Contemporary Mathematics

Nonpositive curvature and complex analysis

Stephen M. Buckley

Abstract. We discuss a few of the metrics that are used in complex analysis and potential theory, including the Poincaré, Carathéodory, Kobayashi, Hilbert, and quasihyperbolic metrics. An important feature of these metrics is that they are quite often negatively curved. We discuss what this means and when it occurs, and proceed to investigate some notions of nonpositive curvature, beginning with constant negative curvature (e.g. the unit disk with the Poincaré metric), and moving on to CAT(k) and Gromov hyperbolic spaces. We pay special attention to notions of the boundary at infinity.

Contents

1. Introduction 3
2. Hyperbolic Geometry 5
3. Other metrics in complex analysis and potential theory 17
4. CAT(k) and related curvature conditions 21
5. Gromov hyperbolicity 32
6. Appendix: terminology of metric geometry 43
References 45

1. Introduction

In this course, we are interested in the geometry of metric spaces which are negatively (or, more generally, nonpositively) curved in some sense. Roughly speaking, this means that in these spaces, if two observers move at the same constant speed from a common origin in different "straight line" directions (more precisely along distinct geodesic paths), then their paths bend away from each other when compared with the Euclidean picture as we move away from the origin. Equivalently, their mutual distance $f(t)$ at time t is a convex function of time, i.e. $t \mapsto f(t)/t$ is an increasing function. We will relate some curvature notions to specific metrics that are important in complex analysis and potential theory.

Suitable definitions of negative or nonpositive curvature lead to a notion of a boundary at infinity in such spaces, which is a central concept in the theory of such spaces. For some of the specific metrics that we consider, we relate this boundary at infinity to a topological boundary of the space.

Let us first briefly discuss the history of Euclidean geometry. Euclid's *Elements* consists of 13 books, written at about 300BC, that are mainly concerned with geometry (although they also contain some number theory and the *method of exhaustion* which is related to integration). It is the earliest known systematic discussion of geometry.

Book 1 begins with 23 definitions (of a point, line, etc.) and 10 axioms. Of these axioms, the following five are termed *Postulates*:

(1) Any two points can be joined by a straight line.
(2) Any straight line segment can be extended indefinitely in a straight line.
(3) Given any straight line segment, a circle can be drawn having the segment as radius and one endpoint as center.
(4) All right angles are congruent.
(5) **Parallel Postulate:** If two lines intersect a third in such a way that the sum of the inner angles on one side is less than two right angles, then the two lines inevitably must intersect each other on that side if extended far enough.

Euclid's other five axioms, his *Common Notions*, are mostly statements about equalities (such as transitivity of equality) and do not concern us.

For two millenia, mathematicians were troubled by the Parallel Postulate of Euclid, principally because it is more complex and rather different from the other Postulates. For most of that time, mathematicians attempted to prove that it followed from the other postulates, and succeeded in finding a large variety of false "proofs" which all fail because they make some assumption that is equivalent to the Parallel Postulate.

One mathematician responsible for several false "proofs" was Farkas Bolyai. When his son, János, also became obsessed with the Parallel Postulate, Farkas wrote to him

> For God's sake, I beseech you, give it up. Fear it no less than sensual passions because it too may take all your time and deprive you of your health, peace of mind, and happiness in life.

But János took a different approach and instead showed that dropping the Parallel Postulate lead to a new, interesting, and seemingly consistent *hyperbolic geometry* which starts by replacing the Parallel Postulate by the axiom stating that there are at least two different lines through a given point a that do not intersect a given line that is disjoint from a.

János Bolyai's important breakthrough was published in 1832 as an 24-page appendix to a mathematics textbook by his father. From there things went downhill for him. First Gauss wrote to János' father about this appendix:

> If I commenced by saying that I must not praise this work you would certainly be surprised for a moment. But I cannot say otherwise. To praise it, would be to praise myself. Indeed the whole contents of the work, the path taken by your son, the results to which he is led, coincide almost entirely with my meditations, which have occupied my mind partly for the last thirty or thirty-five years.

Then János discovered that Lobachevski had published the same advances about three years before him (only in Russian). Furthermore mathematicians were not ready to give proper recognition to either Bolyai's or Lobachevski's work because neither had proven this strange new geometry to be consistent.

The lack of a proof of consistency is not viewed nowadays as a flaw in the work of Bolyai and Lobachevski. In fact, we still do not know whether or not the theories of hyperbolic and Euclidean geometry are consistent![1] In the mid-nineteenth century many mathematicians did not accept hyperbolic geometry because of the lack of a proof of its consistency, but overlooked the same flaw in Euclidean geometry because it had been around for a long time and seemed to correspond to the world around us. In 1868, Beltrami gave what we now call the Poincaré metric in the unit disk, the Poincaré metric in the upper half-plane, and the Klein projective disk metric, as three models of hyperbolic geometry. This implied that hyperbolic geometry was equiconsistent with Euclidean geometry, i.e. it is consistent if and only if Euclidean geometry is consistent. Finally the world was ready to accept hyperbolic geometry, and the theory was developed further by people such as Riemann and Poincaré.

Euclid's set of axioms are an incomplete description of Euclidean geometry, since some of his proofs require the use of "common sense" that does not follow from his axioms. To fix this, we can add such extra assumptions as extra axioms, including the following ones:

[1]Tarski gave a set of axioms for *Elementary Euclidean Geometry* (a substantial part of Euclidean geometry, specifically consisting of all that can be formulated in first order logic with identity, without the use of set theory) and showed it to be consistent, complete, and decidable.

- Of three points on a line exactly one is between the other two.
- Two sides of a triangle and the angle between those sides determine it up to congruence;

Certain continuity assumptions also need to be added, or we cannot prove for instance that two circles, or one line and a circle, intersect in those cases where it it is "obviously" true. There are also many alternative axiom systems for Euclidean geometry, notably those by Hilbert, Birkhoff, and MacLane.

Remarkably, of all the (augmented) set of Euclidean axioms, the only one that fails for the hyperbolic plane—once we give suitable meanings to the basic concepts such as lines and circles—is the Parallel Postulate. If we drop this postulate, the resulting theory of geometry is referred to as *Neutral Geometry* (or *Absolute Geometry*). This theory includes a large part of Euclidean geometry and so all of this theory is valid also for the hyperbolic plane. In planar Neutral Geometry, the Parallel Postulate is equivalent to the following alternative axiom to which we refer later:

Playfair's Axiom: Through a point not on a given straight line, one and only one line can be drawn that never meets the given line.

2. Hyperbolic Geometry

Here we review some of the fundamentals of the theory, concentrating on the hyperbolic plane $\mathbb{H}^2$, and also look at some particular models of $\mathbb{H}^2$ that arise in complex analysis and related areas.

The three models of $\mathbb{H}^2$ that we have chosen to examine each have their own advantages as ways of looking at the hyperbolic plane. In view of the likely background of students taking this course, we will give only a quick overview of those parts of the theory that are covered in the typical introductory graduate course in complex analysis. There are many excellent books that cover most of the hyperbolic geometry parts of this section, for instance the books by Anderson [**3**] and Beardon [**9**].

2.1. The Poincaré metric on a simply connected domain. The *Poincaré* or *hyperbolic metric* in the upper half-plane $H = \{z = x+iy \mid y > 0\}$ is given infinitesimally at a point $z = x + iy \in H$ by

$$ds^2 = \frac{dx^2 + dy^2}{y^2} = \frac{dz d\overline{z}}{y^2},$$

and so the hyperbolic area element is

$$\frac{dxdy}{y^2}.$$

The associated distance function is obtained as always in Riemannian geometry by integrating the infinitesimal distance over paths to define arclength, and then taking an infimum of this arclength over all paths between the desired pair of points. For this

metric, the infimum can be computed and the resulting formula for the distance function is

$$\rho_H(z_1, z_2) = 2 \tanh^{-1} \left| \frac{z_1 - z_2}{z_1 - \overline{z_2}} \right|, \qquad z_1, z_2 \in H\,.$$

The Poincaré metric in the unit disk $D = \{z = x + iy : |z| < 1\}$ is given infinitesimally at a point $z = x + iy \in D$ by

$$ds^2 = \frac{4(dx^2 + dy^2)}{(1 - x^2 - y^2)^2} = \frac{4\,dz d\overline{z}}{(1 - |z|^2)^2}\,.$$

and so the hyperbolic area element is

$$\frac{4\,dxdy}{(1 - x^2 - y^2)^2} = \frac{4\,dxdy}{(1 - |z|^2)^2}\,.$$

The associated distance function is

$$\rho_D(z_1, z_2) = 2 \tanh^{-1} \left| \frac{z_1 - z_2}{1 - z_1\overline{z_2}} \right|, \qquad z_1, z_2 \in D\,.$$

In complex analysis, the most important property of the Poincaré metric is that holomorphic mappings are contractions with respect to it. More precisely, we have:

THEOREM (Schwarz-Pick). *A holomorphic mapping $f : D \to D$ is a contraction with respect to ρ_D. It is an isometry if and only if f is an automorphism (i.e. a Möbius self-map of D).*

A similar result holds in H. More generally, we can define the Poincaré metric ρ_G in a simply connected domain $G \subset \mathbb{C}$ by pulling back the Poincaré metric ρ_D with respect to a Riemann mapping $f : G \to D$. The resulting metric ρ_H on the upper half-plane coincides with the one defined previously.

The following facts about the isometry group G of either the Poincaré disk or Poincaré upper half-space are very useful:

- G is transitive.
- Every $g \in G$ is a Möbius map.

Also useful is the fact that the Möbius map $z \mapsto (z - i)/(z + i)$ acts as a Riemann map for the upper half-plane, and the well-known fact that Möbius maps take circles and lines to circles and lines. The typical use of these facts involves reducing a statement involving a general point $z \in D$ to a statement involving the origin by using an isometry to transport z to 0.

As abstract metric spaces, every simply connected domain with the Poincaré metric attached is the same space, since they are all isometric. For hyperbolic geometry, the most important thing about the Poincaré metrics on H and on D is that they are models for the hyperbolic plane $\mathbb{H}^2$.

2.2. The Klein model. The Klein model of $\mathbb{H}^2$ consists of the unit disk, which we now call K, together with a distance function given by

$$d(z_1, z_2) = \frac{1}{2} \log[z_1^*, z_1, z_2, z_2^*], \qquad z_1, z_2 \in K\,,$$

where the *cross-ratio* $[\cdot, \cdot, \cdot, \cdot]$ is defined by the formula

$$[z_1, z_2, z_3, z_4] = \frac{(z_1 - z_3)(z_2 - z_4)}{(z_1 - z_2)(z_3 - z_4)}\,,$$

and the points z_1^*, z_2^* are obtained as intersection points of the line through z_1, z_2 with the unit circle as illustrated below; we choose z_i^* to be the intersection point closer to z_i, $i = 1, 2$; if $z_1 = z_2$, the points z_1^*, z_2^* are not well-defined but we simply take $d(z_1, z_2) = 0$.

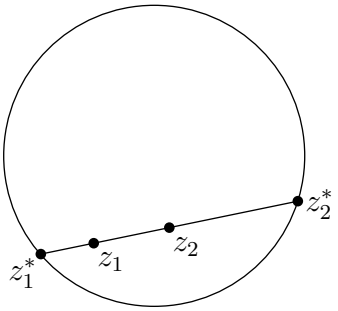

FIGURE 1. The points z_1^*, z_2^*

An explicit analytic formula for $d(z_1, z_2)$ is given by

$$d(z_1, z_2) = \cosh^{-1}\left(\frac{1 - \mathrm{re}(z_1\overline{z_2})}{\sqrt{1 - |z_1|^2}\sqrt{1 - |z_2|^2}}\right).$$

There is a simple isometry f from the Poincaré disk D to the Klein disk K given by $f(z) = 2z/(1 + |z|^2)$. Geometrically this corresponds to the composition of two projections: we place the unit disk inside a unit sphere so that the unit circle is the equator of the sphere, then stereographically project D from the South pole onto the Northern hemisphere, and then do a Euclidean orthogonal projection of the Northern hemisphere back to D.

The Klein model is simpler to use in some situations than the Poincaré models because the geodesics are all line segments rather than circular arcs. However, it has one significant drawback compared with those other models: the notion of hyperbolic angle in this model does not coincide with the Euclidean angle in this model since, unlike the Poincaré models, the Klein metric is not a conformal distortion of the underlying Euclidean metric.

2.3. Gaussian and sectional curvature: a quick guide. Recall that the curvature of an arc at a point is the reciprocal of the radius of the osculating circle. Trivially we can distort a line segment so as to give it nonzero curvature while leaving distance (as measured by arclength) unchanged. If we view the arc as a metric space, then the curvature is a property of the particular imbedding of that arc in Euclidean space, not an intrinsic property of the metric space.

Gauss published his *Theorema egregium* ("Remarkable theorem") in 1828. Here he examined surfaces and defined the principal curvatures to be the maximum and minimum values k_1, k_2 of the signed curvatures at p of all smooth geodesics that pass through p (the sign indicates whether the associated arc bends in the chosen normal direction or not). He then defined what we now call the Gaussian curvature K to be $k_1 k_2$. As in one dimension, the principal curvatures are not intrinsic but Gauss discovered that K is intrinsic, i.e. it is a local isometry invariant. This is why a flat sheet of paper which droops if we hold it only on one side, does not droop if we bend it into a cylinder: the flat paper has Gaussian curvature $K = 0$, so if we bend it like this we are introducing a non-zero k_1 forcing k_2 to be zero in order to preserve K. For the same reason, we naturally bend the sides of a segment of pizza to stop the free end from drooping.

Let X be an open subset of the plane and let $ds = a(z)|dz|$ be a conformal distortion of the Euclidean metric on X by a C^2 function a, i.e. the associated length of a path γ in X is given by $\int_\gamma a(z)|dz|$ and the associated distance $d(z, w)$ is obtained by minimizing this length over all paths from z to w. Then the curvature $K(z)$ with respect to this metric is given by

$$K(z) = -\frac{4\partial\overline{\partial}\log(a(z))}{(a(z))^2} = -\frac{\triangle \log(a(z))}{(a(z))^2}\,. \tag{2.1}$$

EXERCISE 2.2. Use the formula for $K(z)$ to verify that the Poincaré metrics ρ_H and ρ_D in §2.1 satisfy $K(\cdot) \equiv -1$.

Note that it follows from (2.1) that if we dilate the metric by a factor c, then the curvature is multiplied by a factor c^{-2}. This makes it easy to give so-called model surfaces of any desired constant Gaussian curvature. For instance, from the definition of K, it is obvious that $K = 1$ for the 2-sphere of radius 1, so we get a surface M_k of any desired positive curvature k by taking a 2-sphere of radius $k^{-1/2}$. Similarly, we get a surface M_k of any desired negative curvature k by dilating H^2 by a factor $(-k)^{-1/2}$. Finally the Euclidean plane is a surface M_0 of constant zero curvature. These spaces M_k are the *model spaces* that are used to define CAT(k) spaces in Section 4.

The curvature of a higher dimensional Riemannian manifold X (which we implicitly assume to be smooth) is a more complicated beast. Intuitively, a small neighborhood of a point $x \in X$ is almost isometric to a small piece of Euclidean space. *Sectional curvature* of X at x consists roughly of the collection of Gaussian curvatures of all "planar slices" near the point x. Saying that the sectional curvature of X equals, or is

at most, K, means precisely that all of these Gaussian curvatures equal, or are at most, K. There are many good sources for the theory of curvature on manifolds, for instance the book by Chavel [**24**].

2.4. Geodesics in the hyperbolic plane. A path $\gamma : I \to X$ in a metric space is a *geodesic path* and $\gamma(I)$ a *geodesic* if some reparametrization of it is an isometry. We call $\gamma(I)$ a *geodesic segment*, *geodesic ray*, or *geodesic line* if I is of the form $[a, b]$, $[a, \infty)$, or $(-\infty, \infty)$, respectively, for some $a, b \in \Re$, $a < b$. Using the term "geodesic" to describe both paths and their images seems harmless, since we will always indicate explicitly that we are talking about a path if this is so (using terms such as "geodesic path" or "unit speed geodesic").

