) \searrow 0$, and so $\gamma \sim 0$.

Thus, we suppose that $\underline{x} \in C_h(X) \backslash C_h^0(X)$. Since X, is compact, some subsequence of $(f^r(\underline{x}))_{r \in \mathbf{N}}$ converges to some $\underline{y} \in C_h(X)$. Thus $E(f(\underline{y})) = E(\underline{y})$, and by Lemma 3.3.4, we have that $\Gamma(\underline{y}) = \alpha$ is a closed local geodesic.

Now for sufficiently large r, we can homotop $\Gamma(f^r(\underline{x}))$ to α, passing only through loops of length length $\alpha + \epsilon$, where $\epsilon > 0$ is arbitrarily small. For example, if $f^r(\underline{x}) = \underline{z} = (z_1, \ldots, z_n)$, let $\beta_i = [z_i \to y_i]$, and set $z(t) = (\beta_1(t), \ldots, \beta_n(t))$. Then the map $[t \mapsto \Gamma(z(t))]$ gives the desired homotopy.

We can suppose that $L(\underline{x}) > L(f^{2n}(\underline{x}))$ (otherwise $\Gamma(\underline{x})$ is already a local geodesic). Now set $\epsilon = L(\underline{x}) - L(f^{2n}(\underline{x}))$ and choose $r \in \mathbf{N} \cap [2n, \infty)$ accordingly. Thus, by Lemma 3.4.3, we have $\Gamma(\underline{x}) \searrow \Gamma(f^r(\underline{x}))$, and so $\Gamma(\underline{x}) \searrow \Gamma(\underline{y}) = \alpha$. Thus $\gamma \searrow \alpha$. ◊

Corollary 3.1.7 now follows. If X is compact, and $\gamma \in \Omega(X)$, with $\text{length}\,\gamma < \text{sys}(X)$, then either $\gamma \searrow 0$ or else $\gamma \searrow \alpha$ with $\text{length}\,\alpha \leq \text{length}\,\gamma$. But the latter case is impossible by Corollary 2.9.

We have still to prove Theorems 3.1.1–3.1.3 for X locally compact and l-CAT(1) For these, we will need the following lemma.

Lemma 3.4.8 : *Suppose $\underline{x} \in C_h^0(X, 2\pi)$. Then there is a compact π-CAT(1) space (Y, ρ), and points $y_1, \ldots, y_n \in Y$ such that $\rho(y_i, y_{i+1}) = d(x_i, x_{i+1})$ for all $i \in \{1, \ldots, n\}$, and $\rho(y_i, y_j) \geq d(x_i, x_j)$ for all $i, j \in \{1, \ldots, n\}$.*

Proof : In the case where $\underline{x}$ satisfies condition $(*)$, the construction Section 3.3 gives us a singular spherical metric ρ on the disc D, with points $y_1, \ldots, y_n \in \partial D$ satisfying the conclusion of the lemma. (Here $y_i = v(0,1) \in \partial D$.)

In the general situation, we take a product with a small regular euclidean polygon to obtain a family of metrics $\{\rho(\epsilon) \mid \epsilon > 0\}$ on the disc D, with the properties

$$d(x_i, x_{i+1}) \leq \rho(\epsilon)(y_i, y_{i+1}) \leq d(x_i, x_{i+1}) + \epsilon$$

and

$$d(x_i, x_j) \leq \rho(\epsilon)(y_i, y_j)$$

for all $i, j \in \{1, \ldots, n\}$. In this case, the space (Y, ρ) arises as a geometric limit of the spaces $(D, \rho(\epsilon))$ as $\epsilon \to 0$. Thus, (Y, ρ) is a spherical complex.

To make sense of this, note that the $(D, \rho(\epsilon))$ are all singular spherical metrics obtained from the same combinatorial complex G. Thus, each 2-cell of G is a spherical triangle. As $\epsilon \to 0$, some of these 2-cells may degenerate into geodesic segments or points. The remainder converge geometrically to become the 2-cells of Y. We may describe the 1-skeleton $K_1(Y)$ as follows. If $\epsilon > \eta > 0$, there is a natural map $(K_1(G), \rho(\epsilon)) \longrightarrow (K_1(G), \rho(\eta))$ which is linear on each edge. The metrics $\rho(\epsilon)$ thus converge to a limiting pseudometric $\rho(0)$ on $K_1(G)$, so that $(K_1(Y), \rho)$ is the hausdorffification of $(K_1(G), \rho(0))$.

Now, each vertex v in the 0-skeleton $K_0(Y)$ of Y is obtained by collapsing some subcomplex, $G(v)$ of G to a point. Now $G(v)$ must be simply connected (since any short simple closed curve in $(D, \rho(\epsilon))$ bounds a small disc). We consider two cases. If $G(v) \cap \partial D = \emptyset$, then the link of v in Y is a circle. Moreover, this circle has length at least 2π. (To see this, consider the boundary of a small uniform neighbourhood of $G(v)$ in $(D, \rho(\epsilon))$, which we can take to be a circle. By Gauß-Bonnet, this circle must have total turning at least $2\pi - \delta(\epsilon)$, where $\delta(\epsilon) \to 0$ as $\epsilon \to 0$.) On the other hand, if $G(v) \cap \partial D \neq \emptyset$, then the link of v in Y is a disjoint union of points and arcs. Thus, in both cases, the link is π-CAT(1), and so (Y, ρ) is locally CAT(1).

It remains to see that (Y, ρ) is π-CAT(1). By Corollary 3.1.7, it is enough to show that if $\gamma \in \Omega(Y, 2\pi)$, then $\gamma \sim 0$ in Y. However, this follows easily, since we can approximate γ by a loop, γ' in $(D, \rho(\epsilon))$ of $\rho(\epsilon)$-length less than 2π. Now $\gamma' \sim 0$ in $(D, \rho(\epsilon))$, and we may use the (2π)-homotopy of γ' to 0 to construct one for γ ◇

Next we prove Theorem 3.1.2 when X is l-CAT(1). Suppose that T is a triangle in X with $\mathrm{perim}(T) < 2\pi$ and with $\Gamma(T) \sim 0$. Then we claim that T is CAT(1).

Proof of Theorem 3.1.2 : Suppose that $T = (\alpha, \beta, \gamma; x, y, z)$. We choose points $x_1, \ldots, x_{3m}$, cyclically ordered on $\Gamma(T) = \alpha \cup \beta \cup \gamma$, such that $d(x_i, x_{i+1}) \leq h < l$ for all i, and such that $x_m = z$, $x_{2m} = x$ and $x_{3m} = y$. Thus $\underline{x} \in C_h(X, 2\pi)$ and $\Gamma(\underline{x}) = \Gamma(T)$. Now Lemma 3.4.7 tells us that, provided we have subdivided $\Gamma(T)$ finely enough, we have $\underline{x} \in C^0_h(X, 2\pi)$. Thus Lemma 3.4.8 gives us a π-CAT(1) space (Y, ρ), and points $y_1, \ldots, y_{3m} \in Y$ with $d(x_i, x_{i+1}) = \rho(y_i, y_{i+1})$ and $d(x_i, x_j) \leq \rho(y_i, y_j)$ for all $i, j \in \{1, \ldots, 3m\}$. Now since the points $x_1, \ldots, x_m$ lie along the geodesic α in X, we have $\rho(y_0, y_m) \geq d(x_0, x_m) = \sum_{i=1}^m d(x_{i-1}, x_i) = \sum_{i=1}^m \rho(y_{i-1}, y_i)$, and so the points $y_0, \ldots, y_m$ lie along a geodesic $\bar{\alpha}$ in Y. Similarly, $y_m, \ldots, y_{2m}$ lie along a geodesic $\bar{\beta}$ and $y_{2m}, \ldots, y_{3m}$ lie along a geodesic $\bar{\gamma}$. Let $\bar{T}$ be the triangle $(\bar{\alpha}, \bar{\beta}, \bar{\gamma})$ in Y. Now $\mathrm{perim}(\bar{T}) = \mathrm{perim}(T) < 2\pi$, so we may construct a comparison triangle $T' = (\alpha', \beta', \gamma')$ for $\bar{T}$ in (S^2, d_1). This is also a comparison triangle for T. We have a sequence of points $z_1, \ldots, z_{3m}$ around $\Gamma(T')$ with $d_1(z_i, z_{i+1}) = d(x_i, x_{i+1})$ for all i. Since Y is π-CAT(1), we have $d_1(z_i, z_j) \leq \rho(y_i, y_j)$ for all i, j, and so $d_1(z_i, z_j) \leq d(x_i, x_j)$. Since the set $\{x_1, \ldots, x_{3m}\}$ can be chosen to include any two given points of $\Gamma(T)$, we conclude that T is CAT(1). ◇

Now suppose that $\alpha, \beta : [0,1] \longrightarrow X$ are local geodesics with the same endpoints x and y, and with $\mathrm{length}\,\alpha + \mathrm{length}\,\beta < 2\pi$. Thus $\gamma = \alpha \cup -\beta \in \Omega(X, 2\pi)$. We claim that if $\gamma \sim 0$, then $\alpha = \beta$. This is Theorem 3.1.3, in the case where X is l-CAT(1).

