# Alexandrov Spaces

Lecture notes

Spring term 2018

Michael Kunzinger
michael.kunzinger@univie.ac.at
http://www.mat.univie.ac.at/~mike

and

Roland Steinbauer
roland.steinbauer@univie.ac.at
http://www.mat.univie.ac.at/~stein

University of Vienna
Faculty of Mathematics
Oskar-Morgenstern-Platz 1
A-1090 Wien
AUSTRIA

Jannuary 21, 2021

# Preface

The central idea of *metric geometry* is to describe geometric concepts such as length, angles, and curvature purely in terms of metric distances. As it turns out, many notions familiar from differential geometry can indeed be captured in such 'synthetic' terms alone.

The foundational concept is that of a length space, i.e., a metric space where the metric distance between two points is given by the infimum of the length of all connecting curves. Key examples are Riemannian manifolds and polyhedra. Curvature bounds in such spaces are based on comparison with triangles in certain model spaces. E.g., the sphere has positive curvature because triangles are fatter than Euclidean triangles of the same sidelengths. Spaces with a curvature bound in this sense are called Alexandrov spaces.

Metric geometry, and, in particular, the theory of length spaces, is a vast and very active field of research that has found applications in diverse mathematical disciplines, such as differential geometry, group theory, dynamical systems and partial differential equations. It has led to identifying the 'metric core' of many results in differential geometry, to clarifying the interdependence of various concepts, and to generalizations of central notions in the field to low regularity situations.

These notes were written to accompany our lecture course in the spring term of 2018. The main source we have drawn from is the monograph [BBI01], *A course in metric geometry* by Dmitri and Yuri Burago, and Sergei Ivanov. While beeing *the* textbook on the subject it is not an easy read and we have strived to be more verbose to benefit our students. However, in general we rather closely follow its account, covering the main threads of approximately the first half of the book, that is, Chapters 1–6.

In some more detail the contents is as follows: In chapter 1 we lay the foundations by discussing in detail the notion of a length structure and by introducing the theory of length spaces. Chapter 2 is devoted to basic constructions of length spaces and introduces the main classes of examples to be used later on. The central topic of curavture bounds is introduced and extensively studied in Chapter 3. Finally, in Chapter 4, we connect Riemannian geometry to metric geometry by establishing that any Riemannian manifold is a length space.

We are greatful to our students for pointing out the inevitable lapses in earlier versions of these notes. Special thanks go to Benedict Schinnerl who revised the earlier LaTeX-file and coded many of the figures.

M.K. & R.S.
January 2021

# Contents

# Chapter 1

# Length Spaces

The general theme of this chapter is the measurement of length. To begin with, consider two points on the surface of the earth, say Vienna and Auckland, New Zealand. Then the information that the distance between these two cities is approximately 12.000 km, while technically correct, is not very useful in every day terms. In fact, this distance is just the length of a tunnel right through the middle of the earth. Similarly, the information that the two cities are approximately 18.000 km apart as measured by great circle distance is not of big practical value either. In fact, the cheaper flights at the time of writing, feature two stops at e.g. London and Chongqing, China and the actual distance to be traveled is much greater. The message of course is, that "distances" actually depend on the path taken between the point of departure and the final destination.
More abstractly, consider two points on a surface in Euclidean space. Then we can measure their Euclidean distance, but we can also introduce a new distance which is measured by the shortest path within the surface. This is the key idea of this section: A distance function in a metric space is called *(strictly) intrinsic*, if the distance between two points is given by the length of the shortest path connecting them.
In this approach the *length of paths* becomes the primary notion and we begin our endeavour by abstractly defining a minimal mathematical structure which allows formulation of such a notion.

## 1.1 Length structures and length spaces

We begin by introducing the basic notion of a *length structure*. Loosely speaking it consists of a class of so-called *admissible paths* in a topological space together with a notion of *length* associated with each admissible path that together satisfy a number of "natural" properties. Before making these notions precise we fix some basic notations and conventions.

**1.1.1 Convention (Interval, path).** By an *interval* $I$ we mean any connected subset of $\mathbb{R}$ may it be open or closed, finite or infinite. Also a single point counts as an interval. A *path* in a topological space $X$ is a continuous map $\gamma: I \to X$ defined on an interval $I$.

**1.1.2 Definition (Length structure).** *A* length structure *on a topological space $X$ is a subset $A$ of all paths in $X$, called the* admissible paths*, together with a function*

$$L: A \to [0, \infty], \tag{1.1.1}$$

*called the* length*, which satisfy the following list of properties:*

*(A1) The class $A$ is* closed under restrictions*, i.e., if $\gamma: [a,b] \to X$ is admissible and $a \le c \le d \le b$ then the restriction $\gamma|_{[c,d]}$ of $\gamma$ to $[c,d]$ is also admissible.*

*(A2) $A$ is* closed under concatenations*, that is if $\gamma: [a,b] \to X$ is a path such that the restrictions $\gamma_1 = \gamma|_{[a,c]}$ and $\gamma_2 = \gamma|_{[c,d]}$ are admissible for some $a \le c \le b$, then so is their concatenation $\gamma$.*

*(A3)* *A is* closed under linear reparametrisations, *i.e., for* $A \ni \gamma : [a,b] \to X$ *and a homeomorphism* $\varphi : [c,d] \to [a,b]$ *of the form* $\varphi(t) = \alpha t + \beta$ $(\alpha, \beta \in \mathbb{R})$ *the composition* $\gamma \circ \varphi : [c,d] \to X$ *is also an admissible path.*

*(L1)* *L* is additive*: If* $\gamma : [a,b] \to X$ *is admissible then* $L(\gamma) = L(\gamma|_{[a,c]}) + L(\gamma|_{[c,d]})$ *for any* $c \in [a,b]$.