Metric spaces may or may not contain geodesics, but the hyperbolic plane contains a unique geodesic segment between every pair of points. Let us discuss the form of these geodesics for each of our three models of the hyperbolic plane. In each case the form is given in terms of simple Euclidean geometric concepts.

In the Poincaré upper half-plane H, the geodesic lines are precisely the intersections with H of either vertical open lines or circles with centers on the real axis. In the Poincaré disk D, the geodesic lines are precisely the intersections with D of circles that cut the unit circle orthogonally. In the Klein disk K, the geodesic lines are precisely the intersections with K of lines.

2.5. The ideal boundary of the hyperbolic plane. The ideal boundary of a metric space is a type of boundary at infinity which is a very useful concept when dealing with nonpositively curved spaces. Indeed it is useful even in the setting of the hyperbolic plane. We will properly investigate it in later sections. In this section we give an intuitive but somewhat vague introduction to this concept which will suffice for now.

In the two disk models D and K of the hyperbolic plane, there is an underlying Euclidean domain (a disk) and if we put on our Euclidean spectacles, we see that all geodesic lines end at two boundary points of this Euclidean domain. In the upper half-plane model, the underlying Euclidean structure is noncompact, so we instead use its one-point compactification $\overline{H} = H \cup \{\infty\}$. Then we can consider all geodesic lines as having two endpoints in the boundary of $\overline{H}$; the vertical lines are the geodesics that have ∞ as an endpoint. Since geodesic lines have infinite length, we view their Euclidean-type endpoints as "points at infinity" in hyperbolic space. We define the *ideal boundary* of the space to be the collection of all such points at infinity. We denote the ideal boundary of a space X by $\partial_I X$. More explicitly, $\partial_I D$ and $\partial_I K$ are both the unit circle and $\partial_I H$ consists of the one-point compactification of the real line (and so also essentially a circle).

Our definitions of the ideal boundaries of these three models are not intrinsic since they use the underlying Euclidean structure of our spaces. This is just for simplicity

at this stage. In §4.3, we will give an intrinsic definition of the ideal boundary of a nonpositively curved space that agrees with the above definitions and that carries an associated topology which is an isometry invariant and is consistent with the obvious topologies of the ideal boundaries of our three models of $\mathbb{H}^2$. Thus what we have really defined above is the ideal boundary $\partial_I H^2$ of the hyperbolic plane.

The following useful facts, the first of which is a stronger version of Euclid's first postulate in the context of the hyperbolic plane, are rather obvious using our explicit description of geodesics in $\mathbb{H}^2$ (especially in the Klein model).

FACT 2.3. *Between every pair of points $a, b \in H^2 \cup \partial_I H^2$, there is a unique geodesic segment.*

FACT 2.4. *Every geodesic segment in the hyperbolic plane is contained in a unique geodesic line.*

The ideal boundary in the Poincaré models allows us to give an alternative definition of the distance function in those metric spaces that is very similar to the first definition of the distance function in K. Recall that

$$d(z_1, z_2) = \frac{1}{2} \log[z_1^*, z_1, z_2, z_2^*], \qquad z_1, z_2 \in K,$$

where the *cross-ratio* $[\cdot, \cdot, \cdot, \cdot]$ is defined by the formula

$$[z_1, z_2, z_3, z_4] = \frac{(z_1 - z_3)(z_2 - z_4)}{(z_1 - z_2)(z_3 - z_4)},$$

and we can now define the points z_1^*, z_2^* to be the ideal boundary endpoints of the geodesic line L through z_1 and z_2, chosen so that the order of the points induced by L is z_1^*, z_1, z_2, z_2^*.

In a similar fashion, the distance function in the Poincaré disk or upper half-plane is given by

$$\rho(z_1, z_2) = \log[z_1^*, z_1, z_2, z_2^*], \tag{2.5}$$

where z_1^*, z_2^* are the ideal boundary endpoints of the geodesic line L through z_1 and z_2, so that the order of the points induced by L is z_1^*, z_1, z_2, z_2^*. In the half-plane model, we cancel factors involving ∞ in the usual way.

The similarity of these formulae is not a coincidence. Cross-ratio is preserved by Möbius maps, and D and H are isometric via a Möbius map f that respects the ideal boundary (in the sense that if a is an ideal boundary endpoint of a geodesic line L, then $f(a)$ is an ideal boundary endpoint of $f(L)$). From these facts, it is clear that the same cross-ratio formula for either of these models is transported by the identification f to the other model. As for the similarity to the Klein model, recall that an isometry from D to K is obtained by embedding D in $\mathbb{R}^3$, mapping D to the Northern hemisphere S via a stereographic projection p_S from the South pole, and then mapping S onto K via an orthogonal projection p_K.

EXERCISE 2.6. Inversions (and compositions of inversions) take the place of Möbius maps in $\mathbb{R}^n$. The inversion $I_{a,r} : \widehat{\mathbb{R}^n} \to \widehat{\mathbb{R}^n}$ is a mapping on the Riemann n-sphere $\widehat{\mathbb{R}^n}$. Geometrically we invert points through the Euclidean sphere of radius r and center a. Thus $I_{a,r}(x) = a + r^2(x-a)/|x-a|^2$, $x \neq a, \infty$, with $I_{a,r}(a) = \infty$ and $I_{a,r}(\infty) = a$.

(a) Show that inversion maps spheres in $\widehat{\mathbb{R}^n}$ to spheres in $\widehat{\mathbb{R}^n}$ (and so Euclidean spheres and hyperplanes to Euclidean spheres and hyperplanes if we ignore the points a and ∞).
(b) Show that inversion preserves cross-ratio.
(c) Show that the stereographic projection p_S is an inversion via a sphere of radius $\sqrt{2}$ centered at the South pole.

It follows from the above exercise that the cross-ratio formula is preserved by the map p_S. It is with the map p_K that the factor $1/2$ is introduced to the formula for distance in the Klein model, according to the following exercise.

EXERCISE 2.7. Let $a, b \in S$ be distinct points on a semicircular arc with endpoints u, v. Prove that $[u, a, b, v] = \sqrt{[u, p_K(a), p_K(b), v]}^{1/2}$.

2.6. Asymptotic and divergent geodesics. The Poincaré half-plane (H, d) looks a lot like the Euclidean plane and satisfies Euclid's first four postulates if we use suitable definitions of the concepts involved: a "point" is an element of H, a "straight line" is the intersection with H of either a vertical line or a circle centered on the real axis, a "circle" is the set of points of a constant hyperbolic distance from a center point, and the angle between "straight lines" is the Euclidean angle between the geodesics in either of the Poicaré models.

However the Parallel Postulate fails. Since it is essentially equivalent to show the failure of Playfair's Axiom (stated in the Introduction), we do this instead. The argument is most easily seen in the Klein model so we use that.

Suppose we are given a geodesic line L and a point $a \in K \setminus L$. Let $b, c \in \partial_I K$ be the (ideal boundary) endpoints of L. The set $K \setminus L$ consists of two components and a lies in one of them, call it K_1. We define $\partial_I K_1$ in the obvious way: it consists of the largest open arc on the unit circle that is in the Euclidean boundary of K_1 (see the diagram). Any geodesic line through a that ends at two points in $\partial_I K_1 \cup \{b, c\}$ is disjoint from L. It is easy to deduce that *there are infinitely many different geodesic lines that pass through a and do not intersect L* (and so are said to be parallel to L). Three such geodesic lines, M_1, M_2, M_3, are indicated in Figure 2.

Of the infinite number of geodesic lines parallel to L, two are special because they share an ideal boundary endpoint with L (M_1 and M_2 in the diagram). These geodesic lines are said to be *asymptotic* to L, while all the other parallel geodesics (such as M_3 in the diagram) are said to be *divergent* from L.

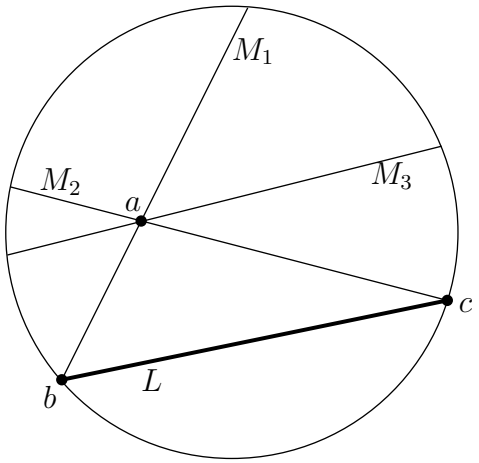

FIGURE 2. Geodesics parallel to L: Klein model

The same distinction between asymptotic and divergent geodesic lines is made in the other models of the hyperbolic plane. In Figures 3 and 4, we look at upper half-plane model for the cases where L is either a vertical half-line or a half-circle. In both diagrams, M_1 and M_2 are the two geodesics asymptotic to L.

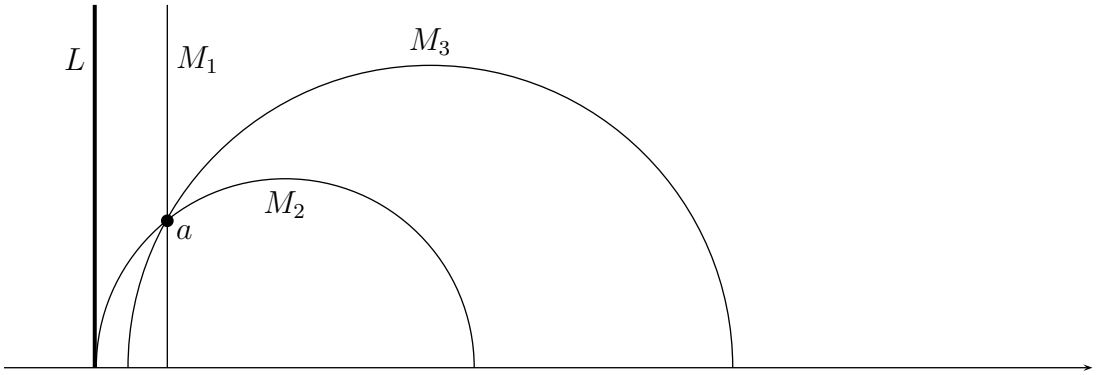

FIGURE 3. Geodesics parallel to L: vertical case

EXERCISE 2.8. Prove that the set of hyperbolic circles in H and the set of Euclidean circles in H coincide. Given that C is a Euclidean circle with center (x_0, y_0) and radius r, $0 < r < y_0$, find the hyperbolic center and radius of C.

Asymptotic and divergent geodesic lines are so-called with good reason. Let us discuss this in the context of a complete discussion of the behavior "near infinity" of nonintersecting unit speed geodesic rays $\gamma : [0, \infty) \to X$ and $\lambda : [0, \infty) \to X$, where X is either the Euclidean or hyperbolic plane. In the Euclidean case, it follows that there

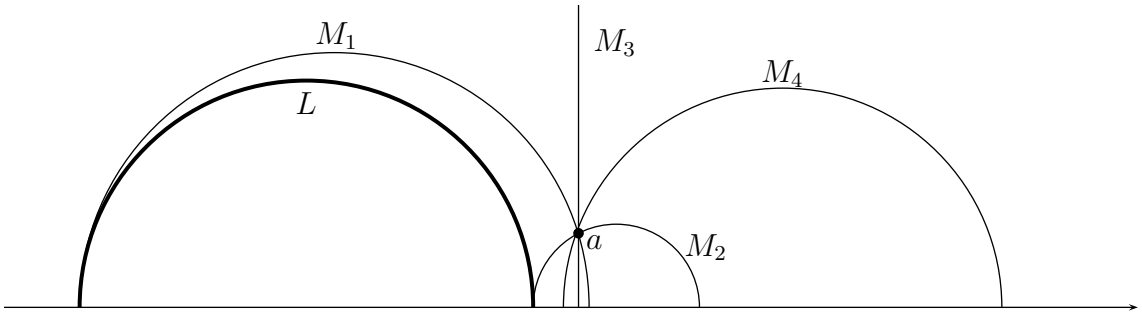

FIGURE 4. Geodesics parallel to L: non-vertical case

are constants $0 \leq a \leq 2$ and $C > 0$ such that

$$| \, |\gamma(t) - \lambda(t)| - at \, | \leq C \, , \qquad t \geq 0 \, . \tag{2.9}$$

All values of a between 0 and 2 are possible in (2.9) by choosing the correct angle between the directions of γ and λ. We get $a = 0$ only if the paths have the same direction and we can then replace (2.9) by the stronger statement that $|\gamma(t) - \lambda(t)|$ is constant. We get $a = 2$ only when the paths have opposite directions.

This continuum of rates of divergence is not found in hyperbolic space. In fact there is a striking dichotomy: either rays are exponentially asymptotic or they eventually move apart about as fast as allowed by the triangle inequality (i.e. they satisfy the hyperbolic analogue of (2.9) with $a = 2$).

Let us make these statements more precise beginning with the asymptotic case. This is the case where both γ and λ have the same endpoint on the ideal boundary. We look at the Poincaré upper half-plane, as this is easiest to analyze. By means of a suitable Möbius map, it suffices to assume that the rays γ and λ are vertical half-lines with ∞ as their ideal boundary endpoint. The fact that $\rho_H(\gamma(t), \lambda(t + t_0))$ tends to zero for some choice of t_0 follows from the fact that for any fixed $u \neq v \in \mathbb{R}$, the distance $\rho_H(u + si, v + si)$ tends to 0 as $s \to \infty$, which in turn follows from the fact that the Euclidean line segment from $u + si$ to $v + si$ has hyperbolic length at most $|u - v|/s$.

EXERCISE 2.10. Fill in the gaps in the above argument. Use it to prove that if $\gamma : [0, \infty) \to H$ and $\lambda : [0, \infty) \to H$ are a pair of nonintersecting unit speed geodesic rays with the same ideal boundary endpoint, then there exist constants $t_0 \in \mathbb{R}$ and $C > 0$ such that

$$\rho_H(\gamma(t), \lambda(t + t_0)) \leq C \exp(-t) \, , \qquad t \geq \max(0, -t_0) \, .$$

We now look at the divergent case. Again we look at the Poincaré upper half-plane model. By means of a suitable Möbius map, we may assume that γ and λ are vertical line segments with real ideal boundary endpoints 0 and $a > 0$, respectively. The key

to proving the desired result is to examine $\rho_H(\epsilon i, a + \epsilon i)$ as $\epsilon \to 0$. According to our formula for ρ_H, this equals $2 \tanh^{-1} |a/(a + 2\epsilon i)|$. Routine estimation shows that this differs from $-2 \log \epsilon$ by at most a constant independent of ϵ.

EXERCISE 2.11. Fill in the gaps in the above argument. Use it to prove that if $\gamma : [0, \infty) \to H$ and $\lambda : [0, \infty) \to H$ are a pair of nonintersecting unit speed geodesic rays with different ideal boundary endpoints, then there exists a constant $C > 0$ such that

$$| \rho_H(\gamma(t), \lambda(t)) - 2t | \le C , \qquad t \ge 0 . \tag{2.12}$$

2.7. Hyperbolic trigonometry. As mentioned in the introduction, a lot of the theory of Euclidean geometry carries over to hyperbolic geometry. For instance, for both the Euclidean plane and the hyperbolic plane, the isometry group G of the space is generated by reflections in geodesic lines (i.e. order 2 elements of G), and the stabilizer of a point is the orthogonal group $O(2)$. For more on this, see [**3**], Chapter 7 of [**9**], and Sections I.2 and I.6 of [**14**].

The trigonometry of hyperbolic geometry is reminiscent of the Euclidean case, but nevertheless some important differences arise. We define hyperbolic triangles in the obvious way: they consist of a set of three points A, B, C together with the geodesic segments between them.