Proof of Theorem 3.1.3 : Given a natural number m sufficiently large, we can take points $x_1, \ldots, x_m$ along α, and $x_m, \ldots, x_{2m}$ along β so that $d(x_i, x_{i+1}) = \mu$ for all $i \in \{1, \ldots, 2m\}$, and $d(x_{i-1}, x_{i+1}) = 2\mu$ for all $i \in \{1, \ldots, 2m\} \setminus \{m, 2m\}$, were $\mu < h$ is some constant. Thus $\underline{x} = (x_1, \ldots, x_{2m}) \in C_h(X, 2\pi)$ and $\Gamma(\underline{x}) = \gamma$. By Lemma 3.4.7, we can suppose that $\underline{x} \in C^0_h(X, 2\pi)$. Let $Y, \rho, y_1, \ldots, y_m$ be as given by Lemma 3.4.8. Thus $\rho(y_i, y_{i+1}) = \mu$ for all i, and $\rho(y_{i-1}, y_{i+1}) = 2\mu$ provided $i \neq m, 2m$. Thus, the points $y_0, \ldots, y_m$ lie along a local geodesic $\bar{\alpha}$ in Y, and $y_m, \ldots, y_{2m}$ lie along a local geodesic $\bar{\beta}$. Now $\mathrm{length}\,\bar{\alpha} + \mathrm{length}\,\bar{\beta} = \mathrm{length}\,\alpha + \mathrm{length}\,\beta < 2\pi$, and Y is π-CAT(1). Thus by Corollary 2.18, we have $\bar{\alpha} = \bar{\beta}$. Thus $y_i = y_{2m-i}$ for all i, and so $x_i = x_{2m-i}$. Thus $\alpha = \beta$ ◇

In Section 3.3, we defined the Birkhoff process for polygonal loops. We can also define a Birkhoff process for piecewise geodesic paths connecting two fixed points $x, y \in X$. Suppose $\underline{x} = (x_0, \ldots, x_n)$ is a sequence of points of X with $x_0 = x$, $x_n = y$, and $M(\underline{x}) = \max\{d(x_i, x_{i+1}) \mid i = 0, \ldots, n-1\} \leq h < l$. Let $f_0(\underline{x}) = (x, x'_1, x'_2, \ldots, x'_n, x_n)$ where $x'_i = \mathrm{mid}(x_{i-1}, x_i)$ for $i \in \{1, \ldots, n\}$. Let

$$f_1 \circ f_0(\underline{x}) = (x, \mathrm{mid}(x'_1, x'_2), \mathrm{mid}(x'_2, x'_3), \ldots, \mathrm{mid}(x'_{n-1}, x'_n), y).$$

Note that $M(f_1 \circ f_0(\underline{x})) \leq M(\underline{x}) \leq h$. Thus we may iterate $f_1 \circ f_0$. Since all metric balls in X are compact (Lemma 2.1), we see that some subsequence $(f_1 \circ f_0)^{r_i}(\underline{x})$ must converge on some $\underline{y} = (y_0, \ldots, y_n)$, with $M(\underline{y}) \leq h$. It's not hard to see (c.f. Lemma 3.3.4) that $f_1 \circ f_0(\underline{y}) = \underline{y}$, and that the points $y_0, \ldots, y_n$ are equally spaced along a local geodesic α joining x to y.

Now suppose that $\gamma : [0,1] \longrightarrow X$ is any path joining points x and y in X. We can choose $0 = u_0 < u_0 < \cdots < u_n = 1$ so that $\mathrm{length}(\gamma|[u_{i-1}, u_i]) \leq h$ for all $i = 1, \ldots, n$. Let $\underline{x} = (\gamma(u_0), \ldots, \gamma(u_n))$ and let $\underline{y}$ and α be as in the previous paragraph. As in the proof of Theorem 3.1.6, we see that $\gamma \searrow \alpha$. We have shown:

Lemma 3.4.9 : *Suppose $\gamma : [0,1] \longrightarrow X$ joins x to y in X. Then there is a local geodesic α joining x to y such that $\gamma \searrow \alpha$.* ◇

We may use this to deduce Theorem 3.1.1. Suppose $\alpha_1, \alpha_2, \alpha_3$ are paths joining x to y, and let $\gamma_i = \alpha_{i+1} \cup -\alpha_{i+2} \in \Omega(X)$. Suppose $\gamma_1, \gamma_2, \gamma_3 \in \Omega(X, 2\pi)$ and that $\gamma_1 \sim 0$ and $\gamma_2 \sim 0$. Then we claim that $\gamma_3 \sim 0$.

Proof of Theorem 3.1.1 : Let α_i, γ_i be as above. By Lemma 3.4.9, there are local geodesics $\alpha'_1, \alpha'_2, \alpha'_2$ joining x to y such that $\alpha_i \searrow \alpha'_i$. Let $\gamma'_i = \alpha'_{i+1} \cup -\alpha'_{i+2}$. Then $\gamma_i \searrow \gamma'_i$ for each i. If $\gamma_1 \sim 0$ and $\gamma_2 \sim 0$, then $\gamma'_1 \sim 0$ and $\gamma'_2 \sim 0$. By Theorem 3.1.3, we see that $\alpha'_2 = \alpha'_3$ and $\alpha'_3 = \alpha'_1$. Thus $\alpha'_2 = \alpha'_1$ and so $\gamma'_3 = \alpha'_1 \cup -\alpha'_2 \sim 0$. It follows that $\gamma_3 \sim 0$. ◇

This concludes the proofs of the main results 3.1.1–3.1.7 of Section 3.1, in the case where X is l-CAT(1) for some $l > 0$. In Section 3.6, we describe how to deal with the general case.

3.5. Lipshitz maps.

In this section, we observe that the results of the last section go through if we restrict attention to Lipshitz maps and homotopies.

Let σ be the standard path-metric on the circle $S^1 = \mathbf{R}/\mathbf{Z}$, so that the total σ-length of S^1 is 1. We may define a *μ-lipshitz loop* as a map $\gamma : S^1 \longrightarrow X$ such that $d(\gamma(t), \gamma(u)) \leq \mu\sigma(t, u)$ for all $t, u \in S^1$. We write $\Omega_L(X)$ for the space of all loops which are μ-lipshitz for some $\mu > 0$. We write $\Omega_L(X, r)$ for the subspace of loops which are μ-lipshitz for some $\mu < r$. Thus $\Omega_L(X) \subseteq \Omega(X)$ and $\Omega_L(X, r) \subseteq \Omega_L(X) \cap \Omega(X, r)$.

Given $\alpha, \beta \in \Omega_L(X)$, a *lipshitz homotopy* from α to β is a path $[t \mapsto \gamma_t] : [0,1] \longrightarrow \Omega_L(X)$ with $\gamma_0 = \alpha$, $\gamma_1 = \beta$, and $d_{sup}(\gamma_t, \gamma_u) \leq \lambda|t - u|$ for some $\lambda \geq 0$. The latter condition is equivalent to saying that the map $[(t, u) \mapsto \gamma_t(u)] : [0,1] \times S^1 \longrightarrow X$ is lipshitz. We may now define the relations $\searrow_L$ on $\Omega_L(X)$ and $\sim_L$ on $\Omega_L(X, 2\pi)$ by restricting to lipshitz homotopies. We make the following observations.

If $\gamma \in \Omega(X, r)$, then we can find a degree-1 homeomorphism $\phi : S^1 \longrightarrow S^1$ such that $\gamma \circ \phi \in \Omega_L(X, r)$.

If $\underline{x} \in C_h(X, r)$, we parameterise $\Gamma(\underline{x})$ proportionately to arc-length so that $\Gamma(\underline{x}) \in \Omega_L(X, r)$. If we normalise so that some fixed point of S^1 gets mapped to x_1 by $\Gamma(\underline{x})$, then the map $\Gamma : C_h(X, r) \longrightarrow \Omega_L(X, r)$ is continuous.

If $\underline{x} \in C_h(X, r)$, then $\Gamma(\underline{x}) \searrow_L \Gamma(f(\underline{x}))$. If $\underline{x} \in C^0_h(X, 2\pi)$, then $\Gamma(\underline{x}) \searrow_L 0$.