*(L2)* *L* depends continuously on the parameter *of the path. Formally, if* $\gamma : [a,b] \to X$ *is admissible and of finite length we set* $L(\gamma, a, t) := L(\gamma|_{[a,t]})$*. We then require* $t \mapsto L(\gamma, a, t)$ *to be continuous and the length of all constant paths to vanish,* $L(t \mapsto x) = 0$ [1]

*(L3)* *L is* invariant under linear reparametrisations*, i.e., for* $\gamma, \varphi$ *as in (A3) we require* $L(\gamma \circ \varphi) = L(\gamma)$.

*(L4)* *The* length structure respects the topology of *X in the following sense: For any neighbourhood* $U_x$ *of a point* $x \in X$ *the length of paths connecting x with a point in the complement of* $U_x$ *is uniformly bounded away from zero, i.e.,*

$$\inf\{L(\gamma) : \ \gamma(a) = x, \gamma(b) \in X \setminus U_x\} > 0. \tag{1.1.2}$$

**1.1.3 Example (Admissible paths and length structures).** There is an abundant variety of natural examples of admissible paths and length structures. A simple one is the piecewise smooth paths in $\mathbb{R}^n$ with the usual length $L(\gamma) = \int |\dot{\gamma}(t)|$. (Recall that $\gamma : [a,b] \to X$ is called a piecewise smooth path, if it is continuous and there is a finite partition $a = t_0 < t_1 < \cdots < t_k = b$ of the domain such that all $\gamma|_{t_i, t_{i+1}}$ are smooth. Also observe that the smooth paths do *not* form an admissible class, since (A2) fails.) Also piecewise $C^1$-paths on $\mathbb{R}^n$ form a class of admissible paths and a length structure with the usual length.

**1.1.4 Remark (On reparametrisations).** Every class of paths comes with its own 'natural' class of reparametrisations, e.g. smooth paths in $\mathbb{R}^n$ with the class of diffeomorphisms. Note, however, that in (A3) and (L3) we only require compatibility with the class of linear reparametrisations. Hence if we deal with such a natural class it is sufficient to know that the natural class of reparametrisations includes all linear ones.

**1.1.5 Notation (Length).** We will often use the notation $L(\gamma, a, b)$ introduced in (L2). Explicitly, if $\gamma : I \to X$ is admissible and $a \le b \in I$ then $L(\gamma, a, b)$ denotes the length of the restriction $\gamma|_{[a,b]}$, i.e., $L(\gamma, a, b) = L(\gamma|_{[a,b]})$.
We will also set $L(\gamma, b, a) = -L(\gamma, a, b)$. With these conventions we have for all $a, b, c \in I$ that $L(\gamma, a, b) = L(\gamma, a, c) + L(\gamma, c, b)$.

Given the notion of a length structure we can define a metric on $X$ associated with this structure, the so-called *length metric*. Note that following [BBI01] we generally allow metrics to take on the value $\infty$. To avoid trivialities it is reasonable to assume that the topology of $X$ is Hausdorff (see 1.1.7 below) and we will henceforth do so.

**1.1.6 Definition (Length metric).** *Let* $(X, A, L)$ *be a length structure with X Hausdorff and let* $x, y \in X$*. We define the* associated distance *or* length metric *to be the infimum of lengths of admissible paths connecting x and y, i.e.,*

$$d_L(x,y) := \inf\{L(\gamma) : \ A \ni \gamma : [a,b] \to X, \gamma(a) = x, \gamma(b) = y\}. \tag{1.1.3}$$

If it is clear from the context which length we are dealing with, we will often drop the subscript in $d_L$ and simply write $d$.

[1] Note that, if we already knew a constant path $\gamma$ to be of finite length, then by (L1) $L(\gamma, a, a) = L(\gamma, a, a) + L(\gamma, a, a)$ and hence $L(\gamma, a, a) = 0$.

**1.1.7 Lemma (Length metric).** *Let $(X, A, L)$ be a length structure with $X$ Hausdorff and with length metric $d_L$. Then $(X, d_L)$ is a metric space.*

**Proof.** We simply check the metric properties.
First $d_L$ is *positive definite* since for $x \in X$ we have $0 \leq d_L(x,x) = \inf\{L(\gamma) : A \ni \gamma : [a,b] \to X, \gamma(a) = x, \gamma(b) = x\} \leq L(t \mapsto x) = 0$. Moreover, if $d_L(x,y) = 0$ but $x \neq y$, then by the Hausdorff property there are disjoint neigbourhoods $U_x$, $U_y$ of $x, y$. Now every path connecting $x$ and $y$ has to leave $U_x$ and by (L4) the infimum of the length of all such paths is non-vanishing[2].
The *symmetry* of $d_L$ is obvious (by (L3)).
Finally the *triangle inequality* holds since for $x, y, z \in X$ we have (in an informal notation, see Figure 1.1)

$$\begin{aligned} d_L(x,z) &= \inf\{L(\gamma) : \gamma \text{ connects } x \text{ with } z\} \leq \inf\{L(\gamma) : \gamma \text{ connects } x \text{ with } z \text{ via } y\} \\ &= \inf\{L(\gamma_1) + L(\gamma_2) : \gamma_1 \text{ connects } x \text{ with } y \text{ and } \gamma_2 \text{ connects } y \text{ with } z\} \\ &= \inf\{L(\gamma_1) : \gamma_1 \text{ connects } x \text{ with } y\} + \inf\{L(\gamma_2) : \gamma_2 \text{ connects } y \text{ with } z\} \\ &= d_L(x,y) + d_L(y,z)\,. \end{aligned}$$