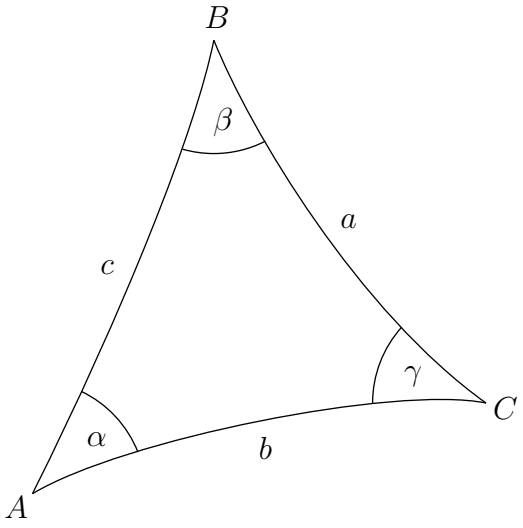

FIGURE 5. A hyperbolic triangle

Suppose we consider a hyperbolic triangle with vertices A, B, C, sidelengths a, b, c, and angles α, β, γ, as pictured in Figure 5. The sidelengths and angles are related by sine and cosine rules reminiscent of those in Euclidean geometry:

Hyperbolic sine rule:

$$\frac{\sin\alpha}{\sinh a} = \frac{\sin\beta}{\sinh b} = \frac{\sin\gamma}{\sinh c}\,.$$

First hyperbolic cosine rule:

$$\cosh a = \cosh b \cosh c - \sinh b \sinh c \cos\alpha\,.$$

However, unlike the Euclidean case, there is greater qualitative symmetry between sidelength data and angular data in the form of a second dual form of the cosine rule.

Second hyperbolic cosine rule:

$$\cos\alpha = -\cos\beta\cos\gamma + \sin\beta\sin\gamma\cosh a\,.$$

If (Euclidean or hyperbolic) triangles T_1 and T_2 have the same sidelengths, there is a natural map $f : T_1 \to T_2$ defined by the requirement that the restriction of f to any one side of T_1 is an isometry. Using the (Euclidean or first hyperbolic) cosine rule twice, we first see that the three sidelengths determine the three angles, and then that any such natural map f is an isometry, i.e. *three sidelengths determine a (Euclidean or hyperbolic) triangle up to isometry.* Using the sine rule and first cosine rule as in the Euclidean case, we similarly see that *a hyperbolic triangle is determined up to isometry by two sidelengths and the angle between them.*

A hyperbolic triangle is also determined by one sidelength and two angles: this is a little harder to show than in the Euclidian case since we do not automatically know the third angle, so let us say a little more. If we know two angles and the side between them, e.g. β, γ, and a, then the second cosine rule gives α, and then the sine rule gives b, c. The other case to be considered involves knowing two angles and an opposite side, e.g. β, γ, b. The sine rule gives c, and by combining the two cosine rules we get a formula for a in terms of b, c, β, and γ.

Lastly, unlike the Euclidean case, the second hyperbolic cosine rule shows that *a hyperbolic triangle is determined up to isometry by three angles.* Thus, with one exception, any three of the six pieces of data $a, b, c, \alpha, \beta, \gamma$ determine a hyperbolic triangle up to isometry. The exception is the same as the "two sides plus one angle" exception in the Euclidean case: if b, c and an angle other than α are given, there is in general two possible values for a. Geometrically, this is because if we have one hyperbolic (or Euclidean) triangle with sidelengths b, c and angle β, then we can get another by reflecting the segment AC in the perpendicular bisector of BC. In terms of the hyperbolic (or Euclidean) sine rule, note that if b, c, and β are given, then we can uniquely solve for $\sin\gamma$, but not normally for γ, because sin takes on all values in $(0, 1)$ twice in $(0, \pi)$.

The fact that a hyperbolic triangle is determined up to isometry by its three angles is tied to the fact that there are no dilations in the hyperbolic plane. More precisely if a map $f : H^2 \to H^2$ takes hyperbolic lines to hyperbolic lines and preserves angles, then it must be an isometry (and so a Möbius map if we are using the Poincaré disk or upper half-plane model of H^2).

2.8. Hyperbolic area of triangles and disks. The absence of dilations means that the area of a triangle or of a disk does not scale up as in the Euclidean case as we scale up the sidelengths or radius. In fact under such rescalings, the area of a triangle increases more slowly and the area of a disk increases quicker than in the Euclidean setting. Let us now say more about both of these.

For triangles, it can be shown that the angles all decrease if we multiply the sidelengths by a factor larger than 1. Moreover it follows from the first cosine rule and the fact that $\lim_{t\to\infty}(\cosh t - \sinh t) = 0$ that all the angles tend to 0 as the sidelengths tend to infinity.

There is a simple and remarkable relationship between angles and area.
Gauss-Bonnet formula: The hyperbolic area of a triangle with interior angles α, β, γ is $\pi - (\alpha + \beta + \gamma)$. This holds even if one or more vertices of the triangle are on the ideal boundary (in which case the associated angles are zero).

The Gauss-Bonnet formula is not hard to prove using the upper-half space model H. First note that a triangle with three vertices in H can be written as a set difference of a triangle with two vertices in H and one on the boundary. By using a Möbius map (which as an isometry, preserves area), we may assume that the ideal vertex is ∞, and then it becomes a rather straightforward computation.

Exercise 2.13. Prove the Gauss-Bonnet formula.

It follows from the Gauss-Bonnet formula that if we rescale upwards the sidelengths of a hyperbolic triangle, its area increases, with a limiting area of π as the sidelengths tend to infinity. It can be shown that the rate of increase of area is always slower than in the Euclidean setting, e.g. doubling the sidelength increases the area by a factor less than 4.

We now turn to disks. The area A_r of a hyperbolic disk of radius r is independent of the center (as is obvious from the transitivity of the isometry group), and is given by $4\pi \sinh^2(r/2)$. The length L_r of the hyperbolic circle of radius r is $2\pi \sinh r$. Both of these can be proven most easily by using the Poincaré disk model and using a Möbius map to assume that the center of the disk is at the origin.

Exercise 2.14. Derive the formulae for A_r and L_r.

By calculus, it follows that both A_r and L_r are very similar to the corresponding Euclidean quantities when r is small. However they increase far faster than in the Euclidean setting when r is large. In fact, for large r, a unit increase in r increases both the area and the circumference by about a factor e.

2.9. n-dimensional hyperbolic space. We will not say much about n-dimensional hyperbolic geometry, since it bears more or less the same relationship to planar hyperbolic geometry as does n-dimensional Euclidean geometry to planar geometry. Higher

dimensional analogues of all three of our earlier models exist for $\mathbb{H}^n$. More explicitly, the Poincaré upper half-space $\{x \in \mathbb{R}^n \mid x_n > 0\}$ has Riemannian metric

$$ds^2 = \frac{ds_E^2}{x_n^2},$$

and the Poincaré ball $\{x \in \mathbb{R}^n : |x| < 1\}$ has Riemannian metric

$$ds^2 = \frac{4ds_E^2}{(1-|x|^2)^2}.$$

In both cases, ds_E denotes the infinitesimal Euclidean metric on the underlying domain. Analogous formulae for the distance function can also be written down (of course we must first rewrite the planar formulae using inner products rather than complex arithmetic). In both cases, the distance function is also given by the same cross-ratio formula (2.5) as before. Note that, as in the planar case, the geodesic lines are circular arcs and half-lines orthogonal to the Euclidean boundary.

The Klein model is such an obvious generalization of the planar case that we will say no more about it.

Let us mention just one basic fact about $\mathbb{H}^n$, namely that lower dimensional hyperbolic spaces are embedded in $\mathbb{H}^n$ just as lower dimensional Euclidean spaces are embedded in $\mathbb{R}^n$. In fact, any set of $m+1$ points in $\mathbb{H}^n$, $1 \leq m \leq n$, lie in an isometric copy of $\mathbb{H}^m$. This is most easily seen by using the Klein model. It follows that if we wish to prove something about hyperbolic triangles in $\mathbb{H}^n$, we may as well assume that $n = 2$.

3. Other metrics in complex analysis and potential theory

3.1. Poincaré, Carathéodory, and Kobayashi metrics. The fundamental *Uniformization Theorem* tells us that every Riemann surface X has as its universal cover one of three simple surfaces: the Riemann sphere, the complex plane, or the unit disk. Moreover, the examples of the first two types are very few, so that "most" Riemann surfaces (including all of genus larger than 1, such as open subsets of the plane with at least two boundary points) have the unit disk D as their universal cover and are termed *hyperbolic* since D can be equipped with the Poincaré metric ρ_D making it a model of the hyperbolic plane. Using the local identification of D and X provided by the covering map, we can transport the infinitesimal Poincaré metric from D to X. By integrating this density, we define a metric on X which is also called the *Poincaré metric*. Since Gaussian curvature is a local isometric invariant, this gives a Riemannian metric of constant Gaussian curvature -1 on X.

The Poincaré metric is our first example of a biholomorphically invariant metric: if $f : X \to Y$ is a biholomorphic map between hyperbolic Riemann surfaces, then $\rho_Y(f(z), f(w)) = \rho_X(z, w)$, $z, w \in X$. This fact follows from the more general result

that the Poincaré metric is distance decreasing with respect to holomorphic maps, which in turn follows from the Schwarz-Pick theorem stated in §2.1.

There are other such invariant metrics, such as the *Carathéodory pseudometric* c_G on a domain $G \subset \mathbb{C}^n$. First let $H(G, D)$ be the class of holomorphic maps from G to the unit disk D and let

$$c_G(z, w) = \sup_{f \in H(G,D)} \rho_D(f(z), f(w)) \qquad z, w \in G\,. \tag{3.1}$$

It is easily seen that $c_D = \rho_D$. Indeed the fact that $c_D \le \rho_D$ follows from the distance decreasing property of the Poincaré metric, and we get equality by picking f to be the identity map. The distance decreasing property of c_G follows immediately from its definition.

The Carathéodory pseudometric is a metric if and only if the space of bounded holomorphic functions, $H^\infty(G)$, separates points in G. For instance if G is biholomorphically equivalent to a bounded domain, then c_G is a metric. Assuming c_G is a metric, the *inner Carathéodory metric* c_G^i is the inner metric on G associated with c_G, as defined in Section 6.

The *Kobayashi pseudometric* on a domain $G \subset \mathbb{C}^n$ is similar to the Carathéodory pseudometric, but defined in terms of mappings from D to G rather than the other way around. For arbitrary $z, w \in G$, we write

$$\begin{aligned} \tilde{k}_G(z, w) &= \inf\{\rho_D(u, v) \mid u, v \in D,\, \exists\, f \in H(D, G) :\ f(u) = z,\ f(v) = w\}\,. \\ k_G(z, w) &= \inf\left\{ \sum_{j=1}^{n} \tilde{k}_G(z_{j-1}, z_j) \right\}\,. \end{aligned} \tag{3.2}$$

Note that in the definition of k_G, we take an infimum over all choices of points $z_0 = z, z_1 \ldots, z_n = w$.

It is straightforward to show that $c_G \le k_G$. Thus k_G is a metric if c_G is a metric. In this case, the fact that k_G is a length metric implies that we also have $c_G^i \le k_G$.

The Kobayashi and (inner) Carathéodory pseudometrics can formally be defined in the same manner on any set G with a complex structure, such as Riemann surfaces and normed spaces.

One disadvantage of c_G^i and k_G compared with ρ_G (when they can all be defined) is that they are not Riemannian metrics. They are however Finsler metrics, meaning that at the infinitesimal level they are given by norms in the same way as a Riemannian metric is given infinitesimally by an inner product. For more on the Kobayashi and (inner) Carathéodory pseudometrics, see the book by Jarnicki and Pflug [**39**]. For more on Finsler geometry, see the books by Bao, Chern, and Shen [**8**], and by Shen [**45**].

3.2. The Hilbert metric in a convex Euclidean domain. Busemann said in [**23**]:

> Plane Minkowskian geometry arises from the Euclidean through replacing the ellipse as unit circle by a convex curve. In a somewhat similar way a geometry discovered by Hilbert arises from Klein's Model of hyperbolic geometry through replacing the ellipse as absolute locus by a convex curve.

Let $G \subset \mathbb{R}^n$ be a bounded convex domain. Then the Hilbert metric on G is defined by

$$h_G(x_1, x_2) = \frac{1}{2} \log[x_1^*, x_1, x_2, x_2^*], \qquad x_1, x_2 \in G\,,$$

with $h_G(x_1, x_2) = 0$ in the special case $x_1 = x_2$. Above, x_i^*, $i = 1, 2$ are the points on the intersection of ∂G and the line through x_1, x_2, with x_1^* being the one that is closer to x_1.

This is a straightforward generalization of the Klein model. Busemann talks about it being a generalization from the case of the ellipse rather than the circle because the cross-ratio of four points on a line is a projective invariant,[2] and so all ellipses give isomorphic Hilbert geometries.

Although these general Hilbert metrics are not related to complex analysis, we feel they are worthy of mention in these notes because they produce an interesting variety of geometries with very simple geodesics. At one extreme, if $G \subset \mathbb{R}^n$ is a sphere (or ellipsoid), then (G, h_G) is isomorphic to H^n as mentioned before. This is the only case where we get a Riemannian metric: in all other cases, the Hilbert metric is merely a Finsler metric, as shown by Socié-Méthou [**46**, 1.3.5].

At the other extreme, de la Harpe [**35**] showed that if $G \subset \mathbb{R}^n$ is a simplex, then (G, d_G) is isometric to $\mathbb{R}^n$ with a polyhedral norm attached (i.e. the unit ball is a polyhedron); in particular when $n = 2$, the resulting space is isometric to the normed plane with a hexagonal unit ball. Moreover, simplices are the only domains for which the Hilbert metric is a normed space, as shown by Foertsch and Karlsson [**29**].

One last point we wish to make about Hilbert geometries is that the Euclidean line segment between pairs of points in G is always a d_G-geodesic, although it may not be unique. The following is a simple criterion for the uniqueness of geodesics [**35**]:

THEOREM 3.3. *Let (G, d_G) be a Hilbert geometry. Then there is a unique d_G-geodesic between every pair of points in G if and only if the following is true: for each $x \in G$ and each plane $\Pi \ni x$, the intersection $\Pi \cap \partial G$ contains at most one nontrivial line segment.*

Figure 6 shows what goes wrong if there are two such line segments. Here z' is the intersection of a line through x', x, y', and a line through x'', y, y'', where x' and

[2]The invariance of cross-ratio of points in a line under projective transforms is central to the study of such transforms, both in the mathematical theory and in their applications to computer visual recognition (they are used to recognize an object that may appear at an angle and distance different from the stored image). Note though that the cross-ratio of four points in general position in $\mathbb{R}^n$, $n > 1$, is not a projective invariant, although it is a Möbius invariant.

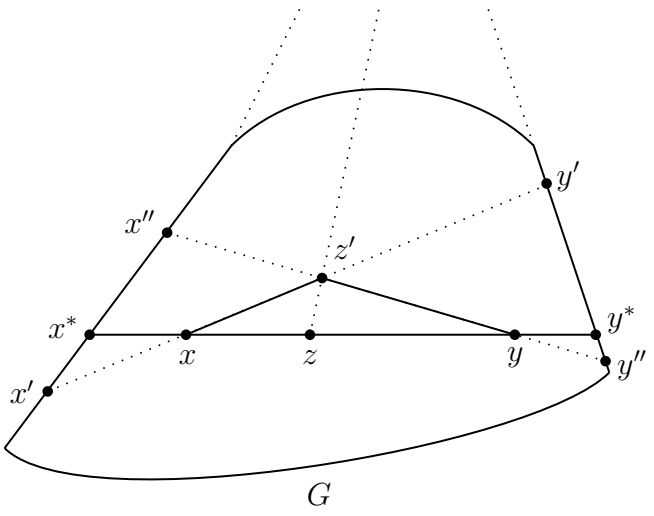

FIGURE 6. Non-unique geodesics

x'' are on opposite sides of the boundary line segment containing x^*, and y' and y'' are on opposite sides of the boundary line segment containing y^*. These line segments intersect at some point (which happens to be off the diagram) and using this point as the center of our perspective, we project z' to some point z on the line through x and y. By the projective invariance of cross-ratio, it follows that $d_G(x,z) = d_G(x,z')$ and that $d_G(z,y) = d_G(z',y)$, and so the polygonal line from x to z' to y is also geodesic.