Suppose $\gamma \in \Omega_L(X)$ is μ-lipshitz, and $u_1, \ldots, u_n \in S^1$ divide S^1 into segments of σ-length at most h/μ. Then $\underline{x} = (\gamma(u_1), \ldots, \gamma(u_n)) \in C_h(X)$ and $\gamma \searrow_L \Gamma(\underline{x})$. If $\gamma \in \Omega_L(X)$ and $\gamma \sim 0$, then $\gamma \searrow_L 0$.

Suppose $\alpha, \beta \in \Omega_L(X, 2\pi)$ and $\alpha \sim \beta$. We can find point $\underline{x}, \underline{y} \in C_h(X, 2\pi)$ such that $\alpha \searrow \Gamma(\underline{x})$ and $\beta \searrow \Gamma(\underline{y})$ and $\underline{x}$ and $\underline{y}$ are connected by a path in $C_h(X, 2\pi)$ (Lemmas 3.4.2 and 3.4.6). Let $[t \mapsto (x_1(t), \ldots, x_n(t))] : [0,1] \longrightarrow C_h(X, 2\pi)$ be a such a path. Given $0 = t_0 < t_1 < \cdots < t_m = 1$, we may approximate $[t \mapsto x_i(t)]$ by the piecewise geodesic path $\gamma_i = \bigcup_j [x_i(t_{j-1}) \to x_i(t_j)]$. Let $\underline{z}(t) = (\gamma_1(t), \ldots, \gamma_n(t))$. If we choose the subdivision $t_0, \ldots, t_m$ fine enough, then we have $M(\underline{z}(t)) < h + \eta$, and $L(\underline{z}(t)) < 2\pi$ for all t, where $\eta < l - h$. Thus $\underline{z}(t) \in C_{h+\eta}(X, 2\pi)$ for all t, an the map $[t \mapsto \Gamma(\underline{z}(t))] : [0,1] \longrightarrow \Omega_L(X, 2\pi)$ gives a lipshitz homotopy from $\Gamma(\underline{x})$ to $\Gamma(\underline{y})$. We conclude that $\alpha \sim_L \beta$.

By a similar argument, we can deduce that if $\alpha, \beta \in \Omega_L(X)$ and $\alpha \searrow \beta$, then $\alpha \searrow_L \beta$.

All of the above statements follow more or less directly from the constructions of Section 3.4. We conclude that working in the category of lipshitz maps, as opposed to continuous maps, amounts to the same thing.

3.6. The general case.

In this section, we describe how the proofs of properties 3.1.1–3.1.5 given in Section 3.4, generalise to the case where X is locally compact and locally CAT(1). Using Lemma 2.4, it is enough to observe that everything of interest goes on inside some compact subset of X.

Lemma 3.6.1 : *Suppose ρ is a π-CAT(1) metric on the disc D, so that ∂D is rectifiable and length $\partial D < 2\pi$. Then $D = N(\partial D, \pi/2)$.*

Proof : By Lemma 3.3.10, there is some point $a \in D$ such that $\rho(a, x) < \pi/2$ for all $x \in \partial D$. We can suppose that $a \notin \partial D$.

We can identify D topologically as a cone over ∂D, i.e. as a quotient of $\partial D \times [0,1]$ where we shrink the circle $\partial D \times \{0\}$ to a point. We define a map $f : D \longrightarrow D$ by $f(x,t) = \alpha_x(t)$ where $\alpha_x = [a \to x] : [0,1] \longrightarrow X$. By Lemma 2.10, f is continuous. Moreover $f|\partial D$ is the identity on ∂D. Thus, by Brouwer degree, f is surjective. In other words, if $y \in D$, there is some $x \in \partial D$ with $y \in \alpha_x([0,1])$, and so $\rho(x, y) < \pi/2$. ◇

Using the construction of Section 3.3, this effectively tells us that if the Birkhoff process starting with a polygonal loop of length less than 2π converges to a point, then it will do so entirely within a metric $(\pi/2)$-neighbourhood of the original polygon.

More formally, suppose $K \subseteq X$ is compact, and $h > 0$. Write

$$C_h(X, 2\pi, K) = \{\underline{x} \in C_h(X, r) \mid x_i \in K \text{ for all } i = 1, \ldots, n\}.$$

Fix any $\epsilon_0 > 0$. By Lemma 2.1, the set $K' > 0 = N(K, \frac{\pi}{2} + \epsilon_0)$ is compact. Lemma 2.4 gives us some $\epsilon > 0$ such that any triangle with vertices in K' and perimeter less than 2ϵ apart may be joined by a unique geodesic in X. We can suppose that $\epsilon \leq \epsilon_0$. Now suppose that $h < \epsilon$, and $\underline{x} \in C_h(X, 2\pi, K)$. The Birkhoff process applied to $\underline{x}$ is well defined provided $f^m(\underline{x})$ remains in $C_h(X, 2\pi, K')$. Let

$$C_h^0(X, 2\pi, K) = \{\underline{x} \in C_h(X, 2\pi, K) \mid \forall m (f^m(\underline{x}) \in C_h(X, 2\pi, K')), \text{ and } L(f^m(\underline{x})) \to 0\}.$$

From Lemma 3.6.1, and the construction described in Section 3.3, we see that if $\underline{x} \in C_h^0(X, 2\pi, K)$, then the image of each loop $\Gamma(f^m(\underline{x}))$ lies inside

a $(\pi/2)$-neighbourhood of $\Gamma(\underline{x})$, and thus inside $K_0 = N(K, \frac{\pi}{2} + \frac{\epsilon}{2})$. Now K_0 is compact and lies in the interior of K'. Thus the argument of Section 3.3 works to show that:

Lemma 3.6.2 : *$C_h^0(X, 2\pi, K)$ is open and closed in $C_h(X, 2\pi, K)$.* ◇

The properties 3.1.1–3.1.5 now all follow easily. For example, to prove Theorem 3.1.5 (if $\gamma \in \Omega(X, 2\pi)$ and $\gamma \sim 0$ then $\gamma \searrow 0$), we argue as follows.

Let K be the image in X of the homotopy $[0,1] \longrightarrow \Omega(X, 2\pi)$ from γ to a constant loop. Let $\epsilon > 0$ be as described above, and $h \leq \epsilon$. Now as with Lemmas 3.4.2 and 3.4.6, we can find $u_1, \ldots, u_n \in S^1$ such that $\gamma \searrow \Gamma(\underline{x})$ and $\underline{x} = (\gamma(u_1), \ldots, \gamma(u_n))$ is connected by a path to a constant cycle in K. Clearly any such constant cycle lies in $C_h^0(X, 2\pi, K)$, and so by Lemma 3.6.2, $\underline{x} \in C_h^0(X, 2\pi, K)$. Thus, as in Lemma 3.4.5, $\Gamma(\underline{x}) \searrow 0$, and so $\gamma \searrow 0$.

The remaining results follow by similar arguments. We just need to observe that we can decide a-priori the compact subset of X in which we are interested.

As observed in Section 3.1, these results allow us to define the *systole* of a locally CAT(1) space as

$$\text{sys}(X) = \inf(\{2\pi\} \cup \{\text{length}\,\gamma \mid \gamma \in \Omega(X, 2\pi), \gamma \not\sim 0\}).$$

Thus, if $l = \frac{1}{2}\text{sys}(X) > 0$, then X is l-CAT(1).

3.7. Convergence of the Birkhoff process.

We have made use of the trivial fact that some subsequence of the Birkhoff curve-shortening process defined on a compact space must converge. This suffices for our applications, though it seems natural to ask when the process itself converges. We shall describe an example of a smooth riemannian 3-manifold where convergence fails for certain polygonal loops. By scaling the metric (if necessary) we can assume that the curvature is everywhere at most 1, and so this gives an example for a locally-CAT(1) space. The convergence of the Birkhoff process seems to be an open question for riemannian 2-manifolds. (See [Ga] for some discussion of the curve shortening flow in this context.) I don't know what, if anything, is known for real-analytic riemannian manifolds.

On a riemannian manifold, the Birkhoff process is closely related to curve shortening flow for smooth curves. Such a flow $\gamma(t,u)$ is defined by the equation $\frac{\partial\gamma}{\partial t} = \frac{DT}{\partial u}\Big/\left|\frac{\partial\gamma}{\partial u}\right|$ where $u \in S^1$, t is the time parameter, and $T = \frac{\partial\gamma}{\partial u}\Big/\left|\frac{\partial\gamma}{\partial u}\right|$ is the unit tangent to the curve $[u \mapsto \gamma(t,u)]$. This seems more natural in this context, though one has to worry about the possibility of running into singularities.