□

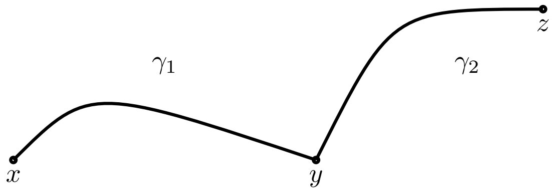



Figure 1.1: paths connecting $x$ and $z$ via $y$

In the following we will be concerned with properties of metric spaces that arise via the above construction from a length structure. Let's give it a name.

**1.1.8 Definition (Length space & metric).** *A metric space $(X, d)$ is called a* length space *if the metric $d$ can be obtained as the length metric $d_L$ of a length structure $(X, A, L)$. In this case we call $d$* intrinsic *or* length metric.

**1.1.9 Remark ($d_L$ may be infinite).** Note that, if $X$ possesses two connected components there will be no continuous paths between points in different components thus the set in the infimum is empty and we use the convention $\inf\{\emptyset\} = \infty$. Also there might be points in $X$ that can be joined by continuous paths but they are all of infinite length. Hence the following notion turns out to be useful.

**1.1.10 Definition (Accessibility component).** *Let $(X, A, L)$ be a length structure. We say two points $x, y \in X$ belong to the same accessibility component if they can be connected by an admissible path of finite length.*

Before deriving the basic properties of accessibility components we prove the following technical but important results.

**1.1.11 Lemma (Continuity of admissible paths).** *In a length structure $(X, A, L)$ any admissible path of finite length is continuous also with respect to the length metric $d_L$.*

[2] Observe that is is actually sufficient for this argument to suppose the topology of $X$ to satisfy the separation axiom $T_0$.

**Proof.** Let $\gamma : [a, b] \to X$ be admissible with $L(\gamma) < \infty$. We show that $\gamma$ is continuous in $t_0 \in [a, b]$ w.r.t. $d_L$. Let $t \in [a, b]$ then

$$d_L(\gamma(t_0), \gamma(t)) \leq L(\gamma, t_0, t) = |L(\gamma, a, t_0) - L(\gamma, a, t)|. \tag{1.1.4}$$

Now by (L2), $L(\gamma, a, t) \to L(\gamma, a, t_0)$ for $t \to t_0$ and hence the claim follows. □

**1.1.12 Lemma (Local path connectedness).** *Any length space is locally path connected, that is, every neighbourhood of any point contains a path connected neighbourhood (i.e., a neighbourhood such that any two points can be connected by a path in this neighbourhood).*

**Proof.** Let $(X, d = d_L)$ be a length space. We show that any open ball $B_\eta(x)$ is path connected. Indeed for any $y \in B_\eta(x)$ we have $d_L(x, y) =: \delta < \eta$. Then there exists an admissible $\gamma_{xy}$ from $x$ to $y$ of length $L(\gamma_{xy}) = \delta + \varepsilon$, where $\varepsilon$ can be chosen such that $\delta + \varepsilon < \eta$. Now by Lemma 1.1.11 $\gamma_{xy}$ is continuous w.r.t. $d_L$ and it is entirely contained in $B_\eta(x)$. □

**1.1.13 Lemma (Accessibility).** *Let $X$ be a length space. Then we have*

*(i) Accessibility by paths is an equivalence relation.*

*(ii) Accessibility components coincide with components of finiteness of $d_L$.*

*(iii) Accessibility components are contained in path connected components (which agree with connected components) but in general are not equal to them.*[3]

**Proof.** (i): By definition $x \sim y$, if there is an admissible finite path $\gamma_{xy}$ from $x$ to $y$. It is easy to see that this actually is an equivalence relation: Reflexivity is clear and so is symmetry (using (A3) and (L3)). Finally transitivity is immediate from (A2) and (L1).

(ii) is clear.

(iii) To begin with, recall that any locally path connected topological space is connected iff it is path connected (see, e.g. [Wil04, Thms. 27.2, 27.5]).
Now if $y$ is in the accessibility component of $x$, then there is an admissible path $\gamma_{xy}$ connecting $x$ with $y$ which is of finite length. By Lemma 1.1.11, $\gamma_{xy}$ is continuous also w.r.t. $d_L$ and so $y$ is in the path connected component of $x$. □

**1.1.14 Remark (Accessibility components vs. path connected components).** In general path connected components are larger than accessibility components: If a point $y$ is in the path connected component of $x$ then there is a continuous (in $(X, d)$) path $\gamma_{xy}$ connecting these two points. Although $\gamma_{xy}$ will in general be continuous w.r.t. the topology of $X$, see Lemma 1.1.15, below, it need not be of finite length! As an example consider $X$ to be the union of the graph of the function $f(x) = x \sin(1/x)$ $(x > 0)$ with $(0, 0)$ with the trace topology of $\mathbb{R}^2$. Then any point $y$ on the graph is in the path connected component of $(0, 0)$ but every connecting path is of infinite length, see Figure 1.2.