3.3. The quasihyperbolic metric and related metrics. The *quasihyperbolic metric* in an incomplete rectifiably connected metric space (X,d) is the metric $k = k_X$ given infinitesimally by the conformal distortion

$$\frac{dx}{\operatorname{dist}(x,\partial X)},$$

where ∂X consists of all points in the metric completion of X that are not in X. In the more concrete setting of a Euclidean domain $G \subsetneq \mathbb{R}^n$, this is a Riemannian metric ds_k given infinitesimally by

$$ds_k^2 = \frac{ds^2}{(\operatorname{dist}(x,\partial G))^2},$$

where ds is the infinitesimal Euclidean metric and ∂G is the Euclidean boundary.

The quasihyperbolic metric k is used extensively in geometric analysis and potential theory; see the survey by Koskela [**41**]. As is well known, k is comparable with the Poincaré metric ρ on a simply connected domain $G \subsetneq \mathbb{C}$. Compared with ρ, it is defined more directly from the geometry of the domain but it has the disadvantage of not being Möbius invariant. There is however a metric which is both Möbius invariant

and bilipschitz equivalent to k and can be defined on any domain $G \subsetneq \mathbb{R}^n$ that has at least two boundary points: the *Ferrand metric* [**27**] is defined infinitesimally by

$$\sigma(z) = \sup_{u,v \in \partial G} \frac{|u-v|}{|u-z|\,|v-z|}, \qquad z \in G\,.$$

The quasihyperbolic metric k is often hard to evaluate but an important lower bound on the distance $k(x,y)$ involves either of two very similar metrics that are often called the j- and $\tilde{j}$-metrics. The $\tilde{j}$-metric, introduced by F. Gehring, is defined on an incomplete metric space (X,d) by

$$\tilde{j}(x,y) = \frac{1}{2}\log\left[\left(1+\frac{d(x,y)}{\operatorname{dist}_d(x,\partial_d D)}\right)\left(1+\frac{d(x,y)}{\operatorname{dist}_d(y,\partial_d D)}\right)\right], \qquad x,y \in X\,,$$

while Vuorinen's j-metric is

$$j(x,y) = \log\left(1+\frac{d(x,y)}{\min[\operatorname{dist}_d(x,\partial_d D), \operatorname{dist}_d(y,\partial_d D)]}\right), \qquad x,y \in X\,.$$

In the following exercises, (X,d) is an incomplete rectifiably connected metric space.

Exercise 3.4. Show that

$$\frac{j(x,y)}{2} \le \tilde{j}(x,y) \le j(x,y), \qquad x,y \in X\,.$$

Exercise 3.5. Show that $j(x,y) \le k(x,y)$, $x,y \in X$.

Exercise 3.6. Show that k is the inner metric associated with either j or $\tilde{j}$ (as defined in Section 6).

Despite the similarity of j and $\tilde{j}$, and their relationship to k, we will see that the question of Gromov hyperbolicity has remarkably different answers for k, j, and $\tilde{j}$; see §5.6.

4. CAT(k) and related curvature conditions

In 1957, Alexandrov introduced several equivalent definitions of what it means for a metric space to have curvature bounded above by k, for any real number k. All involve comparing the space to a well-understood model space. These definitions are nowadays called the CAT(k) condition, a term introduced by Gromov [**32**] in honor of Cartan, Alexandrov, and Toponogov.

These conditions are of great importance for a variety of reasons. They have played an important role in various areas of mathematics, for instance harmonic maps [**34**] and Lipschitz extensions [**43**]. In the context of Riemannian manifolds, the local variant of CAT(k) coincides with the assumption that the sectional curvature is at most k (but it is much simpler to understand than the curvature tensor). CAT(k) itself is a stronger global condition that additionally implies the manifold is simply-connected.

However, unlike sectional curvature, CAT(k) makes sense in any *geodesic metric space* (a metric space where every pair of points can be connected by a geodesic segment). There are many results on metric spaces that involve such a curvature condition as a hypothesis. The fact that they are closed under some important limiting processes, specifically Gromov-Hausdorff limits and ultralimits, adds to their importance.

Here we give a survey of CAT(k) spaces for $k \leq 0$. The case $k > 0$, which we omit, is broadly similar, although there are some differences due to the fact that positively curved spaces, such as spheres, tend to be of finite diameter. We refer the reader to the book by Bridson and Haefliger [**14**] for much more on the theory of CAT(k) spaces for all $k \in \mathbb{R}$. Since much of what is below can be found in [**14**], we mainly give references only to results that are to be found elsewhere.

4.1. CAT(k): introduction and examples. Below, (X, d) is a geodesic metric space. The idea of CAT(k) is simple: intuitively a space X with curvature at most k should have geodesics that move apart at least as fast as the corresponding ones in a simple model space M of constant curvature k. Let us make this statement more precise.

First, we use the spaces (M_k, d_k) introduced in §2.3 as our model spaces. In other words, M_0 is a Euclidean plane and, for all $k < 0$, M_k is the dilation of $\mathbb{H}^2$ by a factor $1/\sqrt{-k}$.

To discuss the rate at which geodesics move apart, we need the notion of a geodesic triangle in X. First, it is convenient to denote a geodesic segment with endpoints $x, y \in X$ as $[x, y]$. This notation is not meant to imply that geodesic segments are unique, but simply refers to a choice of one such geodesic segment. A *geodesic triangle* T with vertices $x, y, z \in X$ is simply the union of three such geodesics $[x, y]$, $[y, z]$, and $[z, x]$.

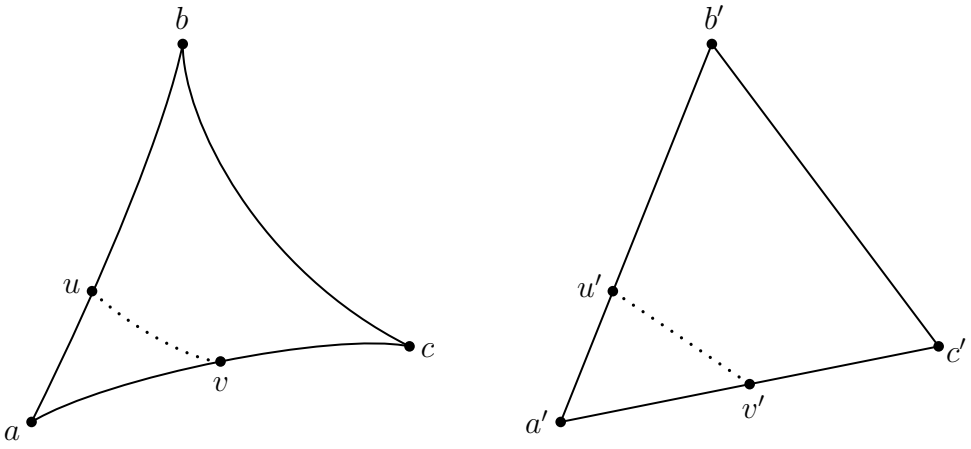

FIGURE 7. A d-triangle and a comparison triangle

We pick a *comparison triangle* T' with vertices a', b', c' in M_k, so that the $d(a,b) = d_k(a',b')$, and similarly for the other two sides. Such a triangle always exists when $k \leq 0$. There is a natural map $f : T \to T'$ with $f(a) = a'$, $f(b) = b'$, and $f(c) = c'$, and such that the restriction of f to any one side is an isometry. We say that T satisfies the *CAT(k) condition* if the following *CAT(k) inequality with data* (T, u, v) holds for all $u, v \in T$:

$$d(u,v) \leq d_k(u',v'), \qquad \text{where } u' = f(u),\ v' = f(v)$$

The space (X,d) is CAT(k) if there is a geodesic segment between every pair of points $x, y \in X$, and all geodesic triangles satisfy the CAT(k) condition. We say that X has curvature $\leq k$ if it is locally CAT(k), i.e. for every $x \in X$, a sufficiently small metric ball $B(x, r_x)$ is CAT(k) when equipped with the subspace metric. In particular, X is said to be *nonpositively curved* if it is of curvature ≤ 0.

There are many conditions equivalent to the CAT(k) conditions mentioned above. Some of them are given in the following result which shows that some seemingly weaker conditions are actually equivalent to CAT(k).

THEOREM 4.1. *Suppose (X,d) is a geodesic space and $k \leq 0$. The following are equivalent.*

(a) *X is CAT(k).*
(b) *The CAT(k) inequality with data (T, x, v) holds whenever T is a geodesic triangle in X with vertices x, y, z, and $v \in [y,z]$.*
(c) *The CAT(k) inequality with data (T, x, m) holds whenever T is a geodesic triangle in X with vertices x, y, z, and m is the midpoint of $[y,z]$.*

EXERCISE 4.2. Show that if T is a Euclidean triangle with vertices u, v, w, and m is the midpoint of $[v,w]$, then

$$|u-v|^2 + |u-w|^2 - 2|u-m|^2 = \frac{|v-w|^2}{2}$$

The above exercise, combined with Theorem 4.1 gives us the following characterization of CAT(0) for a geodesic space (X,d):

$$d(x,y)^2 + d(x,z)^2 - 2d(x,m)^2 \geq \frac{d(y,z)^2}{2}, \qquad \text{whenever } x,y,z,m \in X,\ d(y,z) = 2d(y,m) = 2d(m,z)\,. \tag{4.3}$$

This inequality is called the *CN inequality of Bruhat and Tits*; here, *CN* stands for *courbure négative*.

To appreciate the difference between CAT(k) and curvature $\leq k$, we examine the case of Riemannian spaces. The following result follows by combining II.1.5, II.1A.8, and II.4.1 (2) of [**14**].

THEOREM 4.4. *Suppose $k \leq 0$. A Riemannian manifold has curvature $\leq k$ if and only if its sectional curvature is at most k. It is CAT(k) if and only if it has curvature $\leq k$ and is simply connected.*

A simply connected Riemannian manifold of nonpositive sectional curvature is called an *Hadamard manifold*, so it follows in particular from the above result that the classes of CAT(0) manifolds and Hadamard manifolds coincide.

Note that Theorem 4.4 is one significant difference between the cases $k \leq 0$ and $k > 0$ of CAT(k) theory: for instance, the unit circle is CAT(1) but not simply connected.

The next result indicates the relationship between CAT(k) conditions for different values of k. In particular, it follows that $\mathbb{H}^n$ is CAT(0) for all n (a fact that also follows from Theorem 4.4).

THEOREM 4.5. *Suppose $k < 0$. A metric space is CAT(k) if and only if it is CAT(j) for all $k < j \leq 0$.*

In view of the above theorem, we define a CAT($-\infty$) space to be a space that is CAT(k) for all $k \leq 0$.

We now give some other examples of CAT(k) spaces and spaces of curvature at most k, mostly in the form of exercises.

EXERCISE 4.6. Geodesic graphs are defined in Section 6. Show that the following are equivalent for a geodesic graph (G, d):

(1) G is CAT(0).
(2) G is CAT($-\infty$).
(3) G is a tree.

Deduce that G is always of curvature $\leq k$ for all $k \leq 0$.

An $\mathbb{R}$-*tree* is a metric space T such that between each $x, y \in T$, there is a unique geodesic segment, which we denote $[x, y]$, and such that $[x, y] \cup [y, z] = [x, z]$ whenever $[x, y] \cap [y, z] = \{y\}$. The class of $\mathbb{R}$-trees includes all trees (i.e. all simply connected metric graphs). One example of an $\mathbb{R}$-tree that is not a tree is $\mathbb{R}^2$ with the metric $d(x, y) = x_2 + |x_1 - y_1| + y_2$, where $x = (x_1, x_2)$, $y = (y_1, y_2)$: note that $[x, y]$ is in general the union of three line segments, first vertical, then horizontal, and then vertical again.

EXERCISE 4.7. A metric space is CAT($-\infty$) if and only if it is an $\mathbb{R}$-tree.

After graphs, the next obvious examples to examine are simplicial complexes. These are explored in detail in [**14**, II.5], but suffice it to say here that 2-dimensional complexes are of curvature at most 0 if and only if the set of directions at each vertex equipped with the angle metric contains no loops of length less than 2π.

Normed spaces do not give any interesting CAT(k) examples according to the following result.

THEOREM 4.8. *The only CAT(k) normed spaces are inner product spaces. These are always CAT(0), but are CAT(k) for $k < 0$ only if one-dimensional.*

Because of the scale invariance of the CAT(0) condition, the tangent space at a point of a CAT(0) Finsler space must also be CAT(0). (Much more generally, any ultralimit of a sequence of CAT(0) spaces is CAT(0): see [**14**, II.3.10 (3)].) This fact and the previous theorem together yield the following result.

THEOREM 4.9. *A CAT(0) Finsler space is necessarily Riemannian (and so it is an Hadamard manifold).*

Finally, we look at some ways of getting new CAT(k) spaces from old ones, specifically subsets and products. Suppose first that Y is a nonempty subset of a geodesic space (X, d). Since we want Y to be geodesic also, the appropriate metric on Y is the induced length metric (as defined in Section 6). The simplest situation is when Y is *convex*, meaning that geodesic segments in X connecting pairs of points in Y are fully contained in Y. In this case, the subspace metric on Y is the same as its induced length metric, and it is geodesic. The subset Y is said to be *locally convex* if for every $y \in Y$, there exists $r_y > 0$ such that $B(y, r_y) \cap Y$ is convex. For instance, all open subsets are locally convex.

EXERCISE 4.10. Suppose $k \leq 0$. Show that a convex subset of a CAT(k) space is CAT(k), and that a locally convex subset of a space of curvature $\leq k$ is a space of curvature $\leq k$.

EXERCISE 4.11. Show that the complement of the unit disk in the Euclidean plane, when equipped with the induced length metric, has curvature ≤ 0.

EXERCISE 4.12. Show that the complement of the unit ball in Euclidean 3-space, when equipped with the induced length metric, does not have curvature ≤ 0.

EXERCISE 4.13. Show that the product $Z = X \times Y$ of CAT(0) spaces is CAT(0), if we attach the metric d_Z defined by

$$[d_Z((x, y), (x', y'))]^2 = [d_X(x, x')]^2 + [d_Y(x, x')]^2 .$$

Hint: use the Bruhat-Tits characterization of CAT(0) given by (4.3).

Warped products, as defined for instance by Alexander and Bishop [**1**], are a very useful tool in differential geometry. The following result [**1**] shows that many warped products preserve CAT(0). Note that by taking $f \equiv 1$ it implies Exercise 4.13, at least for complete spaces (although this is like using a sledgehammer to crack a nut!).

THEOREM 4.14. *If B and F are complete CAT(0) spaces and $f : B \to (0, \infty)$ is convex, then the warped product $B \times_f F$ is CAT(0).*

4.2. Angles in CAT(k) spaces. In any metric space (X, d), there is a rather simple-minded way of defining a *three-point angle* $A_x(y, z)$ where $x, y, z \in X$, $x \neq y$, and $x \neq z$: we simply pretend we are computing an angle for a Euclidean triangle at the point x and use the cosine rule to get

$$A_x(y, z) = \cos^{-1}\left(\frac{b^2 + c^2 - a^2}{2bc}\right),$$

where $a = d(y, z)$, $b = d(x, y)$, and $c = d(x, z)$.

However, it is more useful to define a notion of (infinitesimal) angle

$$\angle(\lambda, \nu) \equiv \lim_{t,t' \to 0^+} A_x(\lambda(t), \nu(t')).$$

between two geodesics paths $\lambda : [0, T] \to X$, $\nu : [0, T'] \to X$, satisfying $\lambda(0) = \nu(0) = x$.

In the Euclidean case, $A_x(\lambda(t), \nu(t'))$ is independent of t and t', so no limit is necessary to define $\angle(\lambda, \nu)$.

In the hyperbolic plane, though, $A_x(\lambda(t), \nu(t'))$ is always larger than $\angle(\lambda, \nu)$ for all $0 < t \leq T$, $0 < t' \leq T'$, so employing a limit is essential. The angle $\angle(\lambda, \nu)$ agrees with the Euclidean angle between these geodesics in either of the Poincaré models, but it has the advantage of being an intrinsic definition.

In a general metric space, the limit might not exist, so we define the *upper* and *lower* angles $\overline{\angle}(\lambda, \nu)$ and $\underline{\angle}(\lambda, \nu)$ using lim sup and lim inf, respectively.