We begin with some positive results. We have already observed (Lemma 3.3.6) that if the length of a polygonal loop tends to 0 under the Birkhoff process, then the process must converge to a point. By a similar argument, we shall show that the Birkhoff process converges on any compact non-positively curved manifold, indeed on any space for which the distance function is convex locally.

Suppose X is a compact path-metric space. Suppose that, for some $l > 0$, the map $[(t,u) \mapsto d(\alpha(t),\beta(u))]$ is convex whenever $\alpha,\beta : [0,1] \longrightarrow X$ are geodesics with $\mathrm{diam}(\alpha([0,1]) \cup \beta([0,1])) < l$. (For such a space, we may define $\mathrm{sys}(X)$ to be the length of the shortest geodesic, and we can always take $l = \frac{1}{2}\mathrm{sys}(X)$.) Note that all compact CAT(0) spaces fall into this category.

We fix some $n \in \mathbf{N}$, and define $C_h(X)$ as for CAT(1) spaces. If $h < l$, then we may define Birkhoff curve shortening, $f : C_h(X) \longrightarrow C_h(X)$. Given $\underline{x}, \underline{y} \in C_h(X)$, let $d_{sup}(\underline{x},\underline{y}) = \max\{d(x_i,y_i) \mid i = 1\ldots n\}$. We see that if $d_{sup}(\underline{x},\underline{y}) < l$, then $d_{sup}(f(\underline{x}), f(\underline{y})) \leq d(\underline{x},\underline{y})$.

Proposition 3.7.1 : *Suppose X has a locally convex distance function, and that $h < l$ $(= \frac{1}{2}\mathrm{sys}(X))$. If $\underline{x} \in C_h(X)$, then $f^{2ni}(\underline{x})$ converges as $i \longrightarrow \infty$.*

Proof : Certainly some subsequence $f^{2nk_i}(\underline{x})$ must converge to some $\underline{y} \in C_h(X)$ with $f^{2n}(\underline{y}) = \underline{y}$. Given any $r < l$, we have that $d_{sup}(\underline{y}, f^{2nk_i}(\underline{x})) \leq r$ for some i. Thus, from the discussion above, $d_{sup}(\underline{y}, f^{2n(k_i+j)}(\underline{x})) \leq r$ for all $j \geq 0$. We see that $f^{2ni}(\underline{x})$ must converge to $\underline{y}$. ◊

Counterexample.

We describe a smooth riemannian metric on the 3-torus $T^3 = S^1 \times S^1 \times S^1$ for which the Birkhoff process in general fails to converge. If x, y, z are the coordinates on the respective $S^1 \equiv \mathbf{R}/\mathbf{Z}$ factors, then we take the

infinitessimal riemannian metric ds to have the form

$$ds^2 = dx^2 + dy^2 + \frac{1}{1 - \hat{\theta}(x,y)}dz^2$$

where $\hat{\theta} : S^1 \times S^1 \longrightarrow (0,\infty)$ is a smooth function.

Specifically, define $\theta : [-\frac{1}{2},\frac{1}{2}] \times \mathbf{R} \longrightarrow (0,\infty)$ by

$$\theta(x,y) = e^{-1/x}\left(2 - \cos 2\pi\left(y - \frac{1}{x^2}\right)\right)$$

if $x > 0$. We set $\theta(x,y) = \theta(-x,y)$ for $x < 0$, and $\theta(0,y) = 0$ for all y. We identify each pair of points $\{(-\frac{1}{2},y),(\frac{1}{2},y)\}$ for $y \in \mathbf{R}$, smooth out θ in a neighbourhood of $\{\pm\frac{1}{2}\} \times \mathbf{R}$ and factor out by the translation $[(x,y) \mapsto (x,y+1)]$ to give the map $\hat{\theta} : S^1 \times S^1 \longrightarrow (0,\infty)$. If we now take a polygonal loop with vertices equally spaced about a "fibre", $\{(x,y)\} \times S^1$ for some small $x > 0$, and iterate the Birkhoff process, we end up wandering around infinitely often in the y-direction. (We shall just formally prove that such a process cannot converge.)

To give the idea, suppose $\phi : \mathbf{R}^2 \longrightarrow (0,\infty)$ is a smooth function. Define a riemannian metric on $\mathbf{R}^2 \times S^1$ by $ds^2 = dx^2 + dy^2 + \frac{1}{\phi(x,y)^2}dz^2$. (We assume for the moment that ϕ is defined on the whole of $\mathbf{R}^2$, so that we don't have to worry about falling off the edge of our domain.) The Birkhoff process can be thought of as a discrete approximation to the curve shortening flow on fibres $\{u\} \times S^1$ for $u \in \mathbf{R}^2$. Note that such a fibre has length $\frac{1}{\phi(u)}$, and constant curvature $\frac{1}{\phi(u)}|\mathrm{grad}\,\phi(u)|$ in a direction which projects to $\mathrm{grad}\,\phi(u)$ in $\mathbf{R}^2$. Thus, at time t, the fibre $\{u\} \times S^1$ flows to a fibre $\{\gamma(t)\} \times S^1$, where the curve γ satisfies $\gamma(0) = u$, and $\frac{d\gamma}{dt}(t) = \frac{1}{\phi(\gamma(t))}\mathrm{grad}\,\phi(\gamma(t))$. In other words, curvature flow on fibres reduces to steepest ascent of $\log\phi$ on $\mathbf{R}^2$. It is easy to see from this that the curvature flow of fibres on the 3-torus described above does not in general converge. The Birkhoff process can be thought of as a discrete analogue of the curve shortening flow, and so this approximates to steepest ascent, though the proof in this case will involve us in some messy analysis.

Suppose $[t \mapsto (\beta(t), z(t))]$ is a geodesic in $\mathbf{R}^2 \times S^1$, so that β is a curve in $\mathbf{R}^2$. We may compute the geodesic equation to give

$$\frac{d}{dt}\left(\frac{dz}{dt}(t)\Big/\phi(\beta(t))^2\right) = 0$$

and

$$\frac{D}{dt}\left(\frac{d\beta}{dt}\right)(t) = \frac{1}{2}\left(\frac{dz}{dt}(t)\right)^2 \operatorname{grad}\left(\frac{1}{\phi(\beta(t))^2}\right).$$

Let's suppose that at time $t = 0$, we begin tangentially to the fibre $\{\beta(0)\} \times S^1$, i.e. $\frac{d\beta}{dt}(0) = 0$ and $\frac{dz}{dt}(0) = \phi(\beta(0))$. Then, $\frac{dz}{dt}(t) = \frac{\phi(\beta(t))^2}{\phi(\beta(0))}$ and so

$$\frac{D}{dt}\left(\frac{d\beta}{dt}\right)(t) + \frac{1}{2\phi(\beta(0))^2}\operatorname{grad}\phi(\beta(t))^2 = 0.$$

We thus arrive at the dynamics of a particle moving under a potential $\phi^2/2\phi(\beta(0))^2$. Note that

$$\left|\frac{d\beta}{dt}(t)\right|^2 + \frac{\phi(\beta(t))^2}{\phi(\beta(0))^2} = 1,$$

and so $\phi(\beta(t)) \leq \phi(\beta(0))$ for all $t > 0$.

Let us assume that $\frac{1}{2} \leq \phi^2 \leq 1$, and that there is a bound on the second derivatives of ϕ, so that the norm of the sectional curvature of $\mathbf{R}^2 \times S^1$ is bounded. This means that there is some $n \in \mathbf{N}$ such that any two points of $\mathbf{R}^2 \times S^1$, a distance at most $2/n$ apart, are joined by a unique geodesic. In particular, suppose that $\underline{x} = ((u,0),(u,\frac{1}{n}),\ldots,(u,\frac{n-1}{n}))$ is a cycle of n equally spaced points of the fibre $\{u\} \times S^1$. Since $\phi \geq \frac{1}{2}$, any such fibre has length at most 2, and so Birkhoff curve-shortening, f, applied to $\underline{x}$ is well-defined. In fact, from the S^1 symmetry, we see that $f(\underline{x}) = ((g(u),\frac{1}{2}),(g(u),\frac{1}{n}+\frac{1}{2}),\ldots,(g(u),\frac{n-1}{n}+\frac{1}{2}))$ for some $g(u) \in \mathbf{R}^2$. The point $g(u)$ is uniquely determined by the fact that it is joined to u by a path $\beta : [0,t_0] \longrightarrow \mathbf{R}^2$ with $\beta(0) = g(u)$, $\beta(t_0) = u$, satisfying the differential equation given above, and with $\int_0^{t_0} \phi(\beta(t))^2 dt = \phi(\beta(0))/2n$. This last equation derives from the formula for dz/dt, and the fact that $z(t_0) - z(0) = 1/2n$. From this, it follows that $t_0 \leq 1/n$.

