THE LENGTH SPECTRUM OF A COMPACT CONSTANT CURVATURE COMPLEX IS DISCRETE

NOEL BRADY [1] AND JON MCCAMMOND [2]

In this short note we prove that the length spectrum of a compact constant curvature complex is discrete. After recalling the relevant definitions and reducing to a relatively simple situation, the result follows easily from a foundational result about real semi-algebraic sets and the Morse-Sard theorem. We conclude with a conjecture which remains open, a few remarks and an easy application.

Definition 1 (Constant curvature complexes)**.** A *constant curvature complex* with curvature κ is a polyhedral cell complex where each cell has been assigned a metric of curvature κ with the requirement that the induced metrics agree on the overlaps. In the positive curvature case, there is an additional restriction that each closed polyhedral cell must lie in some open hemisphere. The generic term is M_κ*-complex* where κ is the curvature which all of the cells have in common. In practice, constant curvature complexes can be rescaled so that κ is 1, 0 or -1 and subdivided into simplices. Thus, we only need to consider piecewise spherical, piecewise Euclidean and piecewise hyperbolic simplicial complexes. See [3] for a comprehensive discussion of these types of complexes.

It is an early foundational result of Bridson that compact M_κ-complexes are *geodesic metric spaces*. In other words, every pair of points in a compact M_κ-complex is connected by a geodesic, where a *geodesic* is a length minimizing path. In order to define the length spectrum, we also need the definition of a local geodesic. A *local geodesic* is a local isometric embedding of an interval into X and a *closed geodesic loop* is a local isometric embedding of a metric circle into X.

Definition 2 (Length spectrum)**.** Let X be a compact M_κ-complex and let x and y be points in X. The lengths of all the local geodesic paths from x to y is called the length spectrum from x to y. The lengths of all the closed geodesic loops in X is simply its *length spectrum*.

In Bridson-Haefliger [3, p. 122], a relatively easy compactness argument was able to establish the discreteness of the length spectrum between specific points. That the same statement holds for the set of all closed geodesics, is our main result.

Main Theorem. *The length spectrum of a compact M_κ-complex is discrete.*

Unlike the proof of the result between specific endpoints, however, the closed geodesic situtation requires a very different type of argument, one more reminenscent of the proof of Theorem B in [4]. In that case, the problem of deciding

[1]Partially supported by NSF grant no. DMS-0123344
[2]Partially supported by NSF grant no. DMS-0101506
Date: January 12, 2004.
1991 *Mathematics Subject Classification.* 20F65.
Key words and phrases. nonpositive curvature, CAT(0), decidability.

whether a compact M_κ-complex is non-positively curved in the Alexandrov sense was reduced to a decidability result from real algebraic geometry. Here the proof naturally splits into three parts. First, the theory of constant curvature complexes reduces the problem to a more managable size, then the Morse-Sard theorem from differential topology reduces the problem even further, and finally real algebraic geometry is once again used to complete the proof.

1. Constant curvature complexes

The goal of this section is to reduce the proof of the Main Theorem to the case where everything takes place in a single circular gallery. We begin with some definitions.

Definition 3 (Galleries)**.** Let X be a piecewise constant curvature complex and let γ be a local geodesic path in X. Consider the sequence of open cells through which γ passes. This sequence is essentially what we call a *linear gallery.* If γ is a closed geodesic loop, then this sequence is given a cyclic ordering and the sequence is called a *circular gallery.* Instead of a precise definition, consider the example shown in Figure 1. The upper left shows a very simple complex (which is essentially the boundaries of two Euclidean tetrahedra identified along one edge) along with a local geodesic path from x to y. On the upper right is the corresponding linear gallery. The lower left is the result of concatenating copies of each of the closed cells through which the local geodesic passes rather than the open cells, and the difference between these two is the *boundary* of the gallery shown in the lower right. The circular version constructs a similar complex homotopy equivalent to a circle.