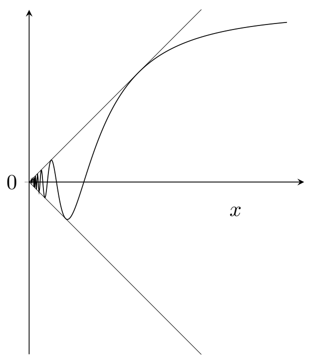



Figure 1.2: The graph of $x \sin\left(\frac{1}{x}\right)$.

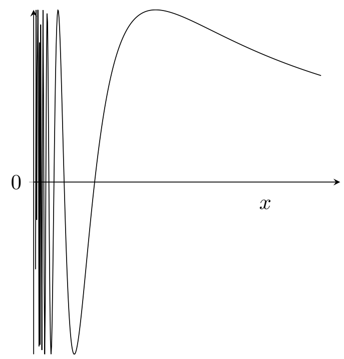



Figure 1.3: The graph of $\sin\left(\frac{1}{x}\right)$.

[3] contrary to the respective statement in [BBI01, Ex. 2.1.3(3)], see also Remark 1.1.14, below.

Observe that in general the topology induced by the length metric $d_L$ on $X$ will be different from the original Hausdorff topology on $X$. A number of respective examples will appear later. However, the topology induced by the length metric is always the finer one.

**1.1.15 Lemma ($d_L$ is finer).** *Let $(X, A, L)$ be a length structure with length metric $d_L$. Then any open set in $X$ (w.r.t. its original topology) is also open in $(X, d_L)$. Hence the topology induced by $d_L$ is finer than the original topology on $X$.*

**Proof.** Let $U$ be open in $X$, $x \in U$ and set $\varepsilon_x := \inf\{d_L(x, y) : y \notin U\}$. Then by (L4) $\varepsilon_x > 0$ and since $\varepsilon_x = d_L(x, X \setminus U)$ we have that $B_{\varepsilon_x}(x) \subseteq U$ so that $U$ is open in $(X, d_L)$. □

It is a fact that not every metric space is a length space. A simple example is $\mathbb{R}^2$ with the unit disc removed, see Figure 1.4. Hence even if $(X, d)$ is a length space and $Y \subseteq X$ then $(Y, d|_Y)$ need not be a length space. An example is the unit circle in $\mathbb{R}^2$: while $\mathbb{R}^2$ with the Euclidean metric is a length space, restricting this metric to the unit circle does not yield a length metric, see Figure, 1.5.

Another relevant obstruction is given by Lemma 1.1.12. In fact, if a metric space is not locally path connected, its metric cannot be induced by *any* length structure. This is the reason underlying the following examples.

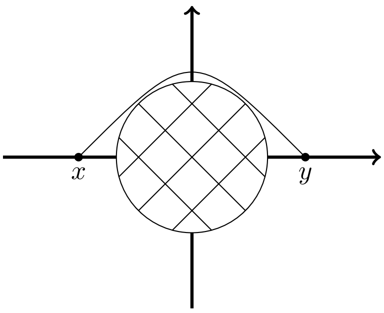



Figure 1.4: $R^2 \setminus B_1(0)$ with the induced metric of $\mathbb{R}^2$ is not a length space. The infimum of the length of paths connecting e.g. $x = (-1.5, 0)$ and $y = (1.5, 0)$ is not $d(x, y) = 3$.

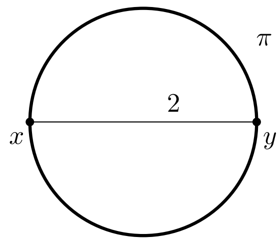



Figure 1.5: The unit circle with the metric induced from $\mathbb{R}^2$ is not a length space. The distance of e.g. antipodal points is 2 whereas any connecting path has at least length $\pi$.

**1.1.16 Example (Not a length space).**

(i) The set of rational numbers $\mathbb{Q} \subseteq \mathbb{R}$ is not homeomorphic to a length space.

(ii) The union of the $y$-axis in $\mathbb{R}^2$ with the graph of the function $f(x) = \sin(1/x)$ $(x > 0)$ with the trace topology of $\mathbb{R}^2$ is not homeomorphic to a length space, see Figure 1.3. In fact, this space is a standard example of a space which is connected but not (locally) path connected.

(iii) The 'fan', i.e., the union of segments $X = \bigcup_{n=1}^{\infty}[(0,0), (\cos(1/n), \sin(1/n))] \cup [(0,0), (1,0)]$ with the induced topology of $\mathbb{R}^2$ is not a length space, see Figure 1.6. The distance of the endpoints of the segments and the point $(1, 0)$ goes to zero while the length of any connecting curve is at least 2. Moreover, $X$ is not locally path connected, hence not homeomorphic to any length space.

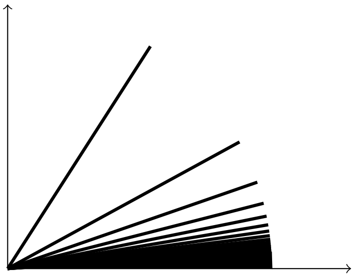


Figure 1.6: The 'fan' is not a length space.

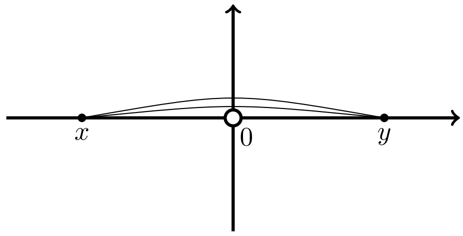



Figure 1.7: $\mathbb{R}^2$ with the origin removed gives an incomplete length structure.