For CAT(k) spaces, it turns out that $\angle(\lambda, \nu)$ always exists. In fact, we have the following result.

THEOREM 4.15. *Suppose (X, d) is a CAT(k) space, $k \leq 0$, and let $\lambda : [0, T] \to X$, $\nu : [0, T'] \to X$, be geodesic paths satisfying $\lambda(0) = \nu(0) = x$. Then the three-point angle $A_x(\lambda(t), \nu(t'))$ is a monotonically increasing function of both t and t'.*

It follows that if X is CAT(k) and the geodesic paths λ, ν have unit speed parametrizations, then

$$\angle(\lambda, \nu) = \lim_{t \to 0} \cos^{-1}\left(\frac{2t^2 - [d(\lambda(t), \nu(t))]^2}{2t^2}\right) = \lim_{t \to 0} 2\sin^{-1}\left(\frac{d(\lambda(t), \nu(t))}{2t}\right).$$

As stated in Theorem 4.8, a normed space is CAT(0) if and only if it is an inner product space. This can be proved by showing that in any other normed space, there exists a pair of directions such that the angle between the geodesics emanating from the origin in those two directions fails to exist. Rather than prove this, let us investigate it in the special setting of the L^p plane.

EXERCISE 4.16. Let $X = \mathbb{R}^2$ with the L^p metric $\|(x, y)\| = (|x|^p + |y|^p)^{1/p}$ attached for some $1 < p < \infty$. Consider the coordinate axis geodesic rays $\lambda(t) = (t, 0)$ and $\nu(t) = (0, t)$, $t \geq 0$.

(a) Show that the associated three-point angles are dilation invariant:

$$A_0(\lambda(ct), \nu(ct')) = A_0(\lambda(t), \nu(t'))\,, \qquad 0 < c \le 1,\ 0 < t, t'\,.$$

Consequently to study the angle $A_0(\lambda(t), \nu(t'))$, it suffices to study

$$f(t) := \cos[A_0(\lambda(t), \nu(1/t))]\,, \qquad 0 < t\,.$$

(b) Show that $f(t) \to 0$ as $t \to 0^+$ (and so by symmetry as $t \to \infty$).
(c) Show that if $1 < p < 2$, then f is strictly increasing on $(0, 1]$ (and so by symmetry, strictly decreasing on $[1, \infty]$).
(d) Show that if $2 < p < \infty$, then f is strictly decreasing on $(0, 1]$ (and so by symmetry, strictly increasing on $[1, \infty]$).

It follows from the above exercise that the angle between the coordinate axis geodesics λ, ν does not exist in the L^p plane, except in the Euclidean case $p = 2$. Moreover,

$$\overline{\angle}(\lambda, \nu) = \begin{cases} \pi/2, & 1 < p \le 2, \\ \cos^{-1}(2^{2/p-1} - 1), & 2 < p < \infty, \end{cases}$$

$$\underline{\angle}(\lambda, \nu) = \begin{cases} \cos^{-1}(2^{2/p-1} - 1), & 1 < p < 2, \\ \pi/2, & 2 \le p < \infty. \end{cases}$$

4.3. The ideal boundary of a CAT(0) space. We already defined the ideal boundary $\partial_I \mathbb{H}^2$ of the hyperbolic plane $\mathbb{H}^2$, although it was model specific. In this section, we give an intrinsic definition of the ideal boundary $\partial_I X$ of a metric space (X, d). In general, this does not have nice properties and is not very useful, but for complete CAT(0) spaces it is well-behaved.

Given a metric space (X, d), we define $\mathrm{GR}(X)$ to be the class of geodesic rays in X parametrized by arclength, and $\mathrm{GR}(X, o)$ to be the class of all rays in $\mathrm{GR}(X)$ with initial point $o \in X$. We say that two geodesic rays γ, ν are equivalent, $\gamma \sim \nu$, if $d_H(\gamma, \nu) < \infty$. Here d_H is the Hausdorff distance associated with the metric d, so that

$$d_H(\gamma, \nu) = \max\{ \sup_{x \in \gamma} \mathrm{dist}(x, \nu),\ \sup_{x \in \nu} \mathrm{dist}(x, \gamma) \}\,.$$

It is easy to see that if $\gamma, \nu \in \mathrm{GR}(X)$, then $\gamma \sim \nu$ if and only if

$$\sup_{t \ge 0} d(\gamma(t)), \nu(t)) < \infty.$$

We define the *ideal boundary* $\partial_I X$ to be $\mathrm{GR}(X)/\sim$. We also write $\overline{X}_I = X \cup \partial_I X$ and $\partial_{I,o} X = \mathrm{GR}(X, o)/\sim$.

For general spaces, this is not such a nice definition because there is no natural way of topologizing the boundary and also because of the examples such as the following one.

EXERCISE 4.17. Let $X = C_+ \cup C_- \cup (\bigcup_{n=0}^{\infty} L_n)$, where the curves $C_\pm$ are given by

$$C_\pm = \{\, z = x + iy \in \mathbb{C} \mid x \geq 0,\ y = \pm 1/(x+1) \,\}$$

and L_n is the line segment $[n^2 + i(n^2+1)^{-1}, n^2 - i(n^2+1)^{-1}]$. Attaching the arclength distance d to X makes X a complete geodesic space. Then $\mathrm{GR}(X, z)$ has either one or two elements depending on whether or not $z \in S$ has non-zero real part. The two distinct geodesic rays λ, ν emanating from real z are inequivalent despite the fact that $\liminf_{t\to\infty} d(\lambda(t), \nu(t)) = 0$.

For complete CAT(0) spaces X, these pathologies disappear: we can attach an intrinsically defined topology (the *cone topology*) to $\partial_I X$ which is consistent with the non-intrinsic topologies that $\partial_I \mathbb{H}^2$ inherits from the Euclidean structure of our previous models of $\mathbb{H}^2$, there is a natural homeomorphism between $\partial_I X$ and $\partial_{I,o} X$ for any $o \in X$, and $\liminf_{t\to\infty} d(\lambda(t), \nu(t)) = \infty$ for any pair of inequivalent rays.

But even complete CAT(0) spaces can have some features not seen in either Euclidean space or the hyperbolic plane, notably the fact that an unbounded space might have an empty ideal boundary.

EXERCISE 4.18. Let $X \subset \mathbb{C}$ consist of all line segments from the origin to $1 + ni$, $n \in \mathbb{N}$, with the arclength metric attached; see Figure 8. Prove that X is an unbounded complete CAT(k) space for all $k \leq 0$, but that $\mathrm{GR}(X)$ is empty.

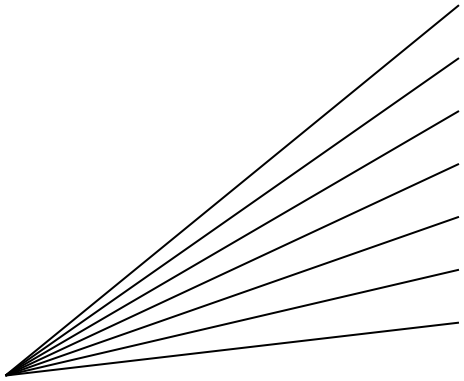

FIGURE 8. An unbounded space with no ideal boundary

The key to the natural homeomorphism between $\partial_I X$ and $\partial_{I,o} X$ for a complete CAT(0) space X and a point $o \in X$ is the following result, whose proof we outline because it illustrates the role of completeness in the theory of CAT(0) spaces: specifically it allows us to make the same sort of limiting arguments that for general metric spaces would require the much stronger assumption that the space is *proper* (meaning that all closed balls are compact).

PROPOSITION 4.19. *Suppose (X, d) is a complete CAT(0) space, $o, x \in X$, and $\lambda \in \mathrm{GR}(X, x)$. Then there exists $\nu \in \mathrm{GR}(X, o)$ such that $\nu \sim \lambda$.*

The idea of the proof is to first let ν_j, $j \in \mathbb{N}$, be the sequence of geodesic segments from o to $\lambda(j)$, parametrizing each ν_j by arclength. Given $s \geq 0$, it follows that $\nu_j(s)$ is defined for all sufficiently large n.

If X were proper, we could extract a subsequence of $(\nu_j(s))$ that converges. By iterating this subsequence procedure, we could construct a geodesic ray. However, things are much easier in a complete CAT(0) space once we prove the following result which is left as an exercise.

EXERCISE 4.20. Prove, using only the CAT(0) condition for (X, d), that $(\nu_j(s))$ is a Cauchy sequence for each $s \geq 0$.

With the above exercise in hand, we simply use completeness and let $\nu(s) = \lim_{j\to\infty} \nu_j(s)$. As the pointwise limit of geodesics, ν is itself a geodesic ray, and we have established Proposition 4.19.

4.4. The cone topology. If (X, d) is a complete CAT(0) space, we attach the *cone topology* τ_{C} to $\overline{X}_I$. This topology is defined using a basepoint $o \in X$, but is independent of the choice of o. For a detailed definition, see [**14**, II.8.5], but we briefly define the concept here. First, in any complete CAT(0) space, there is a unique geodesic γ_x from o to $x \in \overline{X}_I$ parametrized by arclength. This was already mentioned in §4.1 for the case $x \in X$, and is proven in [**14**, II.8.2] for $x \in \partial_I X$; in this latter case, we mean that $\gamma_x \in \mathrm{GR}(X, o)$ and $[\gamma_x] = x$. Let $X_r := \partial_I X \cup (X \setminus \overline{B(o, r)})$, let $p_r : X_r \to S_d(o, r)$ be the "projection" defined by $p_r(x) = \gamma_x(r)$, and let the set $U(a, r, s)$, $r, s > 0$, consist of all $x \in X_r$ such that $d(p_r(x), p_r(a)) < s$. Then τ_{C} is the topology on $\overline{X}_I$ which coincides with the d-topology on X, and has as a local base at $a \in \partial_I X$ the sets $U(a, r, s)$, $r, s > 0$. It is easily verified that τ_{C} is Hausdorff and, since it can be defined as an inverse limit topology, τ_C is compact whenever X is proper.

To help the reader understand this concept, we briefly work through the concepts in the case where X is the Euclidean plane and o is the usual origin. Then $\partial_I X$ consists of the set of rays (or directed lines) in the plane with all lines pointing in the same direction identified. Thus it is the set of all directions. The projection p_r radially retracts all points in $\mathbb{R}^2 \setminus \overline{D(0, r)}$ to $\partial D(0, r)$ and does the same to the set of directions (viewed as rays emanating from the origin). For a given direction a, and positive numbers r and s, the set $U(a, r, s)$ consists of all points in $\mathbb{R}^2 \setminus \overline{D(0, r)}$ and all directions that are pulled back under p_r to the spherical disk consisting of all points on $\partial D(0, r)$ that lie within a Euclidean distance s of $p_r(a)$. Thus $U(a, r, s)$ consists of the part lying outside $\overline{D(0, r)}$ of a cone with vertex at the origin, vertex angle $4 \sin^{-1}(s/r)$, and axis of symmetry in the direction a.

EXERCISE 4.21. Show that $(\overline{\mathbb{R}^2_I}, \tau_{\mathrm{C}})$ is homeomorphic to a closed Euclidean disk.

Using our earlier non-intrinsic definition of ideal boundary, we saw that the ideal boundary of the hyperbolic plane could naturally be given the topology of the circle. This agrees with the cone topology, and is a special case of the following result.

THEOREM 4.22. *If X is an Hadamard manifold, then $\partial_I X$ is homeomorphic to the $(n-1)$-sphere.*

If we want the ideal boundary $\partial_I X$ to have a natural metric that generates the cone topology, we need to assume X is negatively curved, rather than just nonpositively curved, i.e. such a metric exists on the ideal boundary of a CAT(k) space, $k < 0$. We do this in the more general setting of Gromov hyperbolic spaces in the next section.

We note though that there are at least two natural and useful metrics on $\partial_I X$, the angular metric and the Tits metric (the latter being the inner version of the former); see, for instance [**14**, II.9]. These metrics give a topology that, while finer than τ_{C}, might not coincide with τ_{C}. For instance, the angular distance on the Euclidean disk is the usual metric on $\mathbb{S}^1$ and so coincides with τ_{C}, while the angular metric on $\mathbb{H}^2$ is discrete while τ_{C} is still homeomorphic to $\mathbb{S}^1$.

4.5. Weaker notions of nonpositive curvature. CAT(0) was not the first notion of nonpositive curvature for metric spaces. A simpler notion, introduced in 1948 by Busemann [**22**], is now known as Busemann convexity, and spaces with this property are called Busemann spaces.

A geodesic space (X, d) is a *Busemann space* if the metric is convex in the following sense: given any constant speed geodesics $\gamma_i : [0,1] \to X$, $i = 1, 2$, with $\gamma_1(0) = \gamma_2(0)$, then for all $0 \le t \le 1$, we have

$$d(\gamma_1(t), \gamma_2(t)) \le t d(\gamma_1(1), \gamma_2(1)) \,. \tag{4.23}$$

It is *not* required here that the lengths of γ_1 and γ_2 are the same. A *locally Busemann space* is a locally geodesic space where such a convexity condition holds for paths lying in a ball $B(x, r_x)$, for each $x \in X$, where $r_x > 0$; it was this local version that Busemann mainly studied in [**22**].

EXERCISE 4.24. Prove that if (X, d) is Busemann, then it satisfies the following stronger looking condition: given any constant speed geodesics $\gamma_i : [0,1] \to X$, $i = 1, 2$, it follows that

$$d(\gamma_1(t), \gamma_2(t)) \le (1-t) d(\gamma_1(0), \gamma_2(0)) + t d(\gamma_1(1), \gamma_2(1)) \,. \tag{4.25}$$

EXERCISE 4.26. Show that if a geodesic space (X, d) satisfies a condition of the form (4.23), but only for $t = 1/2$, then it is Busemann.

In view of the previous exercise, the Busemann condition can be recast as follows: in a geodesic triangle, the distance between the midpoints of two sides is at most the distance between the corresponding midpoints of a comparison triangle in M_0. In particular, it is trivial that a CAT(0) space is Busemann, and a nonpositively curved space is locally

Busemann. We could similarly define Busemann variants of the CAT(k) condition, but we will not investigate that in these notes.

It is clear that Busemann spaces are *uniquely geodesic*, i.e. there is only one geodesic segment between any given pair of points. Thus all CAT(0) (and CAT(k), $k < 0$) spaces are uniquely geodesic.

Exercise 4.27. Prove that Busemann spaces are contractible (and so the same is true of CAT(k) spaces for all $k \leq 0$). Hence they are simply connected and all of their higher homotopy groups are trivial.

A Riemannian manifold is locally Busemann if and only if it is of nonpositive (sectional or Alexandrov) curvature [**14**, II.1A.8]. In view of Exercise 4.27, it follows that a Riemannian manifold is Busemann if and only if it is CAT(0).

So far, one could be forgiven for suspecting that CAT(0) and Busemann convexity are equivalent. To see that this is not so, we look at normed spaces.

Exercise 4.28. A normed space is Busemann convex if and only if it is uniquely geodesic. Using also Theorem 4.8, we deduce that a nontrivial L^p space is Busemann convex if and only if $1 < p < \infty$, while it is CAT(0) if and only if $p = 2$.

Using [**40**] and the results of §3.2, we can say exactly when a Hilbert geometry is CAT(0) or Busemann. Unfortunately, we do not get any interesting examples.

Theorem 4.29. *The following are equivalent for a Hilbert metric d_G associated with a bounded convex $G \subset \mathbb{R}^n$:*

(a) *d_G is CAT(0).*
(b) *d_G is Busemann convex.*
(c) *G is an ellipsoid (and (G, d_G) is isometric to H^n).*

A central aspect of the local-to-global transition for manifolds of nonpositive curvature is the Cartan-Hadamard theorem. A version of this can be stated for Busemann convexity and for CAT(k), $k \leq 0$. Note that if O is a covering space of a geodesic metric space (X, d), then we can pull back arclength from X to O, and hence define a metric d_O on O by taking an infimum of the length of paths between x and y; this is the *induced length metric* on O.