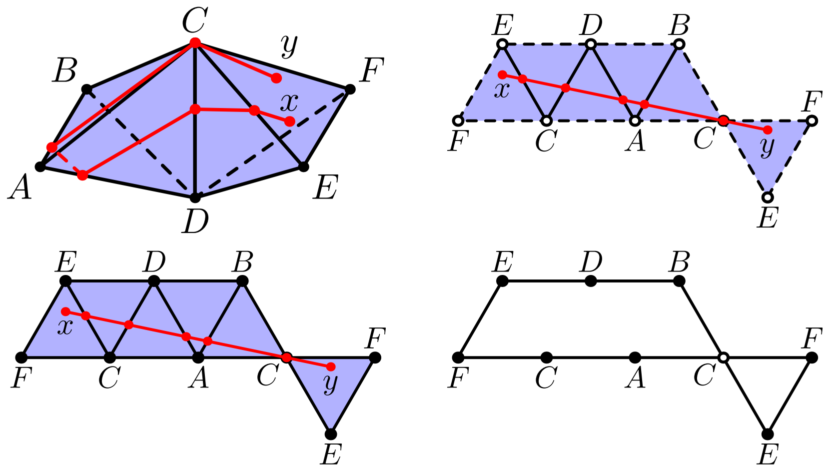

Figure 1. A 2-complex, a linear gallery, its closure and its boundary.

Suppose that X is a compact constant curvature complex whose length spectrum is not discrete. In this case we can choose an infinite sequence of closed geodesic loops $\{\gamma_i\}$, $i = 1, 2, 3, \ldots$ in X with distinct lengths bounded above by some number M. Next, Bridson showed that for a local geodesic in a compact constant curvature complex there is an approximately linear relationship between the number of simplices in its linear gallery and its length (technically, there exist quasi-isometry constants describing this relationship which only depend on the shapes of the metrics given to the cells). As a result, there only exist a finite number of circular galleries which can be determined by closed geodesic loops of length less than M. By passing to a subsequence if necessary, we can assume that all of the γ_i in the sequence determine the same (open) circular gallery $\mathcal{G}$. In other words, all of the closed geodesic loops pass through the exact same sequence of open cells in the exact same order.

Remark. At this point, the piecewise Euclidean and piecewise hyperbolic cases would be easy to complete on their own using the theory of non-positively curved spaces [3]. In particular, it is nearly immediate from this perspective that a piecewise Euclidean circular gallery might contain a continuum of closed geodesic loops, but the non-positively curved nature of the situation ensures that all such closed geodesic loops have exactly the same length. In a piecewise hyperbolic circular gallery the conclusion is even stronger: it contains at most one closed geodesic loop, period. Hence each piecewise Euclidean and piecewise hyperbolic circular gallery contributes at most one number to the length spectrum. While the piecewise spherical case is the only difficult one, it is also the one of most interest in applications. See our remarks in the final section.

Since our proof in the spherical case readily extends to the Euclidean and hyperbolic situations, we continue to view $\mathcal{G}$ as a constant curvature circular gallery without artificially restricting its curvature.

2. Differential topology

In this section we use the Morse-Sard theorem to prove that the length spectrum for a single circular gallery $\mathcal{G}$ has measure zero. We begin by carefully examining the loops in $\mathcal{G}$.

Definition 4 (Simplicially taut loops)**.** As remarked in [4], if $\mathcal{G}$ is a linear (or circular) gallery determined by a local geodesic (or closed geodesic loop) γ then the dimensions of the cells alternately increase and decrease. The cells whose dimension is a local maximum are called *top cells* and the others whose dimension is a local minimum are *bottom cells*. In addition, the intersection of γ with a bottom cell is always a single point and the intersection with a top cell is always an open interval. We call the points of intersection with the bottom cells *transition points*. Note that in order for this description to work in low dimensions, a vertex is considered to be both an closed simplex and an open simplex, and its boundary is empty. Finally, since γ can be completely reconstructed from its transition points, it makes sense to study the collection of all piecewise geodesic paths determined by a sequence of points, one selected at random from each (open) bottom cell in the gallery. We call the piecewise geodesic loop determined by this choice of points a *simplicially taut loop* and we denote the set of all simplicially taut loops in $\mathcal{G}$ by $\textsc{Loops}(\mathcal{G})$. Because each bottom cell is an open spherical / Euclidean / hyperbolic simplex, the set $\textsc{Loops}(\mathcal{G})$ comes equipped with the structure of a real analytic Riemannian manifold.