Finally we remark that it is essential to use the infimum in Definition 1.1.6 rather than the minimum: there need not exist a shortest path between two arbitrary points. Just consider the plane with the origin removed. This is obviously a length space, but there is no shortest path connecting e.g. any point on the positive $x$-axis to any point on the negative $x$-axis. However, it is still possible to approximate the intrinsic distance of such points with arbitrary precision by the length of curves connecting them, see Figure 1.7. (Note that this is the essential feature distinguishing this example from the one with the whole unit disc removed from $\mathbb{R}^2$ above, which does not even yield a length space.)

However, in many applications such a nuisance will be ruled out by some completeness or compactness condition. Hence it is often useful to consider *complete* length structures in the following sense.

**1.1.17 Definition (complete length structure).** *A length structure* $(X, A, L)$ *is called* complete*, if for any pair of points* $x, y \in X$ *there is a path* $\gamma_{xy} \in A$ *connecting* $x$ *and* $y$ *with* $L(\gamma_{xy}) = d_L(x, y)$*.*

In other words a length structure is complete, if there is an admissible shortest path between any pair of its points. An intrinsic metric associated with a complete length structure is often called *strictly intrinsic*. A length space induced by a complete length structure is often called a *geodesic space* and shortest paths are called *geodesics*. These notions will be of great importance later.

In the following we list several key examples of length structures and intrinsic metrics to give an impression how wide the zoo of examples actually is.

**1.1.18 Example (Key length structures, 1).**

(i) *Driving in Manhattan.* We consider the plane $\mathbb{R}^2$ with the usual length of paths but we only consider paths admissible if they are broken lines with all edges parallel to the coordinate axes, see Figure 1.8. A ball in the induced metric is a diamond.

(ii) *Living on an island.* We consider a connected subset $U$ of $\mathbb{R}^2$ and again the usual length of paths which we consider to be admissible if they are piecewise smooth. Then if $U$ is convex, this length structure induces the Euclidean distance. However, if $U$ is not convex the resulting length structure gives a distance which corresponds to the one measured by an individual living on the island $U$ and which cannot swim. Balls w.r.t. the intrinsic metric have a distorted shape, see Figure 1.9. This metric is not intrinsic unless $U$ is closed.

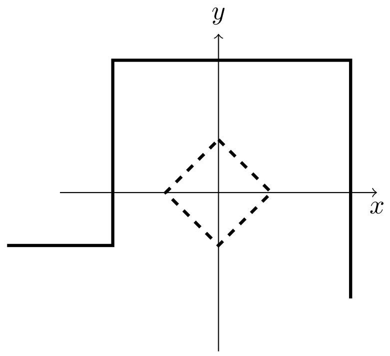



Figure 1.8: Driving in Manhattan: Admissible paths are parallel to the coordinate axes. The (open) unit ball is diamond-shaped.

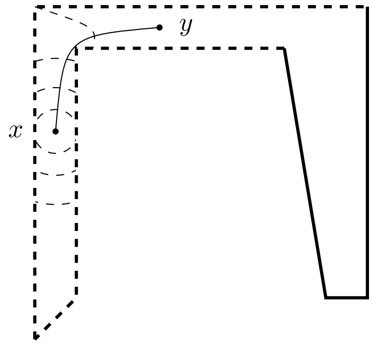



Figure 1.9: Living on an island: The distance balls from $x$ are distorted and there is no shortest path between $x$ and $y$.

(iii) *Crossing a swamp, conformal length.* We again consider the Euclidean plane $\mathbb{R}^2$ with all piecewise smooth paths. Given a positive continuous function $f$ on $\mathbb{R}^2$ we define the length via

$$L(\gamma) = \int_a^b f(\gamma(t))\,|\gamma'(t)|\,dt \tag{1.1.5}$$

The corresponding length structure can be thought of as a weighted Euclidean distance. Intuitively a traveler would assign a high value to $f$ in a region which is hard to traverse, e.g. a swamp. This is the first example of a Riemannian length structure and, in particular, a so-called conformally flat one and hence the title.

(iv) *Finslerian length.* We consider a modification of the above example by allowing $f$ to also depend on the velocity resp. direction of the traveler and define the length by

$$L(\gamma) = \int_a^b f(\gamma(t), \gamma'(t))\,dt \tag{1.1.6}$$

To have the length parametrisation invariant we have to ask for

$$f(x, kv) = |k| f(x, v) \qquad \text{for all } x, v \in \mathbb{R}^2 \text{ and } k \in \mathbb{R}. \tag{1.1.7}$$

A somewhat stronger condition is that at every $x \in \mathbb{R}^2$ the function $f(x, .)$ is a norm on $\mathbb{R}^2$ in which case one speaks of a Finslerian length.

**1.1.19 Remark (Uphill vs. downhill).** We remark the following fact just to set straight our intuition of the traveler measuring length used in the above examples. One feature of such a scenario *cannot* be built into a length structure, namely that is easier to walk downhill than it is to walk uphill (or vice versa, depending on the condition of your knees). In fact distance in a metric space must be symmetric!

We discuss several further examples in a quite informal way. These are nevertheless key examples and we are going to discuss them in a more precise fashion later on—meanwhile, we want to guide the intuition of the reader.