Suppose that for some $r > 0$, we have $\max_{N(g(u),r)} |\operatorname{grad}\phi^2| < r$. Then we claim that $||u - g(u)|| \leq r$, where $||.||$ is the euclidean norm. For if not, set $t' = \inf \beta^{-1}(\mathbf{R}\setminus N(g(u),r)) < t_0 < 1/n$. By the Mean Value Theorem, we arrive at the contradiction $r = ||\beta(0) - \beta(t')|| \leq \operatorname{length}(\beta|[0,t']) \leq \frac{t'^2}{4\phi(\beta(0))}\max_{\beta([0,t'])}|\operatorname{grad}\phi^2| \leq \frac{r}{4n^2} < r$. This proves the claim.

We now restrict attention to the case where $\phi^2 = 1 - \theta$, where $\theta > 0$ is defined on $(0,\frac{1}{2}] \times \mathbf{R}$ by $\theta(x,y) = e^{-1/x}\left(2 - \cos 2\pi\left(y - \frac{1}{x^2}\right)\right)$. (It is irrelevant, for the moment, how θ is defined elsewhere.) We make the

following observations about θ. Clearly, $e^{-1/x} \leq \theta(x,y) \leq 3e^{-1/x} \leq \frac{1}{2}$. In fact, the graph of θ consists of an infinite sequence of "ridges" $\{\rho_m\}$ and "troughs" $\{\tau_m\}$ defined as follows. Given $m \in \mathbf{N}$, define $\tau_m, \rho_m : (0,\frac{1}{2}] \longrightarrow (0,\frac{1}{2}] \times \mathbf{R}$ by $\tau_m(x) = (x, \frac{1}{x^2} - n)$ and $\rho_m(x) = (x, \frac{1}{x^2} - n - \frac{1}{2})$. Thus $\theta(\tau_m(x)) = e^{-1/x}$ and $\theta(\rho_m(x)) = 3e^{-1/x}$. We also have $|\operatorname{grad}\theta(x,y)| \leq H(x)$, where $H(x) = \frac{20}{x^3}e^{-1/x}$.

Now take μ to be some small positive constant ($\mu = \frac{1}{4}$ will do). Note that the functions $H((1+\mu)x)/x$ and $H((1+\mu)x)^2/e^{-1/x}$ both tend to 0 as $x \to 0$. Thus, we can find some $\epsilon > 0$ such that if $0 < x \leq \epsilon$, then $H((1+\mu)x) < \mu x$, and $H((1+\mu)x)^2 \leq \mu\theta(x,y)$ for all $y \in \mathbf{R}$.

Suppose $u = (x',y')$ and $g(u) = (x,y)$ with $x \leq \epsilon$ and $x' \leq \epsilon$. Set $r = H((1+\mu)x)$. If $v = (x'',y'') \in N(g(u),r)$, then $x'' \leq x + r \leq x + H((1+\mu)x) < x + \mu x = (1+\mu)x$. Thus $|\operatorname{grad}\theta(v)| \leq H(x'') < H((1+\mu)x) = r$. In other words, $\max_{N(g(u),r)}|\operatorname{grad}\theta| < r$, and so $||u - g(u)|| \leq r = H((1+\mu)x)$. Now, let $[u,g(u)]$ be the euclidean geodesic joining u to $g(u)$. If $v \in [u,g(u)]$, then by the Mean Value Theorem, we have $|\theta(v) - \theta(g(u))| \leq ||v - g(u)|| \max_{[u,g(u)]}|\operatorname{grad}\theta| \leq H((1+\mu)x)^2 \leq \mu\theta(x,y) = \theta(g(u))$. Thus $(1-\mu)\theta(g(u)) \leq \theta(v) \leq (1+\mu)\theta(g(u))$.

Suppose that $\theta(x,y) = \theta(-x,y)$ for all y. By symmetry, we see that if the x-coordinate of $u \in \mathbf{R}^2$ is positive, then so is that of $g(u)$. Suppose in addition that θ is bounded away from 0 on $(\mathbf{R}\setminus[-\epsilon,\epsilon]) \times \mathbf{R}$. We know that $\theta(g(u)) \leq \theta(u)$ for all u, so if we start sufficiently close to the y-axis and iterate g, then (we can suppose that) the x-coordinates of the iterates tend to 0. (They cannot accumulate in $(0,\epsilon] \times \mathbf{R}$ since θ has no local minimum there.)

Suppose then that we have a sequence of iterates $u_i = (x_i,y_i) = g^i(u_0)$ with $x_i < \epsilon$ for all i. We have $x_i \to 0$, and $||u_{i-1} - u_i|| \to 0$. Also, for all $v \in [u_{i-1},u_i]$ we have $(1-\mu)\theta(u_i) \leq \theta(v) \leq (1+\mu)\theta(u_i)$. We claim that such a sequence cannot converge.

To see this, note that any limit would have to lie on the y-axis, and so the piecewise geodesic path $\bigcup_{i=1}^{\infty}[u_{i-1},u_i]$ would have to cross, in sequence, infinitely many troughs and ridges of θ. More formally, given any $\eta \in (0,\epsilon)$, we can find $m = m(\eta)$, $i = i(\eta)$ and $j = j(\eta)$, with $i \leq j$, such that $x_{i-1}, x_i, x_{j-1}, x_j \in (0,\eta)$, and such that $\tau_m(a) \in [u_{i-1},u_i]$, and $\rho_m(b) \in [u_{j-1},u_j]$ for some $a = a(\eta)$ and $b = b(\eta)$ in $\mathbf{R}$. Now, $3(1-\mu)e^{-1/b} = (1-\mu)\theta(\rho_m(b)) \leq \theta(u_j) \leq \theta(u_i) \leq (1+\mu)\theta(\tau_m(a)) = (1+\mu)e^{-1/b}$, and so $\frac{1}{a} - \frac{1}{b} \geq K = \log\left(\frac{3(1-\mu)}{(1+\mu)}\right) > 0$. Thus $||\tau_m(a) - \rho_m(b)|| \geq \frac{1}{a^2} - \frac{1}{b^2} - \frac{1}{2} \geq K\left(\frac{1}{a} + \frac{1}{b}\right) - \frac{1}{2} \geq \frac{2K}{\eta} - \frac{1}{2}$. Now, $||u_i - \tau_m(a)||$ and $||u_j - \rho_m(b)||$ are bounded, and so $||u_{i(\eta)} - u_{j(\eta)}|| \to \infty$ as $\eta \to 0$. This contradicts the assumption that the sequence (u_i) converges.

We have shown that the Birkhoff process in general fails to converge for such a metric in $\mathbf{R}^2 \times S^1$. The same argument can be carried out for the 3-torus, with $\hat{\theta}$ replacing θ, as we described at the beginning.

4. Area inequalities.

In this chapter we are principally interested in riemannian metrics on the disc. We relate some of the previous discussion to inequalities involving area. In section 4.1, we discuss the spherical isoperimetric inequality, and in Section 4.2, we describe an area comparison theorem for triangular regions.

In dealing with area, we shall confine our attention to smooth riemannian metrics, though there are, no doubt, ways one could attempt to generalise this. A discussion of area for more general path-metrics on surfaces is given in the book [AZ]. Many other geometric inequalities for riemannian manifolds are described in [BurZ].

4.1. The spherical isoperimetric inequality.

In this section, we are interested primarily in "riemannian discs". By this we mean a path-metric ρ on the topological disc D, such that $\rho|\text{int}\, D$ is (derived from) a smooth riemannian metric, and such that $(\partial D, \rho)$ is a rectifiable curve.

We shall continue to use the term "geodesic" in the path-metric sense, namely a length-minimising path, parameterised proportionately to arc-length. Such a geodesic, $\gamma : [0,1] \longrightarrow D$ will be a riemannian geodesic, except, perhaps where it meets the boundary ∂D. If the boundary is smooth, then γ will have zero or outward pointing riemannian curvature at points of $\gamma^{-1}(\partial D)$. In general, we may speak of γ being "concave" where it meets ∂D. (It may be approximated by smooth curves in int D of arbitrarily small inward curvature.)

If the riemannian curvature of ρ in int D is everywhere at most 1, then (D, ρ) is locally CAT(1). When length$(\partial D) < 2\pi$, we have constraints on the possible area of such a disc, which may be expressed in the form of

Theorem 4.1.4. This is a variation on the standard spherical isoperimetric inequality.

We begin with some general observations about locally CAT(1) discs.