Theorem 4.30. *Suppose X is a complete geodesic space, and let U be its universal cover with induced length metric d_U attached.*

(a) *If X is locally Busemann convex, then U is Busemann convex.*
(b) *If X is of curvature $\leq k$, where $k \leq 0$, then U is CAT(k).*

Another condition related to nonpositive curvature is the Ptolemy inequality. We say that a metric space (X, d) is *Ptolemaic* if

$$d(x, y)d(z, w) \leq d(x, z)d(w, y) + d(x, w)d(y, z)\,, \qquad x, y, z, w \in X\,. \tag{4.31}$$

This condition is related to metric space inversions, a tool in metric spaces that is inspired by the concept of spherical inversions in complex analysis; see [**16**] and [**15**]. Note that, unlike CAT(k) and Busemann spaces, Ptolemaic spaces are not required to be geodesic. It is trivial that a subspace of a Ptolemaic space is Ptolemaic.

We list here some features of Ptolemaic spaces which, in particular, show their connection to CAT(0) spaces; these results are taken from [**30**], [**44**], [**15**], and [**20**].

- A metric space is CAT(0) if and only if it is both Busemann and Ptolemaic.
- A normed space is Ptolemaic if and only if it is an inner product space.
- A Riemannian manifold is Ptolemaic if and only if it is CAT(0) (and so an Hadamard manifold).
- A Ptolemaic Finsler space is necessarily Riemannian.
- A simplicial complex with only finitely many isometry classes of simplices is Ptolemaic if and only if it is CAT(0).

5. Gromov hyperbolicity

Gromov hyperbolicity expresses the property of a general metric space to be "negatively curved" in the sense of coarse geometry. Its importance is widely appreciated. Gromov hyperbolicity was introduced by Gromov in the setting of geometric group theory [**32**], [**33**], [**31**], [**25**], but has played an increasing role in analysis on general metric spaces [**12**], [**13**], [**7**], with applications to the Martin boundary, invariant metrics in several complex variables [**6**] and extendability of Lipschitz mappings [**42**]. Here we survey the basics of Gromov hyperbolicity. For detailed expositions, see for instance [**25**], [**31**], [**14**, II.H], or [**47**].

Throughout this section, we write $[x, y]$ to denote a geodesic path form x to y; this is not assumed to be unique.

5.1. Why should complex analysts be interested in Gromov hyperbolicity?

- Many important metrics in complex analysis are frequently Gromov hyperbolic. See §5.6, especially Theorem 5.20 and Theorem 5.21.
- The Gromov boundary is a useful concept, both as an alternative way of treating the topological boundary and as a way of defining boundary extensions of maps. See §5.5.
- For the invariant metrics in complex analysis, it is usually impossible to find the associated geodesics. However, it may be much easier to find quasigeodesics, and geodesics always stay close to quasigeodesics in Gromov hyperbolic spaces. See Theorem 5.7.

5.2. Gromov hyperbolicity: definition and examples. Gromov hyperbolicity can be defined in non-geodesic spaces, but our first definition (the *thin triangles* definition) is valid only in geodesic spaces. It has the virtue of being intuitively simple.

A geodesic space (X, d) is said to have *δ-thin triangles*, $\delta \geq 0$, and all its geodesic triangles are said to be *δ-thin*, if

$$\forall\, x, y, z \in X \;\; \forall\, [x, y],\; [x, z],\; [y, z] \;\; \forall\, w \in [x, z] : \quad d(w, [x, y] \cup [y, z]) \leq \delta\,. \tag{5.1}$$

In other words, a triangle is δ-thin if each of its sides is contained in the δ-neighborhood of the union of the other two sides. We say that X is *Gromov hyperbolic* if it has δ-thin triangles for some $\delta \geq 0$.

Bounded metric spaces are trivially Gromov hyperbolic. We now give some nontrivial examples of Gromov hyperbolic spaces.

Exercise 5.2. Prove that hyperbolic space $\mathbb{H}^n$ has δ-thin triangles with $\delta = \log 3$ for all $n \geq 2$, and with $\delta = 0$ for $n = 1$. *The next paragraph contains a hint that transforms this exercise from a challenging one to something rather routine: you may wish to skip it before you first try the exercise.*

Let us outline how the above exercise can be proved. The case $n = 1$ is rather trivial, since $\mathbb{H}^1$ is isometric to $\mathbb{R}^1$. As for the case $n \geq 2$, it suffices to assume $n = 2$, since any triangle lies in an isometric copy of $\mathbb{H}^2$. Let $R = R(T)$ be the largest hyperbolic radius of a circle that fits inside a hyperbolic triangle T. It suffices to prove that $R(T) \leq (\log 3)/2$. Increasing the sidelength of any one side of T can only increase R (justify this!), so we may as well assume that T is an ideal triangle, all of whose sides are of infinite length. All such triangles are isometric, so we can use the half-plane model and assume that the vertices of T are $-1, 1, \infty$. The circle that maximizes R has Euclidean center $2i$ and radius 1: prove this and find its hyperbolic radius to finish the exercise.

Let us mention an alternative proof of the Gromov hyperbolicity of $\mathbb{H}^2$, although it does not give $\delta = \log 3$. If a triangle in $\mathbb{H}^2$ is not δ-thin, it contains a hyperbolic disk of radius $\delta/2$, which has area $4\pi \sinh^2(\delta/4)$ according to the formula in §2.8. By the Gauss-Bonnet formula, we deduce that the area is at most π, and so $\delta \leq 4 \sinh^{-1}(1/2) \approx 1.925$.

This second proof, although it did not give us as good a constant, hints at the fact that Gromov hyperbolicity can be formulated in terms of a suitably defined concept of area. An account of this can be found in [**14**, III.H.2], for example. Suffice it to say here that Gromov hyperbolicity is equivalent to a coarse linear isoperimetric inequality, i.e. the coarse area of an arbitrary loop γ (defined via triangulations) is at most $K(\mathrm{len}(\gamma) + 1)$ for some fixed K. This is consistent with the fact that the formulae for the perimeter and area of a hyperbolic disk given in §2.8 are comparable.

Exercise 5.3. Deduce from (5.2) that a CAT(k) space, $k < 0$, has δ-thin triangles for $\delta = (\log 3)/\sqrt{-k}$. In particular, the same is true of simply connected Riemannian manifolds of sectional curvature at most k.

The fact that Euclidean n-space for $n > 1$ is not Gromov hyperbolic is simple to prove: the midpoint of a side on a large equilateral triangle is far from all points on the other two sides. It is also trivial that $\mathbb{R}$ has 0-thin triangles.

EXERCISE 5.4. Show that Euclidean space $\mathbb{R}^n$ is Gromov hyperbolic only for $n = 1$, in which case it has 0-thin triangles. Thus CAT(0) spaces are not necessarily Gromov hyperbolic.

We next consider graphs. The following exercise should be compared with Exercise 4.6.

EXERCISE 5.5. Show that a geodesic graph G has δ-thin triangles, $\delta > 0$, if and only if all its loops (meaning isometric copies of Euclidean circles) have length at most 4δ. Moreover, the following are equivalent:

(1) G is CAT(0).
(2) G has 0-thin triangles.
(3) G is a tree.

Gromov hyperbolicity is a rough negative curvature assumption, so it is incomparable with CAT(0): on the one hand, Euclidean space is CAT(0) but not Gromov hyperbolic, on the other hand, spheres, cylinders, and certain graphs are examples of Gromov hyperbolic spaces that are not CAT(0).

5.3. Tripods and geodesic stability. A *tripod* T is a union of three segments in the Euclidean plane, that have only the origin in common; the segments are allowed to have length 0. We attach arclength metric to T. Thus only the lengths of the segments in T are important: the angles between segments are irrelevant.

A *tripod map* for a geodesic triangle S in a geodesic space X is a map from S to a tripod, so that the restrictions of f to each side of S are isometries.

EXERCISE 5.6. Show that such tripod maps always exist, and that if $f_i : S \to T_i$ are two tripod maps, then T_1 and T_2 are isometric.

In a geodesic space (X, d), Gromov hyperbolicity is quantitatively equivalent to the following stronger looking *tripod condition*: there exists some number $A \geq 0$ such that every geodesic triangle S in the geodesic space (X, d) and associated tripod map $f : S \to T$ have the property that $d(x, y) \leq A$ whenever $f(x) = f(y)$.

The tripod condition says that any configuration of three points in X are roughly isometric to the vertices of a tree (namely, the tripod). This can be generalized to any finite number of points: if (X, d) has δ-thin triangles, and S consists of a set of points $\{x_1, \ldots, x_n\} \subset X$ together with geodesic segments $[x_j, x_k]$, $1 \leq j < k \leq n$, then there exists a metric tree (T, d') with finitely many vertices and edges and a map $f : S \to T$ which takes the points x_i to vertices of T, and for which $|d'(f(a), f(b)) - d(a, b)| \leq A$, where A depends only on δ and n.

We next discuss the connection between hyperbolicity and geodesic stability. In the complex plane, there is only one optimal way of getting from z to w: a straight line segment. However if we allow "limited suboptimality", the set of "reasonably efficient paths" are well spread. For instance, if we split the circle $\partial D(0,R) \subset \mathbb{C}$ into its two semicircles between the points R and $-R$, then we have two such reasonably efficient paths between these endpoints such that the point Ri on one of the semicircles is far from all points on the other semicircle if R is large. Even an additive suboptimality can lead to paths that fail to stay close together. For instance, the union of the two line segments in $\mathbb{C}$ given by $[0, R+\sqrt{R}i]$ and $[R+\sqrt{R}i, 2R]$ gives a path of length less than $2R+1$, and so is "additively inefficient" by less than 1. However, its corner point is very far from all points on the line segment $[0, 2R]$ when R is very large.

The situation in Gromov hyperbolic spaces is very different: all such reasonably efficient paths stay within a bounded distance of each other:

THEOREM 5.7. *A geodesic space with δ-thin triangles, $\delta \geq 0$, is geodesically stable in the sense that if λ, ν are (a,h)-quasigeodesics for some $a \geq 1$, $h \geq 0$, then $d_H(\lambda, \nu) \leq R$, where d_H is Hausdorff distance (as defined in §4.3) and $R = R(\delta, a, h)$. Conversely such geodesic stability implies that the space has δ-thin triangles, with δ dependent only on the function $R = R(a,h)$.*

The first part of the above result can be found in [**25**], while the converse was proven by Bonk [**11**]; see also [**19**]. See Section 6 for the definition of a quasigeodesic.

We next discuss a related result for local quasigeodesics. It is said that if a person tries to walk straight ahead in a large featureless desert, he will eventually end up back at the same point. The idea is that our gait is not perfectly symmetrical so that we naturally tend to go in a big circle. However if we walk straight enough in a desert shaped like the hyperbolic plane, this phenomenon cannot occur. In fact, the following useful companion to Theorem 5.7 says that it also cannot occur if our desert is shaped like any Gromov hyperbolic space: "walking straight enough" means following a (a,h,L)-local quasigeodesic for a close to 1, h close to 0, and L large.

THEOREM 5.8. *Suppose (X,d) is a geodesic space with δ-thin triangles, $\delta \geq 0$. Given $a \geq 1$ and $h \geq 0$, there exist numbers L, a', h' such that every (a,h,L)-local quasigeodesic is an (a',h')-quasigeodesic.*

5.4. Gromov product and quasi-isometries. In any metric space (X,d), we define the *Gromov product* with basepoint $p \in X$ by

$$\langle x,\, y\rangle_p = (d(x,p) + d(p,y) - d(x,y))/2, \qquad x, y \in X\,.$$

The Gromov product can be used to give an alternative definition of Gromov hyperbolicity:

$$\langle x,\, z\rangle_w \geq \min(\langle x,\, y\rangle_w\,, \langle y,\, z\rangle_w) - \delta, \qquad x, y, z, w \in X. \tag{5.9}$$

We say that a metric space (X, d) is *δ-hyperbolic* if it satisfies the above condition. Now, δ-hyperbolicity and δ-thin triangles are quantitatively equivalent: either condition implies the other one with δ replaced by 3δ [**47**, 2.34, 2.35].

The Gromov product definition makes it clear that if a space X is Gromov hyperbolic, then so are all spaces that are roughly isometric to X. Thus Gromov hyperbolicity is a concept of "rough geometry".

The big advantage of the Gromov product definition is that it does not involve geodesics or even paths, so it allows hyperbolicity to be defined in a more general context. This is also a rather natural thing to do since the class of paths is not closed under rough isometries.

The disadvantage of the Gromov product definition is that its geometric meaning is unclear at first, whereas the thin triangles definition is very easy to understand geometrically.

EXERCISE 5.10. Show that if a metric space (X, d) satisfies the inequality in (5.9) for all $x, y, z \in X$ and one particular $w \in X$, then it is (2δ)-hyperbolic.

The following useful estimate [**47**, 2.33] is the key to understanding the geometric meaning of the Gromov product definition: if x, y lie in a δ-hyperbolic space (X, d), and γ is an arc from x to y with $\operatorname{len}(\gamma) \le d(x, y) + h$, then

$$\operatorname{dist}_d(w, \gamma) - 2\delta - h \le \langle x,\, y\rangle_w \le \operatorname{dist}_d(w, \gamma) + h/2. \tag{5.11}$$

Indeed only the lower bound requires hyperbolicity.

In particular if γ is a geodesic between x and y then, modulo an additive fudge of 2δ, $\langle x,\, y\rangle_w$ equals $\operatorname{dist}_d(w, \gamma)$. Suppose now that (X, d) is a geodesic space and that T is a geodesic triangle with vertices x, y, z. Then (5.9) is equivalent to the statement that the distance from a point w to the union of the two sides $[x, y] \cup [y, z]$ is, modulo a bounded additive fudge, no larger than its distance to $[x, z]$. This last statement is easily seen to be equivalent to (5.1), with quantitative dependence of parameters.

The following exercise is reminiscent of Exercise 4.7; see also Exercise 5.5.

EXERCISE 5.12. Show that a (not necessarily geodesic) metric space is 0-hyperbolic if and only if it is isometric to a subspace of an $\mathbb{R}$-tree. Show also that a geodesic space is 0-hyperbolic if and only if it is an $\mathbb{R}$-tree.

We now state a fundamental result about hyperbolicity and quasi-isometries; see Section 6 for the definitions of quasi- and rough isometries and rough similarities.

THEOREM 5.13. *Suppose (X, d) and (Y, d') are metric spaces, that (Y, d) is Gromov hyperbolic, and that $f : X \to Y$ is some map.*

(a) *If f is a rough isometry, or more generally a rough similarity, then X is also Gromov hyperbolic.*
(b) *If f is a quasi-isometry and X is a geodesic space, then X is Gromov hyperbolic.*

The original context of hyperbolicity was hyperbolic groups. Suppose a group G is finitely generated, considering the inverse of a generator to be a generator also. We can attach a metric to G by the rule that $d(g,h)$ equals the minimum number of generators that need to be multiplied to produce gh^{-1}. The metric space (G,d) is not geodesic but it is easy to find a rough isometry from G to a geodesic space X: whenever $d(g,h)=1$, add an edge of length 1 between the vertices g,h. This gives the *Cayley graph* (X,d_X). The rough isometry is the natural identification of group elements with vertices in the Cayley graph.

We call the group hyperbolic if the metric d is hyperbolic. The metric d is dependent on our choice of generators, but the concept of a hyperbolic group is not:

EXERCISE 5.14. Suppose the metrics d and d' on a group G are defined as above by two different finite sets of generators.

(a) Prove that d and d' are quasi-isometric.
(b) Deduce that (G,d) is Gromov hyperbolic if and only if (G,d') is. *Note that this does not follow immediately from Theorem 5.13(b) because these metric spaces are not geodesic.*

There are many examples in the literature of hyperbolic groups. Here we mention only that all free groups are Gromov hyperbolic.