The connection with differential topology comes from the fact that the closed geodesic loops are critical points of a suitably chosen function. Recall that a *critical point* for a smooth map between smooth manifolds is one where the tangent map is not onto and that the set of *critical values* for this map is the image of its set of critical points. We only need the weak version of the Morse-Sard Theorem which asserts that for any smooth map between smooth manifolds, the set of critical values has measure zero (see for example [5]). Thus, in order to show that the length spectrum of $\mathcal{G}$ has measure zero, it is sufficient to find a smooth function $d : \textsc{Loops}(\mathcal{G}) \to \mathbb{R}$ whose critical points include all closed geodesic loops. We claim that the function $d : \textsc{Loops}(\mathcal{G}) \to \mathbb{R}$ which records the length of the simplicially taut geodesic corresponding to a point in $\textsc{Loops}(\mathcal{G})$ has both of these properties.

To see that d is smooth (and in fact real analytic) it is enough to show that each of the functions involved in this finite sum are smooth, but this is immediate since each such function is the composition of two smooth maps. The first map is the projection of $\textsc{Loops}(\mathcal{G})$ onto the product of the two bottom cells, say σ and σ', involved in this particular distance calculation, and the other is a restriction of the distance function $\text{dist} : M \times M \to \mathbb{R}$ where M is either $\mathbb{S}^n$, $\mathbb{R}^n$ or $\mathbb{H}^n$ and n is the dimension of the top cell connecting σ and σ' in $\mathcal{G}$. More precisely, there are isometric embeddings of σ and σ' as faces of the top cell between them, which is itself embedded in M. In this context, it is clear that the function we care about is the restriction of the distance function on M. Each of these distance functions is well-known to be real analytic away from the diagonal (i.e. pairs of identical points), and in the spherical case, away from pairs of points which are antipodal. Because the bottom cells σ and σ' are embedded disjointly in M (recall they are open) and do not include antipodal points (because of the open hemisphere restriction on the size of cells in the spherical case), the distance function is real analytic on this restriction and thus the function d is real analytic as well.

Now that smoothness has been established, the fact that closed geodesic loops in $\mathcal{G}$ correspond to points in $\textsc{Loops}(\mathcal{G})$ which are critical points of d follows readily. Let γ be a closed geodesic loop in $\mathcal{G}$ and let x be the point in $\textsc{Loops}(\mathcal{G})$ to which it corresponds. Since d is smooth and $\textsc{Loops}(\mathcal{G})$ is open, it is sufficient to show that the partials at x are zero in a sufficient number of directions. Consider, for example, the change in d when one of the transition points of γ is varied in a specific direction in its bottom cell (i.e. along a short local geodesic arc in its bottom cell) and all of the other transition points are left fixed. The geodesics from the previous transition point to this short arc all lie in a common geodesic triangle embeddable in $\mathbb{S}^2$, $\mathbb{R}^2$, or $\mathbb{H}^2$. Similarly, the geodesics connecting this arc to the following transition point lie in a common geodesic triangle. These two geodesic triangles share the edge of variation, and it is clear that a path through this configuration which is locally geodesic must pass through a point which determines an element of $\textsc{Loops}(\mathcal{G})$ at which the partial of d in the direction of variation is zero. In the Euclidean and hyperbolic cases, this point is a local minimum of d in this direction, but in the spherical case, it might be a local minimum, locally flat, or a local maximum depending on whether the distance from the previous transition point through this special point to the following transition point is less than, equal to, or more than π. By varying the direction and the transition point under consideration, it is easy to see that the point x corresponding to γ must be critical.