**1.1.20 Example (Key length structures, 2).**

(i) *Induced length structure.* Consider topological Hausdorff spaces $X, Y$ with $Y$ endowed with a length structure and $f : X \to Y$ a continuous function. Then one may define a length structure on $X$ induced by the one of $Y$. Indeed admissible paths $\gamma$ in $X$ are defined as those for which $f \circ \gamma$ is admissible in $Y$ and the length in $X$ is defined to be the length in $Y$ of the composition. However, condition (L4) is not automatic and we only use the term induced length structure, if it indeed holds.

Observe that an induced length structure on $X$ can be dramatically different from say a metric which was used to define the topology on $X$. A historically important key example is that of a surface in $\mathbb{R}^3$. We also mention the following fact which is surprising and not easy to prove: Every Riemannian length structure on $\mathbb{R}^n$ can be induced by a map $f : \mathbb{R}^n \to \mathbb{R}^n$ (which is *not* smooth).

(ii) *Cobweb.* Begin with several disjoint line segments in Euclidean space and glue (some of) their endpoints together. You obtain the picture of a cobweb. It is natural to consider the length structure given by all (continuous) paths. The length of each segment is just the Euclidean length. And the length of a path is defined to be the sum of the (countable many) restrictions to intervals such that the image is contained in one of the segments only, see figure 1.10. This is an example of so-called *metric graph* and the construction is a particular case of *gluing* which we will discuss further in 2.1.2

(iii) Notebook. A similar example can be constructed by 'gluing' several closed half-spaces in Euclidean space along their boundary lines. This gives rise to a so-called *polyhedral length space* (see section 2.2) and can be visualised like an open notebook.

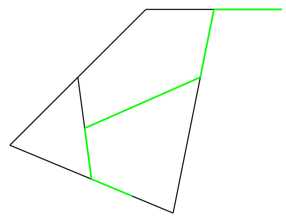


Figure 1.10: Length in a cobweb: The length of the green path is the summed up length of the green segments.

## 1.2 Length structures induced by metrics

In this section we will consider a special class of examples that have already frequently occurred above: Those in which the original topology on $X$ on which the length structure is built upon is given by a metric. Indeed this class contains some of the most important length spaces of all.
In this case the class of admissible paths can be taken to be just all (continuous) paths. Sometimes it is, however, preferable to use Lipschitz paths which we define now.

**1.2.1 Definition (Lipschitz maps & paths).** *A map $f : X \to Y$ between metric spaces is called* Lipschitz continuous *with Lipschitz constant $C(> 0)$ if*

$$d_Y(f(x_1), f(x_2)) \leq C\, d_X(x_1, x_2) \quad \textit{for all } x_1, x_2 \in X. \tag{1.2.1}$$

*In particular, a* Lipschitz path *$\gamma : I \to X$ satisfies $d(\gamma(t), \gamma(s)) \leq C|t - s|$ for all $s, t \in I$.*

### 1.2.1 Length of curves in metric spaces

To begin with we define the length of a path in a metric space. Observe that the following is a direct generalization of the notion of length of paths in Euclidean space, which is given in terms of the length of inscribed polygons.

**1.2.2 Definition (Variational length).** *Let $\gamma : [a,b] \to X$ be a path in a metric space. By a* partition *of $[a,b]$ we mean a finite set of points $\sigma = \{t_0, \dots, t_n\}$ such that $a = t_0 < t_1 < \cdots < t_n = b$. We define the* total variation *of $\gamma$ with respect to a partition $\sigma$ by*

$$V_\sigma(\gamma) := \sum_{i=0}^{n-1} d(\gamma(t_i), \gamma(t_{i+1})). \tag{1.2.2}$$

*The* (variational) length *of $\gamma$ is defined as the supremum over all total variations, i.e.,*

$$L_d(\gamma) = \sup_\sigma V_\sigma(\gamma) = \sup \left\{ \sum_{i=0}^{n-1} d(\gamma(t_i), \gamma(t_{i+1})) : \begin{array}{l} \sigma = \{t_0, \dots, t_n\} \text{ a partition} \\ \text{of } [a,b], \text{ and } n \text{ in } \mathbb{N} \end{array} \right\}. \tag{1.2.3}$$

*Finally $\gamma$ is said to be* rectifiable *if $L_d(\gamma)$ is finite.*

The idea behind the definition of the variational length, i.e., approximation of a path $\gamma$ by polygons is illustrated in Figure 1.11. We note the following simple fact, which is immediate from the triangle inequality: For a path $\gamma : [a,b] \to X$ the length $L_d(\gamma)$ is always bounded below by the distance of its endpoints $d(\gamma(a), \gamma(b))$ which is the length of the 'zeroth approximation'.

**1.2.3 Lemma (Generalized triangle inequality).** *The variational length of a path $\gamma$ in a metric space $(X,d)$ satisfies*

$$L_d(\gamma) \geq d(\gamma(a), \gamma(b)). \tag{1.2.4}$$

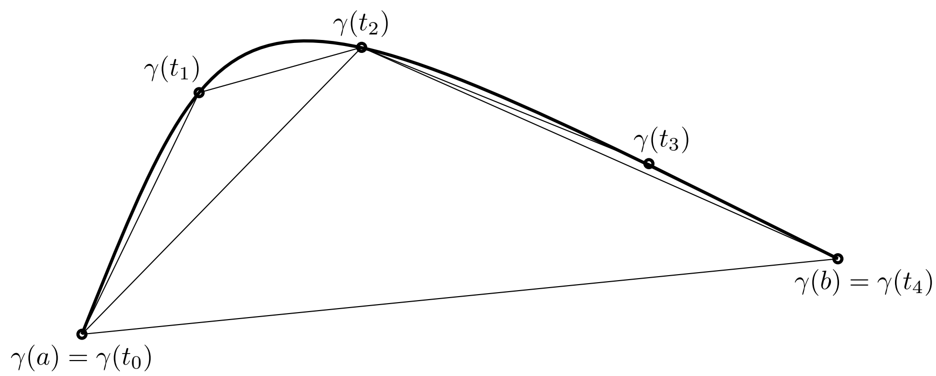



Figure 1.11: The idea behind the variational length is approximation by polygons.