5.5. The Gromov boundary. Let (X,d,o) be a pointed metric space, and let $\langle\cdot,\cdot\rangle \equiv \langle\cdot,\cdot\rangle_o$ denote the Gromov product with respect to the basepoint o. A sequence $x=(x_j)$ in X is a *Gromov sequence* if $\langle x_j,\, x_k\rangle \to \infty$ as $j,k\to\infty$. Intuitively this says that, for large j,k, the points x_j and x_k are much closer to each other than to o, and one should think of a Gromov sequence as a sequence that marches off to infinity. As such, it is a replacement for the geodesic rays that we considered for CAT(0) spaces.

We define a binary relation E on the set of Gromov sequences as follows:

$$x\ E\ y \quad\Longleftrightarrow\quad \liminf_{i,j\to\infty} \langle x_i,\, y_j\rangle = \infty.$$

Intuitively, xEy if x and y are marching off to infinity in the same direction. Thus this relation is a replacement for the equivalence relation that we used to get the ideal boundary from the collection of geodesic rays. The relation E is indeed an equivalence relation when X is Gromov hyperbolic, but we wish to talk about the Gromov boundary of more general spaces later in this section, so we use E to define an equivalence relation in the obvious way: $x\sim y$ if there is a finite chain of sequences x^k, $0\le k\le k_0$, such that

$$x=x^0, \quad y=x^{k_0}, \quad \text{and} \qquad x^{k-1}\ E\ x^k, \quad 1\le k\le k_0.$$

EXERCISE 5.15. Prove that $\sim$ is an equivalence relation. Given an example of a metric space where E is not an equivalence relation. *Hint: you need look no further than sequences of points along lines in the Euclidean plane.*

The *Gromov boundary* $\partial_G X$ is the set of all equivalence classes $[x]$ of Gromov sequences x, and we write $\overline{X}_G = X \cup \partial_G X$.

If (X, d) is a proper geodesic hyperbolic space, it is well known that the Gromov boundary $\partial_G X$ and the ideal boundary $\partial_I X$ can be identified as sets; see, for instance, [**14**, III.H.3.13]. Indeed, the term "Gromov boundary" is sometimes used for the ideal boundary in this setting.

We extend the Gromov product with basepoint o to $\overline{X}_G \times \overline{X}_G$ via the equations

$$\langle a,\, b\rangle = \inf\,\{\liminf_{i,j\to\infty} \langle x_i,\, y_j\rangle :\ [x] = a,\ [y] = b\}, \qquad a, b \in \partial_G X, \tag{5.16}$$

$$\langle a,\, b\rangle = \inf\,\{\liminf_{i\to\infty} \langle x_i,\, b\rangle :\ [x] = a\}, \qquad a \in \partial_G X,\ b \in X. \tag{5.17}$$

Whenever $\partial_G X$ is nonempty, we define the functions $\rho_\epsilon, d_\epsilon : \partial_G X \times \partial_G X \to [0, \infty)$ for $\epsilon > 0$ by the equations

$$\rho_\epsilon(a, b) = \exp(-\epsilon\, \langle a,\, b\rangle), \qquad a, b \in \partial_G X, \tag{5.18}$$

$$d_\epsilon(a, b) = \inf \sum_{j=1}^{n} \rho_\epsilon(a_{j-1}, a_j), \qquad a, b \in \partial_G X, \tag{5.19}$$

where the infimum is taken over all finite sequences $a = a_0, \ldots, a_b = b$, in $\partial_G X$. Clearly, d_ϵ is a pseudometric but it can happen that there are distinct points a, b with $d_\epsilon(a, b) = 0$. However if X is δ-hyperbolic and $\epsilon\delta \le 1/5$, then d_ϵ is actually a metric and in fact

$$\rho_\epsilon(a, b)/2 \le d_\epsilon(a, b) \le \rho_\epsilon(a, b), \qquad a, b \in \partial_G X.$$

The metrics d_ϵ, together with all metrics on $\partial_G X$ bilipschitz equivalent to some power of some d_ϵ, form a *canonical gauge of metrics* on $\partial_G X$. All these metrics generate the same *canonical topology* τ_G on $\overline{X}_G$. We automatically associate τ_G with $\overline{X}_G$ and $\partial_G X$. If X is a proper geodesic space then $\overline{X}_G$ is compact.

At this stage we should pause for some examples. First, the Gromov boundary of Euclidean space has only one point. This is very different from its ideal boundary (a sphere), but such differences occur when the domain is not Gromov hyperbolic.

The Gromov boundary of $\mathbb{H}^n$ is homeomorphic to $\mathbb{S}^{n-1}$. The canonical gauge of metrics includes the usual spherical metric ρ, but most metrics in the gauge have no associated rectifiable paths (since they are comparable to ρ^t for some $0 < t < 1$).

The unit ball in $\mathbb{C}^n$, with the Carathéodory metric attached, is Gromov hyperbolic. Its Gromov boundary is homeomorphic to $\mathbb{S}^{2n-1}$. However, the metrics in the canonical gauge are a lot more complicated than for hyperbolic space: one of them is a *sub-Riemannian metric* whose tangent space is the Heisenberg group. For more on sub-Riemannian geometry, see the survey by Bellaïche [**10**].

The Gromov boundary of the free group on two generators is a Cantor set.

One of the most important features of the transition from a Gromov hyperbolic space to its Gromov boundary is that it is functorial. If $f : X \to Y$ is in a certain class of

maps between two Gromov hyperbolic spaces X and Y, then there is a boundary map $\partial f : \partial X \to \partial Y$ which is in some other class of maps.

One does not expect ∂f to be in the same class of maps as f, since the transition from the metric on X to the one on ∂X involves an exponential which changes additive fudges to multiplicative fudges, and multiplicative fudges to exponentiated fudges. Thus it is not surprising that if f is a rough isometry, then ∂f is a bilipschitz map. If f is a rough similarity, then ∂f is a snowflake map (i.e. $d_Y(f(x), f(y))$ is comparable to some fixed power of $d_X(x, y)$), and if f is a quasi-isometry, then ∂f is a power quasisymmetry (meaning that the ratio $d_X(x, y)/d_X(x, z)$ is bounded above and below by constant multiples of fixed powers of the corresponding ratio on the image side).

As one application of this functoriality, consider the case of the Carathéodory metric on a bounded suitably smooth strictly pseudoconvex domain in $\mathbb{C}^n$ which is known to be Gromov hyperbolic, and the Gromov boundary is homeomorphic to the topological boundary (see Theorem 5.21 below for more precise statements). Thus the boundary map ∂f associated with a map between two such domains X and Y, is essentially a boundary extension of f and as such is a coarse geometry version of the famous boundary extension result of C. Fefferman [**26**] for biholomorphic mappings between strictly pseudoconvex domains. Fefferman's result gives much more precise information, but the *Gromov functor* gives information about classes of maps that are much more general than biholomorphic maps.

In applications to various areas of mathematics, the Gromov boundary can similarly be shown (under appropriate conditions) to coincide with other "finite" boundaries, such as the Euclidean or inner Euclidean boundary, or the Martin boundary, so we obtain a variety of boundary extension results as above.

5.6. Gromov hyperbolicity of some important metrics. Here we discuss when some of the metrics defined previously are Gromov hyperbolic, and when their Gromov boundaries can be identified with some natural "finite" boundaries. As explained above, such identifications are important as they allow us to define boundary extensions of certain mappings.

In general, although Gromov hyperbolicity has a simple definition (especially the thin triangles version), it is far from simple to verify for many of the metrics that arise in analysis. Let us begin with an easy case: the Poincaré and quasihyperbolic metrics on a simply connected domain $\Omega \subset \mathbb{C}$. We already know that the Poincaré metric is of constant curvature -1, and so it is CAT(-1) on any simply connected domain $\Omega \subset \mathbb{C}$ by Theorem 4.30(b). Thus by Exercise 5.3, it is Gromov hyperbolic. Since the quasihyperbolic metric k is bilipschitz equivalent to the Poincaré metric on any simply connected planar domain, and since is a geodesic metric, it is also Gromov hyperbolic in this setting by Theorem 5.13(b).

Much more generally, Bonk, Heinonen, and Koskela [**12**] and Balogh and Buckley [**7**] explore the quasihyperbolic metric in a metric space setting. It is beyond the scope

of these notes to properly describe the results in those papers, so we only give a vague description. To define the quasihyperbolic metric in a metric space (X, d), we first need to define its boundary $\partial_d X$. This can be done in a simple intrinsic manner that is consistent with the topological boundary of X whenever X is an open subspace of a complete space Y: we simply take $\partial_d X$ to be the collection of elements of the metric completion of X that are not in X (under the natural identification of X with a subset of its completion).

It is shown in [**12**] that proper geodesic Gromov hyperbolic spaces are intimately connected with bounded uniform spaces. A *uniform space* is a locally compact metric space (X, d) in which every pair of points x, y can be joined by a path γ that is at most a fixed multiple of $d(x, y)$ in length, and the distance of any point on γ from the boundary is at least some fixed multiple of the length of the shorter of the two subpaths obtained by cutting the path at this point. Thus the Poincaré geodesics in the unit disk can be used to show that the unit disk with the Euclidean metric is uniform, but the Euclidean geodesics cannot (since they do not necessarily move away from the boundary at a linear rate). The link between uniform and Gromov hyperbolic spaces is, roughly speaking, a generalization of the connection between the Euclidean disk (a uniform space) and the Poincaré disk (a Gromov hyperbolic space).

Bonk, Heinonen, and Koskela [**12**] show that if (X, d) is a uniform space, then the associated quasihyperbolic metric is a proper geodesic Gromov hyperbolic space that has a certain property called *roughly starlike*, and also $\partial_d X$ is naturally equivalent to the Gromov boundary of (X, d). Conversely we can "dampen" a proper geodesic roughly starlike Gromov hyperbolic space (X, ρ) to obtain a bounded uniform domain (X, d_ϵ), dependent on a sufficiently small parameter $\epsilon > 0$, and we can identify the Gromov boundary of (X, ρ) with the metric boundary $\partial_{d_\epsilon} X$. Moreover the two processes are in some sense quasi-inverses of each other. This means that, subject to certain technical assumptions, we can transport questions about typical Gromov hyperbolic spaces to questions about bounded uniform spaces and vice versa. It also implies that in some sense all the typical Gromov hyperbolic spaces that arise in finite dimensional analysis are bilipschitz equivalent to some space with a quasihyperbolic metric attached. Although the above statements of the main results in [**12**] are very vague, suffice it to say that [**12**] provides a fundamentally important toolbox for the study of Gromov hyperbolicity.

Balogh and Buckley [**7**] prove that the quasihyperbolic metric on any of a large class of bounded metric spaces (X, d) is equivalent to the combination of a separation condition and a Gehring-Hayman condition. These latter conditions are often considerably easier to investigate.

The *separation condition* says that if w is a point on a quasihyperbolic geodesic segment between points $x, y \in X$, then a ball of some fixed multiple of the distance from w to $\partial_d X$ "separates" x from y in the sense that any path in X from x to y must intersect the ball.

The Gehring-Hayman condition says that every quasihyperbolic geodesic segment from x to y has d-length at most a constant multiple of $d(x, y)$.

For special classes of domains and Riemann surfaces, there have been quite a number of papers giving necessary or sufficient conditions for the Gromov hyperbolicity of the Poincaré metric. We mention just one of these, which deals with Denjoy domains [**37**]; recall that a *Denjoy domain* in $\mathbb{C}$ is a domain whose boundary is contained on the real axis.

THEOREM 5.20. *Let G be a Denjoy domain with $G \cap \mathbb{R} = (-\infty, 0) \cup \bigcup_{n=1}^{\infty}(a_n, b_n)$, where $a_1 > 0$ and $b_n \le a_{n+1}$ for every n, and $\lim_{n\to\infty} a_n = \infty$. Then the quasihyperbolic and Poincaré metrics on G are Gromov hyperbolic if and only if*

$$\liminf_{n\to\infty} \frac{b_n - a_n}{a_n} > 0\,.$$

Although the j- and $\tilde{j}$-metrics defined in §3.3 are bilipschitz equivalent, they are very different from the viewpoint of Gromov hyperbolicity. Hästö [**36**] showed that, among domains $G \subsetneq \mathbb{R}^n$, the j-metric is only Gromov hyperbolic for $G = \mathbb{R}^n \setminus \{z\}$ whereas $\tilde{j}$ is always Gromov hyperbolic. The quasihyperbolic metric k is intermediate in the sense that it is Gromov hyperbolic for many domains but not for many others: for instance it is Gromov hyperbolic for all uniform domains but, among domains G obtained by removing a countable closed set of points from some uniform domains, it is Gromov hyperbolic if and only if G is uniform, making it easy to construct examples that are not Gromov hyperbolic.

The fact that the j- and $\tilde{j}$-metrics are bilipschitz equivalent, but are very different from the viewpoint of Gromov hyperbolicity is a striking reminder of the importance of the geodicity assumption in Theorem 5.13(b).

Finally, we discuss the invariant metrics on domains in $\mathbb{C}^n$. The outstanding result in this context is the following result of Balogh and Bonk; for proofs and definitions of terms used, see [**5**], [**6**], and [**4**, 4.1].

THEOREM 5.21. *Suppose $\Omega \subset \mathbb{C}^n$, $n \ge 2$, is a bounded strictly pseudoconvex domain, with a C^2-smooth boundary. Equip Ω with the Kobayashi, Carathéodory, inner Carathéodory, or Bergman metric d. Then (Ω, d) is Gromov hyperbolic, and the Euclidean boundary of Ω can be identified with its Gromov boundary. Under this identification, the sub-Riemannian metric induced on $\partial\Omega$ by d is in the canonical quasisymmetric gauge of metrics associated with the Gromov boundary.*

5.7. Comparing boundaries at infinity. Here we discuss how three types of boundary at infinity relate to each other. As we will see, for most Gromov hyperbolic spaces, they are the same, but for more general spaces, they can be quite different.

We have already defined two of these boundaries at infinity, namely the ideal and Gromov boundaries. The third type of boundary at infinity, the conformal boundary, is

motivated by the transition from the Euclidean metric d to the Poincaré metric ρ in the unit disk. When we go from d to ρ, we conformally blow up distances near the topological boundary by a factor $(1-|z|^2)^{-1}$. The integral of $(1-t^2)^{-1}$ on the interval $[0,r]$ grows logarithmically, so we must distort ρ conformally by an exponentially decaying function to recover d. The metric d "sees" the topological boundary of the unit disk in an intrinsic way: it consists of all points in the metric completion of the disk that are not in the disk. When we go to ρ, we have a complete metric, so it has no boundary in this metric sense. But by using an exponentially decaying conformal distortion, we recover the "boundary at infinity" of ρ.

We now generalize this idea. Recall first that the boundary $\partial_d X$ of a metric space (X,d) is the collection of elements in the metric completion of X that are not in X.

Next, given a pointed unbounded length space (X,l,o), we define a new metric σ to be the conformal distortion of l given by $d\sigma(z) = g(|z|)\,|dz|$, where $g:[0,\infty)\to(0,\infty)$ is a nonincreasing function, $|x| = d(x,o)$, and $|dz|$ indicates the length element for l. Thus the σ-length of a path γ is $\int_\gamma g(|z|)\,|dz|$ and we define $\sigma(x,y)$ to be the infimum of the σ-length of paths connecting x and y.

Not all measurable functions $g:[0,\infty)\to(0,\infty)$ are of interest to us. The functions should have a finite integral over $[0,\infty)$ so as to "drag the boundary at infinity back to a finite distance from the origin." They should also satisfy the condition

$$g(t) \le C g(s), \qquad \text{whenever } s,t \ge 0,\ s-1 \le t \le 2s+1\,. \tag{5.22}$$

Valid examples include exponential decay functions, which have been used extensively in geometric analysis (for instance in [**12**]), and $g(t) = t^{-2}$, which has been used in geometric group theory (for instance in [**28**]).

We now have two metrics l and σ (taking the place of ρ and d, respectively, in our above discussion of the unit disk), and so we have two boundaries. There is a natural identification of $\partial_l X$ with a subset of $\partial_\sigma X$. The *conformal boundary* or g-boundary, $\partial_g X$, is simply $\partial X_\sigma \setminus \partial X_l$. Note that $\partial_g X$ inherits a metric, and so a topology, from σ.