3. Real algebraic geometry

In this section we use real algebraic geometry to show that the set of critical points of $d : \textsc{Loops}(\mathcal{G}) \to \mathbb{R}$ has the topology of a real semi-algebraic set. Recall that a *real semi-algebraic set* is a subset of Euclidean space which can be described using a Boolean combination of polynomial equations and inequalities in the standard set of real variables. Once this is established, the main result follows quickly. It is a standard result in real algebraic geometry that a real semi-algebraic set has only a finite number of connected components [1], so its image under a continuous map such as d (i.e. the set of critical values) also has only finitely many components. In the previous section we established that the critical values of d have measure zero in $\mathbb{R}$. Combining these two results implies that d has only a finite number of critical values. As a consequence, the length spectrum of $\mathcal{G}$ is finite and, by the argument in Section 1, the length of the original constant curvature complex is discrete. Thus, it only remains to show that the critical points form a real semi-algebraic set. To do this we first embed $\textsc{Loops}(\mathcal{G})$ into a high dimensional Euclidean space.

Definition 5 (Standard embeddings)**.** Let σ be one of the bottom cells of $\mathcal{G}$ of dimension n and consider the portion of $\mathbb{R}^{n+1}$ defined by the equation $\kappa(x_1^2 + \ldots + x_n^2) + x_{n+1}^2 = 1$ and the inequality $x_{n+1} > 0$ where κ is the curvature. When κ is 1 / 0 / -1, respectively, then this describes the upper open hemisphere of the standard embedding of $\mathbb{S}^n$ into $\mathbb{R}^{n+1}$ / the projective embedding of $\mathbb{R}^n$ into $\mathbb{R}^{n+1}$ defined by the equation $x_{n+1} = 1$ / the standard hyperboloid used as a model for hyperbolic n-space inside $\mathbb{R}^{n+1}$. Depending on the curvature κ we choose an embeding of σ into the appropriate model space so that the image is isometric to σ when the standard spherical / Euclidean / hyperbolic metric is used to measure distances within the model.

Notice that in each case the image a bottom cell can be described by a finite list of polynomial equations and inequalities in the $n+1$ variables, x_i, $i = 1, \ldots, n+1$. For example, a spherical / Euclidean / hyperbolic triangle can be described using three standard variables (x_1, x_2, and x_3), one equation ($\kappa(x_1^2 + x_2^2) + x_3^2 = 1$) and three linear inequalities to specify the sides. The equations of the sides really are *linear* inequalities because in all three cases, the "lines" in the model are found by intersecting a plane through the origin with the model space. Notice that the inequality $x_3 > 0$ is not needed because it is a consequence of the other three inequalities.

Definition 6 (Variables)**.** The first n variables are called the *intrinsic variables* of this bottom cell, the $n+1$-st variable is the *special variable* for this bottom cell and the set of all $n+1$ variables are the *standard variables.* When we need to distinguish between intrinsic and standard variables we use u_i for the former and x_i for the latter.

Combining the chosen embeddings for each bottom cell creates an embedding of $\textsc{Loops}(\mathcal{G})$ into R^N for some large value of N (equal to the sum of the dimensions of the bottom cells plus the number of bottom cells). Moreover, the image can be described using a finite list of polynomial equations and inequalities in the N standard variables. Thus, we have parameterized the set of all simplicially taut loops by a real semi-algebraic subset of Euclidean N-space. One might wonder whether the choice of particular embeddings mattered, but since the isometries of $\mathbb{S}^n$, $\mathbb{R}^n$, and $\mathbb{H}^n$ interpreted inside $\mathbb{R}^{n+1}$ are *linear* transformations of $\mathbb{R}^{n+1}$, the net effect on the polynomial equations and inequalities would be a linear substitution on the set of variables. Thus, changing the chosen embeddings of the bottom cells has only a minimal effect on the resulting equations and inequalities. With this in mind, fix once and for all a specific embedding of $\text{Loops}(\mathcal{G}) \subset \mathbb{R}^N$.