Lemma 4.1.1 : *Suppose ρ is a locally CAT(1) path-metric on the disc D, such that ∂D is rectifiable and* length$(\partial D) < 2\pi$. *Then (D, ρ) is π-CAT(1) if and only if $\partial D \sim 0$.*

Proof : If (D, ρ) is π-CAT(1), then $\partial D \sim 0$ by Corollary 3.1.7.

Suppose that $\partial D \sim 0$. There are various ways to see that (D, ρ) must be π-CAT(1). For example, suppose to the contrary that (D, ρ) is not π-CAT(1). Then it must contain an embedded closed geodesic γ. Given $h < \frac{1}{2}\text{sys}(X) \leq \frac{1}{2}\text{length}\,\gamma$, by Lemma 3.4.7, we can find $\underline{x} = \{x_1, \ldots, x_n\} \in C_h^0(X, 2\pi)$, with $x_j \in \partial D$ for all j. However, we see easily that γ acts as a barrier to the curves $\Gamma(f^i(\underline{x}))$ as $i \to \infty$, contradicting the fact that they must converge to a point. ◇

Lemma 4.1.2 : *If (D, ρ) is π-CAT(1), then there is some $a \in D$ such that $d(a, x) < \pi/2$ for all $x \in D$.*

Proof : As with Lemma 3.6.1. ◇

Note in particular that diam$(D, \rho) < \pi$, and so any pair of point are joined by a unique geodesic.

Lemma 4.1.3 : *Suppose (D, ρ) is a riemannian disc of curvature ≤ 1, and with* length$(\partial D) < 2\pi$. *Then (D, ρ) is π-CAT(1) if and only if* area$(D, \rho) < 2\pi$.

Proof : If (D, ρ) is not π-CAT(1), then it contains an embedded closed geodesic (in the path-metric sense) of length $< 2\pi$. This bounds a closed disc $D_0 \subseteq D$. Now ∂D_0 is concave (in the riemannian sense), and so the Gauß-Bonnet formula tells us that area$(D_0, \rho) \geq 2\pi$. Thus area$(D, \rho) \geq 2\pi$ as required.

Now suppose that (D, ρ) is π-CAT(1). By Lemma 4.1.3, there is some $a \in D$ such that $d(a, x) < \pi/2$ for all $x \in \partial D$. We can assume that $a \in \text{int}\, D$. If ∂D is smooth, then it is the image of a smooth embedding $\gamma : S^1 \longrightarrow D$. Fix $t_0 \in S^1$, and for $t \in S^1$, let $A(t)$ be the area of the "sector" bounded by the geodesics $[a \to \gamma(t_0)]$ and $[a \to \gamma(t)]$. (More precisely, $A(t)$ is the area of the union of the images of $[a \to \gamma(u)]$ for all u in the positively oriented interval of S^1 joining t_0 to t.) Applying the Rauch comparison theorem, we find that $A(t)$ is differentiable in t, and that $dA(t)/dt \leq |d\gamma(t)/dt|$. Integrating over S^1 we find that $\text{area}(D, \rho) \leq \text{length}(\partial D) < 2\pi$. The general case follows by approximating ∂D by smooth curves. ◊

We may refine the above result by quoting the following isoperimetric inequality.

Suppose (D, ρ) is a riemannian disc of curvature ≤ 1, and with $\text{area}(D) \leq 4\pi$. Then

$$\text{length}(\partial D) \geq L(\text{area}(D))$$

where $L(A) = \sqrt{A(4\pi - A)}$. Note that the extremal case is that of a spherical cap in (S^2, d_1) of area A, bounded by a round circle (of length equal to $L(A)$).

This inequality, in a variety of forms, has a long history, and one could attach to it a long list of names. For an exposition of this, and many related inequalities, we refer to Osserman's articles [O1,O2].

We may express the above inequality in terms of the "dual" problem of spanning a circle of a given length by a disc of curvature ≤ 1. Thus, if $\text{length}(\partial D) = L < 2\pi$, then either $\text{area}(D) \leq A_-(L)$ or else $\text{area}(D) \geq A_+(L)$, where $A_\pm(L) = 2\pi \pm \sqrt{4\pi^2 - L^2}$. Thus $A_-(L)$ and $A_+(L)$ are, respectively, the areas of the small and large spherical caps in the unit 2-sphere, bounded by a round circle of length L. Clearly $A_-(L) < 2\pi < A_+(L)$, and so by Lemmas 4.1.1 and 4.1.3, we see that this dichotomy can be expressed in terms of the shrinkability of the boundary:

Theorem 4.1.4 : *Suppose that (D, ρ) is a riemannian disc of curvature ≤ 1, and that $L = \text{length}(\partial D) < 2\pi$. If $\partial D \sim 0$, then $\text{area}(D) \leq A_-(L)$, whereas if $\partial D \not\sim 0$, then $\text{area}(D) \geq A_+(L)$, where $A_\pm(L) = 2\pi \pm \sqrt{4\pi^2 - L^2}$.*

4.2. Area comparison for triangles.

In this section we show (Proposition 4.2.7) that the area of a riemannian disc of curvature ≤ 1 bounded by a shrinkable triangle is less than or equal to the area of the small region in the unit 2-sphere bounded by a comparison triangle.

We begin with some observations from spherical geometry. Given two non-antipodal points $x, y \in S^2$, we write $m(x, y)$ for the midpoint of the geodesic joining x and y.

Lemma 4.2.1 : *There is some universal constant $k > 0$ such that if (x, y, z) is a spherical triangle, then*

$$d_1(m(x,y), m(x,z)) \leq \frac{1}{2} d_1(y, z) + k\,\text{perim}(x, y, z)^2.$$

Proof : One can obtain an explicit value for using the spherical cosine formula.

Alternatively, note that the exponential map to S^2 based at some point $x \in S^2$ is analytic. Thus, if $\alpha, \beta : [0,1] \longrightarrow S^2$ are geodesics of length $< \pi$ parameterised proportionately to arc-length with $\alpha(0) = \beta(0) = x$, then $d(\alpha(t), \beta(t))$, as a function of t, can be written in the form $at(1 + tg(t))$, where the constant a and analytic function g are confined to a compact set. ◊

Given a non-degenerate spherical triangle $T_0 = (x_1, x_2, x_3)$ in S^2, we may subdivide the small region $R(T_0)$ bounded by $\Gamma(T_0)$ into four smaller regions $R(T_i)$, $i = 1, 2, 3, 4$, by joining the midpoints of the edges. More formally, set $y_i = m(x_{i+1}, x_{i+2})$, and set $T_i = (x_i, y_{i+1}, y_{i+2})$ for $i = 1, 2, 3$, and $T_4 = (y_1, y_2, y_3)$. (Figure 4a.) Clearly, $\text{perim}(T_i) \leq \text{perim}(T_0)$ for all $i \in \{1, 2, 3, 4\}$. In fact,

Lemma 4.2.2 : *Given $\epsilon > 0$, there is some $\mu < 1$, such that if $\text{perim}(T_0) \leq 2(\pi - \epsilon)$, then $\text{perim}(T_i) \leq \mu\,\text{perim}(T_0)$ for $i \in \{1, 2, 3, 4\}$.*

Proof : Since $d_1(x_i, x_j) \leq 2d_1(y_i, y_j)$ for $i, j \in \{1, 2, 3\}$, we have $\text{perim}(T_i) \leq \text{perim}(T_4)$ for $i \in \{1, 2, 3\}$. Thus, it suffices to verify that $\text{perim}(T_4) \leq \mu\,\text{perim}(T_0)$. This is certainly true for small triangles (with μ close to $\frac{1}{2}$), so we can suppose, without loss of generality, that $d_1(x_1, x_2)$ and $d_1(x_1, x_3)$ are greater than some fixed constant. If $\text{perim}(T_4)$ were almost equal to $\text{perim}(T_0)$, then we would have $d_1(y_2, y_3)$ almost equal to $d_1(y_2, x_1) + d_1(x_1, y_3)$, so that the angle at x_1 would be abitrarily close to π. If we now assume, in addition, that $\text{perim}(T_0)$ is bounded away from 2π, then it follows that $d_1(x_2, x_3)$ is also bounded away from 0. So, by a similar argument we see that the angles at x_2 and x_3 are also close to π. We arrive at the conclusion that $\text{perim}(T_0)$ is arbitrarily close to 2π, contrary to our hypothesis. ◇

Note that the construction of the triangles T_1, T_2, T_3, T_4 makes sense, and Lemma 4.2.2 remains valid, even if T_0 is degenerate.