Buckley and Kokkendorff [**17**] examined the relationship between these three types of boundary at infinity for general metric spaces, obtaining the following pair of results in which it is assumed that l, σ, and g are as above.

THEOREM 5.23. *Suppose that the numbers $\delta \ge 0$ and $\epsilon > 0$ are such that $\epsilon\delta \le 1/5$. Suppose also that (X,l) is geodesic.*

(a) *If $\partial_I X$ is nonempty, then so is $\partial_G X$, and there is a natural map $J_1 : \partial_I X \to \partial_G X$.*
(b) *If (X,l) is proper, then $\partial_I X$ is nonempty and $J_1 : \partial_I X \to \partial_G X$ is surjective.*
(c) *If (X,l) is δ-hyperbolic, and $\partial_I X$ is nonempty, then $J_1 : \partial_I X \to \partial_G X$ is injective.*
(d) *If (X,l) is δ-hyperbolic, complete, and CAT(0), and $\partial_I X$ is nonempty, then J_1 is a homeomorphism from $(\partial_I X, \tau_C)$ to its image in $(\partial_G X, d_\epsilon)$.*

THEOREM 5.24. *Suppose that the numbers $\delta \geq 0$ and $\epsilon > 0$ are such that $\epsilon\delta \leq 1/5$.*

(a) *If $\partial_G X$ is nonempty, then $\partial_g X$ is nonempty, and there is a natural map $J_2 : \partial_G X \to \partial_g X$.*
(b) *If (X, l) is proper then $\partial_G X$ is nonempty, and if the natural map $J_2 : \partial_G X \to \partial_g X$ exists, then it is surjective.*
(c) *If (X, l) is δ-hyperbolic, $\partial_G X$ is nonempty, and g satisfies the decay condition $g(t)\exp(\epsilon_0 t) \geq K > 0$ for sufficiently small $\epsilon_0 = \epsilon_0(\delta) > 0$, then we have a natural map $J_2 : (\partial_G X, d_\epsilon) \to (\partial_g X, \sigma)$, which is a homeomorphism.*

The natural map in Theorem 5.23 is induced by taking any sequence of points (x_n) on a ray γ that "tend to infinity" (i.e. the distance from x_n to the initial point of γ tends to infinity). The natural map in Theorem 5.24 is induced by taking σ-limits of Gromov sequences.

It is also shown in [**17**] that the above pair of theorems are sharp in the sense that if we drop any assumption anywhere, we get a false statement. The decay condition $g(t)\exp(\epsilon_0 t) \geq K > 0$ in Theorem 5.24 is needed for some $\epsilon_0 \leq 1/\delta$. In fact it is not hard to show that if X is the dilation of $\mathbb{H}^2$ by a factor δ and $g(r) = \exp(-\epsilon r)$ for any $\epsilon > 1/\delta$, then $\partial_g X$ is a singleton set, and so J_2 is not a homeomorphism.

Also, for a general unbounded pointed length space (X, l, o), the *cardinality triple of* X (a, b, c) which is defined as the 3-tuple of cardinalities $(\#(\partial_I X), \#(\partial_G X), \#(\partial_g X))$, can take on any value subject only to the two constraints given by the above theorems, namely that if $a > 0$ then $b > 0$, and if $b > 0$ then $c > 0$.

One would hope that more could be said about the relationship between these different types of boundary at infinity for a CAT(0) space than for a general space, although one cannot expect to get a homeomorphism or a bijection since fort $\mathbb{R}^n$, $n > 1$, the ideal boundary is a sphere but the Gromov and conformal boundaries are singleton sets.

This general question is investigated by Buckley and Kokkendorff in [**18**] where the conformal boundary of a general warped product is found, and this allows one to give examples where the ideal and conformal boundaries are rather different in interesting ways. Let us mention two such examples. Recall from Theorem 4.22 that the ideal boundary of an Hadamard n-manifold is homeomorphic to $\mathbb{S}^{n-1}$. However, certain warped products result in Hadamard n-manifolds whose conformal boundary is either the one-point join of two $(n-1)$-spheres, or a closed $(n-2)$-ball. The ideal and conformal boundaries of an Hadamard n-manifold for $n \geq 3$ are shown to have one property in common: they are both simply connected.

6. Appendix: terminology of metric geometry

Here we gather together some basic terminology and notation for metric spaces (X, d). Recall that $d : X \times X \to [0, \infty)$ is called a *metric* or *distance function* if the following conditions hold for all $x, y, z \in X$:

(a) $d(x,x) = 0$.
(b) $d(x,y) > 0$ if $x \neq y$.
(c) $d(x,y) = d(y,x)$.
(d) $d(x,y) \leq d(x,z) + d(z,y)$.

If we merely assume (a), (c), and (d), then d is called a *pseudometric*.

The length of a path $\gamma : [0,T] \to X$ in a metric space can be defined as in the Euclidean setting, i.e. associated with a partition $0 = t_0 < t_1 < t_2 < \cdots < t_n = T$, we have the sum $\sum_{j=1}^{n} d(\gamma(t_{j-1}), \gamma(t_j))$, and now the length of the curve is the supremum of these sums over all possible partitions. Most of these metrics in these notes have the property that points x, y can always be joined by a path of length arbitrarily close to $d(x,y)$. We call such a metric space a *length space*. If moreover there always exists a path between x, y of length $d(x,y)$, we call (X,d) a *geodesic space*.

Given a rectifiably connected subset Y of a metric space (X,d), the *induced length metric* d_Y on Y is defined by letting $d_Y(x,y)$ be the infimum of the d-lengths of paths in Y from x to y. Note that Y is a length space, but it might not be a geodesic space even if X is geodesic, for example if X is the Euclidean plane and Y is the punctured plane. The *inner metric* d_X *on* X *(associated with* d*)* is another name for the induced length metric when $Y = X$; trivially $d_X \geq d$.

We say that X is *proper* if all closed balls are compact (or equivalently if all closed bounded sets are compact). This is a common assumption in analysis on metric spaces because it allows us to extract convergent sequences in all sorts of situations. It is very frequently true for the metrics that arise in analysis. In particular, Riemannian and Finsler manifolds are proper. A proper space is complete and locally compact, while the converse is true in a length space according to the Hopf-Rinow theorem which also says that a proper length space is a geodesic space.

Given $a \geq 1$, $h \geq 0$, an (a,h)-*quasi-isometry* f between metric spaces (X,d) and (X',d') is a map $f : X \to X'$ such that every $x' \in X'$ lies within a distance h of $f(X)$, and

$$a^{-1}d(x,y) - h \leq d'(f(x),f(y)) \leq ad(x,y) + h, \qquad x,y \in X.$$

If $a = 1$, we say that f is a h-*rough isometry*; if $a = 1$ and $h = 0$, we say that f is an *isometry*. An (a,h)-quasi-isometry f is an (r,h)-*rough similarity*, $r > 0$, $h \geq 0$, if

$$rd(x,y) - h \leq d'(f(x),f(y)) \leq rd(x,y) + h, \qquad x,y \in X.$$

A *quasigeodesic*, *rough geodesic*, or a *geodesic path* is a path $\gamma : I \to X$ in a metric space (X,d) which can be reparametrized to obtain a map from I to $(g(I),d)$ that is quasi-isometric, rough isometric, or isometric, respectively; we associate with these new concepts the same parameters a, h. An (a,h,L)-*local quasigeodesic* is a path such that all subpaths of length at most L are (a,h)-quasigeodesics. Local L-geodesics are defined similarly. The parameters a, h, and L are suppressed in all of this terminology if they are irrelevant to the matter at hand.

A *geodesic graph* is a metric space (G, d) consisting of a (not necessarily finite) set $V \subset G$ of *vertices* and a set E of *edges* where each $e \in E$ is a subset of G isometric to an interval $[0, L]$ for some $L > 0$, and the endpoints of each $e \in E$ lie in V. All edges are disjoint except at their endpoints, and distance in G is given in the obvious way by arclength. We also assume that all edges are of length at least ϵ for some fixed $\epsilon > 0$. Since d is a metric, it follows that all geodesic graphs are connected and that there is a path between any pair of vertices that involves only a finite number of edges. A *tree* is a geodesic graph G which is simply connected.

Although the above terminology is for the most part standard in the theory of metric spaces, there are some differences in the use of terminology in the special case of Riemannian manifolds. First, the term "metric" is commonly used there to refer to the infinitesimal quantity which must be integrated over paths to define arclength and hence distance. For that reason, we use the terms *infinitesimal metric* and *distance* function when discussing manifolds to make it clear which of the two we are talking about. One other difference is that "geodesic" is commonly used in a Riemannian context to refer to what we call a local geodesic; for us a geodesic always has its global meaning.

References

[1] S. Alexander and R.L. Bishop, *Warped products of Hadamard spaces*, Manuscripta Math. **96** (1998), 487–505.

[2] A.D. Alexandrov, *Über eine Verallgemeinerung der Riemannschen Geometrie*, Schriftreihe des Forschinstituts für Mathematik **1** (1957), Berlin, 33–84.

[3] J.W. Anderson, Hyperbolic geometry, 2nd edition, Springer-Verlag, London, 2005.

[4] Z.M. Balogh, *Aspects of quasiconformality and several complex variables*, Habilitationschrift, U. Bern, 1999.

[5] Z.M. Balogh and M. Bonk, *Pseudoconvexity and Gromov hyperbolicity*, C. R. Acad. Sci. Paris Sér. I Math. **328** (1999), 597–602.

[6] Z.M. Balogh and M. Bonk, *Gromov hyperbolicity and the Kobayashi metric on strictly pseudoconvex domains*, Comment. Math. Helv. **75** (2000), 504–533.

[7] Z.M. Balogh and S.M. Buckley, *Geometric characterizations of Gromov hyperbolicity*, Invent. Math. **153** (2003), 261–301.

[8] D. Bao, S.S. Chern, and Z. Shen, An introduction to Riemann-Finsler geometry, Springer-Verlag, 2000.

[9] A.F. Beardon, The geometry of discrete groups, Springer-Verlag, New York, 1983.

[10] A. Bellaïche, *The tangent space in sub-Riemannian geometry*, Progress in Math. **144**, Birkhäuser, Boston, 1996, 4–78.

[11] M. Bonk, Quasi-geodesic segments and Gromov hyperbolic spaces, *Geom. Dedicata* **62** (1996), 281–298.

[12] M. Bonk, J. Heinonen, and P. Koskela, *Uniformizing Gromov hyperbolic spaces*, Astérisque **270** (2001), 1–99.

[13] M. Bonk and O. Schramm, *Embeddings of Gromov hyperbolic spaces*, Geom. Funct. Anal. **10** (2000), 266–306.

[14] M.R. Bridson and A. Haefliger, Metric spaces of non-positive curvature, Springer-Verlag, Berlin, 1999.

[15] S.M. Buckley, K. Falk, and D.J. Wraith, *Ptolemaic spaces and CAT(0) spaces*, Glasgow Math. J. **51** (2009), 301–314.
[16] S.M. Buckley, D. Herron, and X. Xie, *metric space inversions, quasihyperbolic distance, and uniform spaces*, Indiana U. Math. J. **57** (2008), 837–890.
[17] S. Buckley and S.L. Kokkendorff, *Comparing the ideal and Floyd boundaries of a metric space*, Trans. Amer. Math. Soc. **361** (2009), 715–734.
[18] S. Buckley and S.L. Kokkendorff, *Warped products and conformal boundaries of CAT(0) spaces*, J. Geom. Anal. **18** (2008), 704–719.
[19] S. Buckley and S.L. Kokkendorff, *Detours and Gromov hyperbolicity*, Internat. J. Pure Appl. Math. **47** (2008), 313–323.
[20] S.M. Buckley, J. MacDougall, and D.J. Wraith, *On Ptolemaic metric simplicial complexes*, preprint.
[21] D. Burago, Y. Burago, and S. Ivanov, A Course in Metric Geometry, Graduate Studies in Mathematics **33**, AMS, 2001.
[22] H. Busemann, *Spaces with non-positive curvature*, Acta Math. **80** (1948), 259–310.
[23] H. Busemann, The Geometry of Geodesics, Academic Press Inc., New York, 1955.
[24] I. Chavel, Riemannian Geometry: A modern introduction, Cambridge University Press, 1993.
[25] M. Coornaert, T. Delzant, and A. Papadopoulos, Géometrie et théorie des groupes, Lecture Notes in Mathematics 1441, Springer-Verlag, Berlin, 1990.
[26] C. Fefferman, *The Bergman kernel and biholomorphic maps of pseudoconvex domains*, Invent. Math. **26** (1974), 1–65.
[27] J. Ferrand, *A characterization of quasiconformal mappings by the behavior of a function of three points*, Complex Analysis, Joensuu 1987 (Berlin), Lecture Notes in Math., no. 1351, Springer-Verlag, 1988, 110-123.
[28] W.J. Floyd, *Group completions and limit sets of Kleinian groups*, Invent. Math. **57** (1980), 205–218J.
[29] T. Foertsch and A. Karlsson, *Hilbert geometries and Minkowski norms*, J. Geometry **83**, 22–31 (2005).
[30] T. Foertsch, A. Lytchak, and V. Schroeder, *Non-positive curvature and the Ptolemy inequality*, Int. Math. Res. Not. IMRN (2007), article ID rnm100, 15 pages.
[31] E. Ghys and P. de la Harpe (Eds.), Sur les groupes hyperboliques d'aprés Mikhael Gromov, Progress in Math. 38, Birkhäuser, Boston, 1990.
[32] M. Gromov, *Hyperbolic Groups*, Essays in Group Theory, S. Gersten, Editor, MSRI Publication, Springer-Verlag, 1987, 75–265.
[33] M. Gromov, Asymptotic invariants of infinite groups, Geometric Group Theory, London Math. Soc. Lecture Notes Series **182**, 1993.
[34] M. Gromov and R. Schoen, *Harmonic maps into singular spaces and p-adic superrigidity for lattices in groups of rank one*, Inst. Hautes tudes Sci. Publ. Math. **76** (1992), 165–246.
[35] P. de la Harpe, *On Hilbert's Metric for Simplices*, Geometric Group Theory, Vol.1, (Sussex, 1991), Cambridge Univ. Press, 1993, 97–119.
[36] P. Hästö, *Gromov hyperbolicity of the j_G and $\tilde{j}_G$ metrics*, Proc. Amer. Math. Soc. **134** (2006), 1137-1142.
[37] P. Hästö, H. Lindén, A. Portilla, J.M. Rodríguez, and E. Touris, *Gromov hyperbolicity of Denjoy domains with hyperbolic and quasihyperbolic metrics*, preprint.
[38] D. Herron, Z. Ibragimov and D. Minda, *Geodesics and curvature of Möbius invariant metrics*, Rocky Mountain J. Math. **38** (2008), 891–921.
[39] M. Jarnicki and P. Pflug, Invariant distances and metrics in complex analysis, de Gruyter, Berlin, 1993.

[40] P. Kelly and E. Straus, *Curvature in Hilbert geometries*, Pacific J. Math. 8 1958 119–125.
[41] P. Koskela, *Old and new on the quasihyperbolic metric*, Quasiconformal mappings and analysis: A collection of papers honoring F.W. Gehring (New York), Springer-Verlag, 1998, 205-219.
[42] U. Lang, *Extendability of large-scale Lipschitz maps*, Trans. Amer. Math. Soc. **351** (1999), 3975–3988.
[43] U. Lang, B. Pavlović, and V. Schroeder, *Extensions of Lipschitz maps into Hadamard spaces*, Geom. Funct. Anal. **10** (2000), 1527–1553.
[44] I.J. Schoenberg, *A remark on M.M. Day's characterization of inner-product spaces and a conjecture of L.M. Blumenthal*, Proc. Amer. Math. Soc. **3** (1952), 961–964.
[45] Z. Shen, Lectures on Finsler geometry, World Scientific, 2001.
[46] É. Socié-Méthou, Comportement asymptotiques et rigidités en géométries de Hilbert, dissertation, Université de Strasbourg, 2000
(`http://www-irma.u-strasbg.fr/annexes/publications/pdf/00044.pdf`).
[47] J. Väisälä, *Gromov hyperbolic spaces*, Expo. Math. **23** (2005), 187–231.