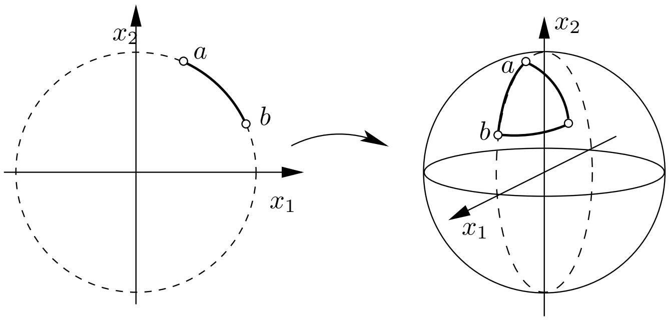

Figure 2. Parameterizing a circular arc bottom cell and its embedding into the boundary of a spherical triangle top cell.

Next, we also choose similar embeddings of the top cells into model spaces. Once we have fixed such an embedding of a top cell, there is a unique linear map from the Euclidean space containing the previous bottom cell so that its image lies in the boundary of the embedded top cell in the appropriate way. Such a map does in fact exist precisely because the lower dimensional model spaces embed in the higher dimensional ones in an obvious way and, once there, the image can be moved into the correct position by a linear transformation of this higher dimensional Euclidean space. Notice that the summand of the function d which depends on this particular top cell can now be written explicitly as either $\cos^{-1}(\vec{x}\cdot\vec{y})$, $\sqrt{\vec{x}\cdot\vec{y}-1}$, or $\cosh^{-1}(-\vec{x}\circ\vec{y})$ where the coordinates of $\vec{x}$ and $\vec{y}$ are linear functions of the standard variables for the appropriate bottom cell. In these equations the notations $\vec{x}\cdot\vec{y}$ and $\vec{x}\circ\vec{y}$ refer to the standard Euclidean and Lorenztian inner products, respectively. In other words, this portion of d is can be defined by evaluating a quadratic polynomial in the standard variables in a rather simple function. Finally, notice that the partial derivative with respect to one of the standard variables of either bottom cell is, in all three cases, a linear polynomial divided by the square root of a quadratic polynomial. Moreover, the polynomial under the square root sign is independent of the variable involved in the differentiation.

The final step is to show that the restriction of $\text{Loops}(\mathcal{G}) \subset \mathbb{R}^N$ to those points which are critical points for the function d can also be described using only polynomial equations and inequalities in the standard variables. One difficulty here is that we have used an excessive number of variables to embed $\text{Loops}(\mathcal{G})$ in a Euclidean space. To fix this we can also parameterize $\text{Loops}(\mathcal{G})$ using only the intrinsic variables. More explicitly, the map which projects the points in each model space onto its first n coordinates has an inverse which sends the point $\vec{u} = (u_1, \ldots, u_n)$ to the point

$$\vec{x} = (x_1, \ldots, x_n, x_{n+1}) = \left(u_1, \ldots, u_n, \sqrt{1 - \kappa(u_1^2 + \cdots + u_n^2)}\right)$$

This parameterization map is particularly well-behave in that every partial derivative is a rational polynomial in the standard variables. Concretely,

$$\frac{\partial x_i}{\partial u_j} = \begin{cases} \delta_{ij} & \text{if } i \leq n \\ \dfrac{-\kappa x_j}{x_{n+1}} & \text{if } i = n+1 \end{cases}$$