By a "locally CAT(1) triangular region", we shall mean a locally CAT(1) path-metric ρ on the disc D, such that ∂D consists of three geodesic segments $\alpha_1, \alpha_2, \alpha_3$. In other words, $D = R(T)$ and $\partial D = \Gamma(T)$, where T is the triangle $(\alpha_1, \alpha_2, \alpha_3)$. Suppose $\text{perim}(T) < 2\pi$, then by Theorem 3.1.2, and Lemma 4.1.1, the following three conditions are equivalent:

(1) The space $(R(T), \rho)$ is π-CAT(1),

(2) The triangle T is CAT(1),

(3) $\Gamma(T) \sim 0$.

Suppose that $(R(T_0), \rho)$ is such a π-CAT(1) triangular region. As observed after Lemma 4.1.2, any pair of points of $R(T_0)$ are joined by a unique geodesic. Thus, as in the case of a spherical triangle, we may join together the midpoints of the edges of T_0 to obtain four well-defined triangles T_1, T_2, T_3, T_4.

If $\Gamma(T_i)$ happens to be an embedded curve, then it bounds a closed disc $R(T_i)$. Since $R(T_i)$ is obtained from $R(T_0)$ by cutting along geodesics, we see that the metric ρ restricted to $R(T_i)$ is already a path-metric, and is locally CAT(1). In fact, $(R(T_i), \rho)$ is π-CAT(1), since any embedded closed geodesic in $R(T_i)$ would be an obstruction to shrinking $\Gamma(T_0)$ (as in the proof of Lemma 4.1.1). Thus $R(T_i)$ is itself a π-CAT(1) triangular region.

In general, there are several ways in which $\Gamma(T_0)$ may be degenerate, though in all cases, we obtain a π-CAT(1) "region" $R(T_i)$ "bounded by" $\Gamma(T_i)$. Thus, $R(T_i)$ might consist of an arc, three arcs connected together at a common endpoint, or a (genuine) triangular region with arcs attached to one or more of the vertices. This is a somewhat tedious complication, which we shall not worry about too much.

We may iterate this procedure to obtain, at the nth stage, a subdivision of $R(T_0)$ into 4^n triangular regions. (Note that triangular regions may also degenerate into points, on iteration.)

More formally, we write $\mathcal{T}_1(T_0) = \{T_1, T_2, T_3, T_4\}$, and define inductively $\mathcal{T}_n(T_0) = \bigcup\{\mathcal{T}_1(T) \mid T \in \mathcal{T}_{n-1}(T_0)\}$. By Lemma 4.2.2, we see that $\max\{\text{perim}(T) \mid T \in \mathcal{T}_n(T_0)\}$ tends to 0 geometrically in n. Also,

Lemma 4.2.3 : *The quantity $\sum_{T\in\mathcal{T}_n(T_0)} \text{perim}(T)^2$ is bounded (by some function of $\text{perim}(T_0) < 2\pi$).*

Proof : By Lemma 4.2.2 and the CAT(1) inequality, we see that if $T \in \mathcal{T}_n(T_0)$, then $\text{perim}(T) \le 2\pi\mu^n$. Let $\mathcal{T}_1(T) = \{T_1, T_2, T_3, T_4\}$. By Lemma 4.2.1 and the CAT(1) inequality, we have that for each i,

$$\text{perim}(T_i) \le \frac{1}{2}\text{perim}(T) + 3k\,\text{perim}(T)^2,$$

and so

$$\sum_{i=1}^{4} \text{perim}(T_i)^2 \le \frac{1}{2}\text{perim}(T)^2(1 + 12k\,\text{perim}(T) + 36k^2\text{perim}(T)^2)$$
$$\le \text{perim}(T)^2(1 + K\mu^n)$$

where $K = 24k\pi + 144k^2\pi^2$. By induction, it follows that

$$\sum_{T\in\mathcal{T}_{n+1}(T_0)} \text{perim}(T)^2 \le (1 + K\mu^n) \sum_{T\in\mathcal{T}_n(T_0)} \text{perim}(T)^2$$
$$\le \prod_{i=1}^{n}(1 + K\mu^i)$$

Thus $\mathcal{T}_n(T_0)$ is bounded by $\prod_{i=1}^{\infty}(1 + K\mu^i) < \infty$. ◇

Given a triangle T, with $\text{perim}(T) < 2\pi$, we write T' for the comparison triangle in S^2. The area, $\text{area}(R(T'))$, of the small region bounded by T' is well-defined.

Lemma 4.2.4 : *Suppose $(R(T), \rho)$ is a π-CAT(1) triangular region. Let $\mathcal{T}_1(T) = \{T_1, T_2, T_3, T_4\}$, and let T', T_i' be comparison triangles in S^2. Then*

$$\sum_{i=1}^{4} \text{area}(R(T_i')) \le \text{area}(R(T')).$$

Proof : We can assume that the comparison triangles have the form $T_i' = (x_i', y_{i+1}', y_{i+2}')$ for $i = 1, 2, 3$ and $T_4' = (y_1', y_2', y_3')$, and that the interiors of $R(T_i')$ are disjoint. Thus $R' = \bigcup_{i=1}^4 R(T_i')$ is a spherical hexagonal region. Let σ be the induced path-metric on R'. Since $R = R(T)$ is π-CAT(1), the natural map $(\partial R', \sigma) \longrightarrow (\partial R, \rho)$ is distance non-increasing. In particular, $\sigma(x_{i+1}', x_{i+2}') = \sigma(x_{i+1}', y_i') + \sigma(y_i', x_{i+2}')$, so the interior angle of R' at y_i is at least π, for all i. (Figure 4b.) Applying Proposition 1.2(7) three times, we see that $\sum_{i=1}^4 \text{area}(R(T_i')) = \text{area}(R') \leq \text{area}(R(T'))$ as required. ◇

By induction, we obtain:

Corollary 4.2.5 : *If $R(T_0)$ is a π-CAT(1) triangular region, then*

$$\sum_{T \in \mathcal{T}_n(T_0)} \text{area}(R(T')) \leq \text{area}(R(T_0')).$$

◇

By a *riemannian triangular region*, we mean a path-metric ρ on the disc, such that $\rho|\text{int}\, D$ is riemannian, and such that ∂D consists of three geodesics arranged in a triangle. To avoid technical complications, we shall assume that (D, ρ) is a riemannian manifold with smooth boundary and corners at the vertices of the triangle. This implies some lower bound on the curvature.

In general the edges of the triangle will be concave in the riemmannian sense. However, we shall tacitly assume in what follows that they are all riemannian geodesics. In this way, the triangles obtained in the above subdivision will all be non-degenerate. The general case can be dealt with modifying the metric in a neighbourhood of the boundary so that it has this property, or by describing the manner in which triangles may degenerate.

Lemma 4.2.6 : *Suppose $R(T)$ is a riemannian triangular region with curvature between $-\kappa^2$ and 1, and with $\text{perim}(T) < 2\pi$, and $\Gamma(T) \sim 0$. Then*

$$|\text{area}(R(T)) - \text{area}(R(T'))| \leq \text{perim}(T)^2 \eta_\kappa(\text{perim}(T))$$

where $\eta_\kappa \to 0$ monotonically as $t \to 0$.

Proof : Let $\lambda = \text{perim}(T) = \text{perim}(T')$. We scale the metric ρ by a factor $1/\lambda$ so that $(R(T), \rho/\lambda)$ has curvature between $-\lambda^2\kappa^2$ and λ^2. Since there are no conjugate points, the inverse exponential map $\log = \exp^{-1} : R(T) \longrightarrow \mathbf{R}^2 \equiv (\mathbf{E}^2, d_0)$, based at some vertex of $R(T)$ is well-defined and injective. As $\lambda \to 0$, this map becomes arbitrarily close to area-preserving, and the images $\log(R(T))$ become arbitrarily close (in the Hausdorff topology) to a euclidean comparison triangle, $R(T'')$ for T. Thus $|\text{area}(R(T), \rho/\lambda) - \text{area}(R(T''), d_0)|$ is arbitrarily small, depending on λ. It's not hard to see that this convergence is uniform (independent of the shape of $R(T''))$, so that this quantity is bounded by some function, $\frac{1}{2}\eta_\kappa(\lambda)$, of κ and λ. We therefore also have $|\text{area}(R(T'), d_1/\lambda) - \text{area}(R(T''), d_0)| \leq \frac{1}{2}\eta_\kappa(\lambda)$, and so $|\text{area}(R(T), \rho) - \text{area}(R(T'), d_1)| \leq \lambda^2 \eta_\kappa(\lambda)$ as required. ◇