We also have the following useful observation.

**1.2.4 Lemma (Constant paths & zero length).** *The length of a path $\gamma : [a,b] \to X$ vanishes iff $\gamma$ is a constant path, (i.e., iff there is $x \in X$ such that $\gamma(t) = x$ for all $t \in [a,b]$).*

**Proof.** If $\gamma$ is constant then $V_\sigma(\gamma) = 0$ for any partition and so the length vanishes.
Conversely suppose that $L(\gamma) = 0$. Choose some $t \in (a,b)$ and consider the partition $\sigma = \{a, t, b\}$. Then

$$V_\sigma(\gamma) = d(\gamma(a), \gamma(t)) + d(\gamma(t), \gamma(b)) \leq L(\gamma) = 0. \tag{1.2.5}$$

So $d(\gamma(a), \gamma(t)) = 0$ hence $\gamma(t) = \gamma(a)$ for all $t \in (a,b)$, which gives that $\gamma$ is constant. □

While the above definition makes sense for any metric space it only gives something really useful if $(X,d)$ is chosen appropriately. In fact, if $X$ is a discrete space then any path in $X$ is constant with zero length.

Also there exist *non-rectifiable* curves in nice spaces as e.g. $\gamma : [0,1] \to \mathbb{R}$ with $\gamma(t) = x\sin(1/x)$ for $x \neq 0$ and $\gamma(0) = 0$, see Figure 1.2 above, or Koch's curve in $\mathbb{R}^2$, see e.g. [Pap14, Ex. 1.1.11].

As in the Euclidean case the variational length actually is the limit of the total variation as the edges of the polygons approach each other. To formalise this we call the number

$$|\sigma| := \max_{1\leq i\leq n-1} |t_i - t_{i+1}| \tag{1.2.6}$$

the *modulus* of the partition $\sigma = \{t_0, \ldots, t_n\}$. We then have the following statement.

**1.2.5 Lemma (Variational length as a limit).** *For every path in a metric space $\gamma : [a,b] \to X$ the length is given by*

$$L_d(\gamma) = \lim_{|\sigma|\to 0} V_\sigma(\gamma). \tag{1.2.7}$$

**Proof.** Let $M < L_d(\gamma)$ (which we have not supposed to be finite). It suffices to show that there is $\eta > 0$ such that for all $\sigma$ with $|\sigma| < \eta$ we have $M \leq V_\sigma(\gamma) \leq L_d(\gamma)$. Since the right inequality always holds we only have to prove the left one.
To this end let $\varepsilon > 0$ such that $M + \varepsilon < L_d(\gamma)$. By definition of $L$ there is a partition $\tau = \{t_0, \ldots, t_n\}$ of $[a,b]$ with $V_\tau(\gamma) > M + \varepsilon$. Moreover since $\gamma$ is uniformly continuous on $[a,b]$ there is $0 < \eta < (1/4)\inf(t_{i+1} - t_i)$ such that

$$d(\gamma(r), \gamma(s)) \leq \frac{\varepsilon}{2(n-1)} \quad \text{provided that } |r - s| \leq \eta. \tag{1.2.8}$$

Let now $\sigma$ be a partition of $[a,b]$ with $|\sigma| \leq \eta$ and define for all $1 \leq i \leq n-1$

- $t_i'$ to be the vertex of $\sigma$ closest to $t_i$ with $t_i' \leq t_i$, and
- $t_i''$ to be the vertex of $\sigma$ closest to $t_i$ with $t_i'' > t_i$.

Now by the choice of $|\sigma|$ we have

$$t_i' \leq t_i < t_i'' < t_{i+1}' \leq t_{i+1} < t_{i+1}'', \tag{1.2.9}$$

see Figure 1.12. Then since 'between' $t_i''$ and $t_{i+1}'$ the two partitions $\sigma \cup \tau$ and $\sigma$ coincide we have

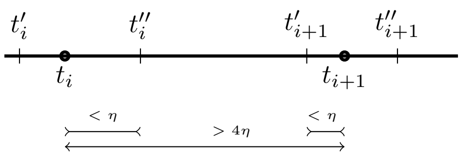



Figure 1.12: The choice of $t_i$, $t_i'$ and $t_i''$.