If we let U denote the product of the projections of the embedded bottom cells onto their intrinsic coordinates, then these parameterization maps combine to give a smooth homeomorphism f between U and $\textsc{Loops}(\mathcal{G})$. Thus, a point in $\textsc{Loops}(\mathcal{G})$ is a critical point of d if and only if it is the image of a critical point in U of $d \circ f$. These latter points are characterized by the fact that the partial of $d \circ f$ with respect to each intrinsic variable is zero. Finally, if we are differentiating with respect to an intrinsic variable u_i in a bottom cell of dimension n, then the only standard variables which depend on u_i are the variables x_i and x_{n+1} assigned to the same bottom cell, and the only summands of d which depend on these variables are the terms involving the neighboring top cells. As a result, the partial derivative of $d \circ f$ with respect to u_i can be written as the sum of four terms, each of which involves a rational polynomial divided by the square root of polynomial - all using the standard variables. (We are using here both the characterization of the partials of x_i with respect to u_j and the partials of a term of d with respect one of the standard variables that we derived above.) Moreover, the first two terms and the last two terms involve the same polynomial under the square root sign. Thus, this can be simplified to two terms with rational polynomials divided by square roots of polynomials. We get one such expression in the standard variables for each of the intrinsic variables. Setting each of these to zero, we have found the exact conditions which need to be satisfied by a point in $\textsc{Loops}(\mathcal{G})$ for it to be a critical point for the map d. It is now an easy exercise to rewrite each of these equations involving only two terms with square roots into polynomial equations and inequalities in the same set of variables, where the new system has the same solution set as the old. As a consequence, the set of critical points for d as a subset of $\textsc{Loops}(\mathcal{G}) \subset \mathbb{R}^N$ is a real semi-algebraic set. As argued at the beginning of the section, this completes the proof of the main theorem.

4. Final comments

We conclude this note with a conjecture, a few additional remarks, and an application. The conjecture is motivated by the facts noted above about the closed geodesic loops in piecewise Euclidean and piecewise hyperbolic circular galleries.

Conjecture (Trivial length spectrum)**.** *The length spectrum of a piecewise spherical circular gallery $\mathcal{G}$ should contain at most one value.*

An even stronger conjecture, which is known to be true in the Euclidean and hyperbolic cases, is that the set of closed geodesic loops in $\mathcal{G}$ should correspond to a connected subset of $\textsc{Loops}(\mathcal{G})$. In other words, if α and β are closed geodesic loops in $\mathcal{G}$, then the length of α and β should be equal and there should be a homotopy from α to β which only passes through closed geodesic loops with this particular length. At present, the only bounds we know of on the size of the length spectrum in the spherical case come from an examination of the degrees of the polynomials which define the real semi-algebraic set of critical points.

Turning our attention from the set of critical values to the size of the set of critical points, we note that the Main Theorem should definitely *not* be interpreted to mean that there are only a finite number of closed geodesic loops. In fact, it is very easy to construct examples where the dimension of the set of critical points, and of closed geodesic loops, is large.

Remark (Number of closed geodesics)**.** Consider a sequence of n all-right spherical tetrahedra (i.e. all edges have length $\frac{\pi}{2}$), and let $\mathcal{G}$ be the circular gallery formed by identifying an edge in one tetrahedron to an edge in the next in such a way that the two edges used in each tetrahedron are non-adjacent. In this example, every simplicially taut loop has the same length, namely $\frac{n\pi}{2}$. In particular, its image under d is trivial, every point in the n-manifold $\textsc{Loops}(\mathcal{G})$ is critical, and every simplicially taut loop is a closed geodesic.

It is also instructive to consider a slight modification of this example where every edge which is not shared between two tetrahedra has length $\frac{\pi}{2} + \epsilon$. When ϵ is small and positive $\mathcal{G}$ contains no closed geodesic loops, and when ϵ is small and negative $\mathcal{G}$ contains exactly one. This extreme sensitivity to shape is one difficulty inherent in problems of this type. Our next remark is about another, more subtle, difficulty that our proof as given narrowly avoids.