Proposition 4.2.7 : *Suppose $R(\Delta)$ is a riemannian triangular region of curvature ≤ 1, with $\text{perim}(T) \leq 2\pi$, and with $\Gamma(\Delta) \sim 0$. Let Δ' be a spherical comparison triangle for Δ. Then*

$$\text{area}(R(\Delta)) \leq \text{area}(R(\Delta')).$$

Proof : We are assuming that there is some lower bound, $-\kappa^2$, on curvature. Given $\epsilon > 0$, there is some $\delta > 0$ such that $\eta_\kappa(\delta) < \epsilon$. There is some natural number n such that $\text{perim}(T) < \delta$ for all $T \in \mathcal{T}_n(\Delta)$. Thus, by Lemma 4.2.6, if $T \in \mathcal{T}_n(\Delta)$,

$$|\text{area}(R(T)) - \text{area}(R(T'))| \leq \epsilon \text{perim}(T)^2.$$

Now $\text{area}(R(\Delta)) = \sum_{T \in \mathcal{T}_n(\Delta)} \text{area}(R(T))$, and so

$$\left|\text{area}(R(\Delta)) - \sum_{T \in \mathcal{T}_n(\Delta)} \text{area}(R(T'))\right| \leq \epsilon \sum_{T \in \mathcal{T}_n(\Delta)} \text{perim}(T)^2 \leq C\epsilon,$$

where C is constant (Lemma 4.2.3). By Corollary 4.2.5, we have

$$\sum_{T \in \mathcal{T}_n(\Delta)} \text{area}(R(T')) \leq \text{area}(R(\Delta')).$$

Thus

$$\text{area}(R(\Delta)) \leq \text{area}(R(\Delta')) + C\epsilon.$$

The result follows by letting $\epsilon \to 0$. $\diamond$

The conclusion is independent of the lower curvature bound, $-\kappa^2$, so it's not hard to see that this condition can be dropped.

The same argument can be used to show that if $R(\Delta)$ is a triangular region of curvature ≤ 0 (≤ -1), then $\text{area}(R(\Delta)) \leq \text{area}(R(\Delta''))$ where Δ'' is a euclidean (hyperbolic) comparison triangle. Here, the bound on $\text{perim}(\Delta)$, and the hypothesis that $\Gamma(\Delta) \sim 0$ are redundant.

By analogy with Theorem 4.1.4, it seems reasonable to conjecture that if $R(\Delta)$ is a riemannian triangular region of curvature ≤ 1, with $\text{perim}(\Delta) < 2\pi$ and with $\Gamma(\Delta) \not\sim 0$, then we have $\text{area}(R(\Delta)) \geq 4\pi - \text{area}(R(\Delta'))$.

References.

[AZ] A.D.Aleksandrov, V.A.Zalgaller, *Intrinsic geometry of surfaces* : Amer. Math. Soc. Translations of Mathematical Monographs (1967).

[Ba] W.Ballmann, *Singular spaces of non-positive curvature* : Chapter 10 of "Sur les groupes hyperboliques d'après Mikhael Gromov" (ed. E.Ghys, P.de la Harpe) Progress in Maths. 83, Birkhäuser (1990).

[BaGS] W.Ballmann, M.Gromov, V.Schroeder, *Manifolds of non-positive curvature* : Progress in Maths. 61, Birkhäuser (1985).

[Br] M.R.Bridson, *Geodesics and curvature in metric simplicial complexes* : in "Group theory from a geometric viewpoint" (ed. E.Ghys, A.Haefliger, A.Verjovsky), World Scientific (1991) 373–463.

[BurZ] Yu.D.Burago, V.A.Zalgaller, *Geometric inequalities* : Grundlehren der Mathematischen Wissenschaften 285, Springer Verlag (Translation of Russian edition of 1980).

[Bus] H.Busemann, *The geometry of geodesics* : Pure and Applied Maths. 6, Academic Press (1955).

[CD] R.Charney, M.Davis, *Singular metrics of non-positive curvature on branched covers of riemannian manifolds* : preprint, Ohio State (1990).

[E] D.B.A.Epstein et. al. *Word processing in groups* : Jones and Bartlett (1992).

[Ga] M.E.Gage, *Curve shortening on surfaces* : Annal. Scient. de l'Éc. Norm. Sup. **23** (1990) 229–256.

[Gr] M.Gromov, *Hyperbolic groups,* in "Essays in group theory" : (ed. S.M.Gersten) M.S.R. I. Publications No. 8, Springer (1988) 75–263.

[HR] C.D.Hodgson, I.Rivin, *A characterisation of compact convex polyhedra in hyperbolic 3-space* : preprint (1991).

[O1] R.Osserman, *Bonnesen-style inequalities* : Amer. Math. Monthly **86** (1977) 1–29.

[O2] R.Osserman, *The isoperimetric inequality* : Bull. Amer. Math. Soc. **84** (1978) 1182–1238.

[P] F.Paulin, *Constructions of hyperbolic groups via hyperbolisations of polyhedra* : in "Group theory from a geometric viewpoint" (ed. E.Ghys, A.Haefliger, A.Verjovsky), World Scientific (1991) 313–372.

[S] J.J.Stoker, *Geometric problems concerning polyhedra in the large* : Comm. Pure and Applied Math. **21** (1968) 119–168.

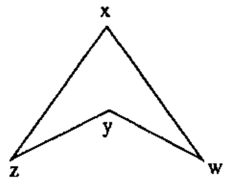

Figure 1a

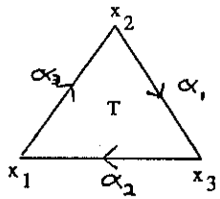

Figure 2a

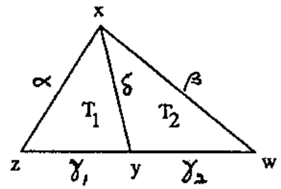

Figure 2b

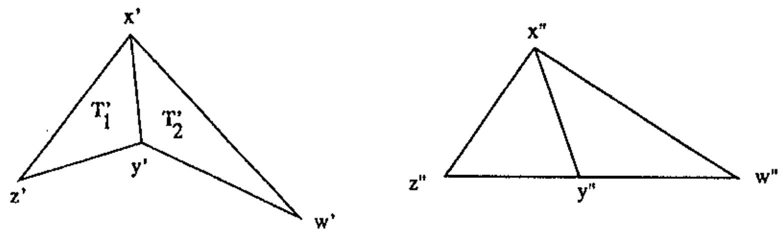

Figure 2c

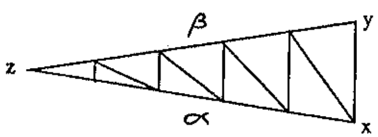

Figure 2d

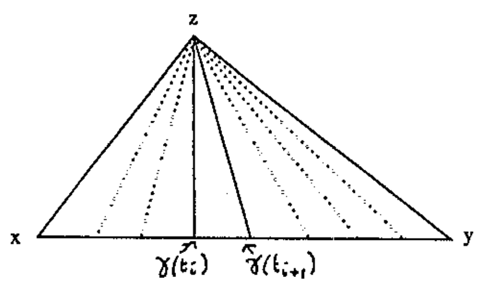

Figure 2e

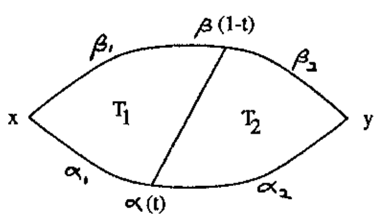

Figure 2f

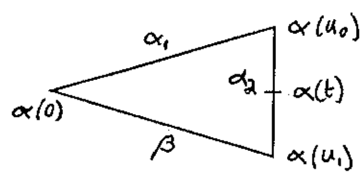

Figure 2g

(i)

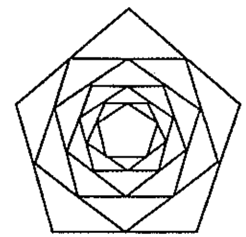

Figure 3a

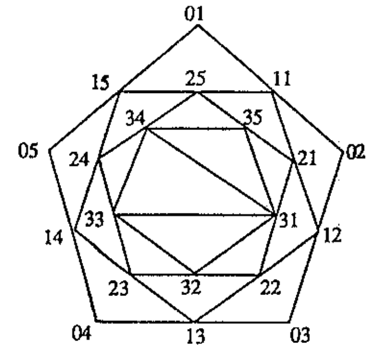

Figure 3b

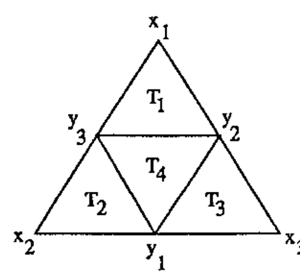

Figure 4a

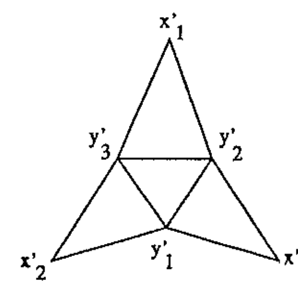

Figure 4b

(ii)