Remark (Morse-Sard for real analytic functions)**.** At one point we believed that we could eliminate the need to use real algebraic geometry by establishing the main result via the real analytic version of the Morse-Sard theorem. As stated and proved in [8], this asserts that if $f : U \to \mathbb{R}$ is a real analytic function defined over an open set $U \subset \mathbb{R}^N$ and $C \subset U$ is compact, then f restricted to C has only a finite number of critical values. Unfortunately, even though the function d is a real analytic function on an open subset of $\mathbb{R}^N$ containing $\textsc{Loops}(\mathcal{G})$, it is not real analytic on any open subset containing the closure of $\textsc{Loops}(\mathcal{G})$. The problem is that there are points in the boundary of $\textsc{Loops}(\mathcal{G})$ where d is not real analytic, or even C^1. Geometrically, these bad points arise because we are extending the appropriate distance function to a place where points collide, and there are no neighborhoods of these points where the distance functions are differentiable.

Finally, we conclude with a remark about an application of the Main Theorem to the study of curvature in locally geodesic surfaces.

Remark (Locally geodesic surfaces)**.** Let $f : D \to X$ be an isometric immersion of a metric surface D. We call this immersion *locally geodesic* if for all points $d \in D$, the link of d is sent to a local geodesic in the link of $f(d)$. Recall that the *link of a point* is the set of unit tangent at that point. In a 2-dimensional piecewise Euclidean complex, every null-homotopic curve bounds a locally geodesic surface, but this fails in dimensions 3 and higher. This is also one of the key reasons why some theorems which hold in dimension 2 fail to easily generalize to higher dimensions. For example, Chris Hruska proves in [6] that in dimension 2 the isolated flats property is equivalent to the relative fellow travel property and to the relatively thin triangles property. Moreover, he conjectures that the same result is true in all dimensions. In his proof, however, he heavily uses the fact that null-homotopic curves bound locally geodesic surfaces and that the curvature of points in locally geodesic surfaces mapped into finite 2-dimensional complexes is bounded away from 0. By the *curvature of a point* in the interior of D we mean 2π minus the length of its link. See also [7] where a more general definition of curvature is discussed. An immediate corollary of Main Theorem is that the second portion of his proof immediately generalizes to arbitrary dimensions.

Corollary (Quantantizing curvature)**.** *If $D \to X$ is a locally geodesic surface in a compact piecewise Euclidean complex X, then those points in the interior of D with nonzero curvature have curvatures uniformly bounded away from 0. Moreover, this bound depends only on X and not on D. As a consequence, if the total amount of negative (or positive) curvature is known, then there is a bound on the number of points in D whose curvature is less than (or more than) zero.*

References

[1] Saugata Basu, Richard Pollack, and Marie-Françoise Roy. *Algorithms in real algebraic geometry*, volume 10 of *Algorithms and Computation in Mathematics*. Springer-Verlag, Berlin, 2003.

[2] Brian H. Bowditch. Notes on locally CAT(1) spaces. In *Geometric group theory (Columbus, OH, 1992)*, pages 1–48. de Gruyter, Berlin, 1995.

[3] Martin R. Bridson and André Haefliger. *Metric spaces of non-positive curvature*. Springer-Verlag, Berlin, 1999.

[4] Murray Elder and Jon McCammond. CAT(0) is an algorithmic property. To appear in Geometriae Dedicata.

[5] Morris W. Hirsch. *Differential topology*, volume 33 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 1994. Corrected reprint of the 1976 original.

[6] G. Christopher Hruska. *Nonpositively curved spaces with isolated flats*. PhD thesis, Cornell University, 2002.

[7] Jonathan P. McCammond and Daniel T. Wise. Fans and ladders in small cancellation theory. *Proc. London Math. Soc. (3)*, 84(3):599–644, 2002.

[8] J. Souček and V. Souček. Morse-Sard theorem for real-analytic functions. *Comment. Math. Univ. Carolinae*, 13:45–51, 1972.

Dept. of Mathematics, University of Oklahoma, Norman, OK 73019
E-mail address: `nbrady@math.ou.edu`

Dept. of Mathematics, U.C. Santa Barbara, Santa Barbara, CA 93106
E-mail address: `jon.mccammond@math.ucsb.edu`