$$\begin{aligned} V_{\sigma\cup\tau}(\gamma) - V_\sigma(\gamma) &= \sum_{i=1}^{n-1} \Big( d\big(\gamma(t_i'), \gamma(t_i)\big) + d\big(\gamma(t_i), \gamma(t_i'')\big) - d\big(\gamma(t_i'), \gamma(t_i'')\big)\Big) \\ &\leq \sum_{i=1}^{n-1} \Big( d\big(\gamma(t_i), \gamma(t_i'')\big) + d\big(\gamma(t_i), \gamma(t_i'')\big)\Big) \\ &\leq 2(n-1)\, \frac{\varepsilon}{2(n-1)} = \varepsilon\,, \end{aligned} \tag{1.2.10}$$

by (1.2.8) and hence $V_{\sigma\cup\tau}(\gamma) \leq V_\sigma(\gamma) + \varepsilon$.
Now since $\tau \subseteq \sigma \cup \tau$ we have by the triangle inequality $V_\tau(\gamma) \leq V_{\sigma\cup\tau}(\gamma)$ and so

$$M + \varepsilon \leq V_\tau(\gamma) \leq V_{\sigma\cup\tau}(\gamma) \leq V_\sigma(\gamma) + \varepsilon, \tag{1.2.11}$$

which gives $M \leq V_\sigma(\gamma)$ as desired. □

Next we note the fact that the variational length of differentiable curves in Euclidean space is also compatible with the usual length, i.e., the formula saying that 'length is the integral of the speed'. More precisely we have.

**1.2.6 Proposition (Length as integral of speed).** *Let* $(V, |\ |)$ *be a finite-dimensional normed vector space and let* $\gamma : [a, b] \to V$ *be a differentiable map. Then the variational length is given by*

$$L_d(\gamma) = \int_a^b |\gamma'(t)|\, dt. \tag{1.2.12}$$

We will not prove this statement, since we will later deal with a more general situation in section 4.1 and in particular in theorem 4.1.2. A direct proof can be found in any textbook on elementary differential geometry, see e.g. [Bär10, Prop. 2.1.18], and in [Pap14, Prop. 1.3.1].

## 1.2.2 Length structures induced by metric spaces

The variational length introduced above for (continuous) paths in metric spaces gives rise to a length structure. In this section we are going to investigate the properties of this structure which we now make explicit.
To begin with we specify the class of reparametrizations we are going to use. For reasons that will be discussed below in Section 1.4.1 we will not use the seemingly natural class of homeomorphisms, but rather the more general one of (not necessarily strictly) monotonous surjective maps. We will also prove invariance of the variational length under these reparametrizations although that would not strictly be needed at this stage.

**1.2.7 Definition & Lemma (Length structure induced by a metric).** *Let* $(X, d)$ *be a metric space. Then the class of all (continuous) paths* $C$ *in* $X$ *with (monotonous surjective reparametrisations and) the variational length* $L_d$ *defines a length structure* $(X, C, L_d)$ *which we call the length structure* induced by $d$.

**Proof.** We briefly check that all the required properties hold. Clearly (A1)–(A3) hold for $C$.
(L1): We show that for all $c \in [a, b]$ we have $L_d(\gamma) = L_d(\gamma|_{[a,c]}) + L_d(\gamma|_{[c,b]})$. Note that by the triangle inequality we have for any partition $\sigma$ of $[a, b]$ that $V_{\sigma\cup\{c\}}(\gamma) \geq V_\sigma(\gamma)$ and hence

$$\begin{aligned} L_d(\gamma) &= \sup_\sigma V_\sigma(\gamma) = \sup_\sigma V_{\sigma\cup\{c\}}(\gamma) \\ &= \sup_\sigma \Big( V_{\sigma\cup\{c\}\cap[a,c]}(\gamma|_{[a.c]}) + V_{\sigma\cup\{c\}\cap[c,b]}(\gamma|_{[c.b]}) \Big) \\ &= \sup_\sigma \Big( V_{\sigma\cup\{c\}\cap[a,c]}(\gamma|_{[a.c]}) \Big) + \sup_\sigma \Big( V_{\sigma\cup\{c\}\cap[c,b]}(\gamma|_{[c.b]}) \Big) = L_d(\gamma|_{[a,c]}) + L_d(\gamma|_{[c,b]}). \end{aligned} \tag{1.2.13}$$

(L2): We even show that for a rectifiable path $\gamma$, $L_d(\gamma|_{[r,s]}) = L_d(\gamma, r, s)$ is continuous w.r.t. $r, s$ (with $a \leq r < s \leq b$). Since $\gamma$ is rectifiable we may choose a partition $\sigma$ with $L_d(\gamma) - V_\sigma(\gamma) < \varepsilon/2$ and suppose w.l.o.g. that $r, s \in \sigma$. But then by splitting the length up as in (1.2.13), one obtains $L_d(\gamma, r, s) - d(\gamma(r), \gamma(s)) < \varepsilon/2$. Further by uniform continuity of $\gamma$ we have $d(\gamma(r), \gamma(s)) < \varepsilon/2$ provided $|r - s|$ is small enough. Hence in this case $L_d(\gamma, r, s) < \varepsilon$ and we are done.
(L3): Let $\psi : [c, d] \to [a, b]$ be monotonous and surjective and set $\gamma' = \gamma \circ \psi$. Any partition $\sigma'$ of $[c, d]$ gives rise to a partition $\sigma$ of $[a, b]$ (via $t_i = \psi(t'_i)$) with $V_{\sigma'}(\gamma') = V_\sigma(\gamma)$ (since for $t'_i \in \sigma'$ we have $\gamma'(t'_i) = \gamma(\psi(t'_i))$. Now $V_{\sigma'}(\gamma') \leq L_d(\gamma)$ and by taking supremum over $\sigma'$ we obtain $L_d(\gamma') \leq L_d(\gamma)$.
For the converse inequality[4] note that for any partition $\sigma = \{t_0 < \cdots < t_n\}$ of $[a, b]$ we may associate a partition $\sigma'$ of $[c, d]$ by choosing $t'_i$ for all $i$ arbitrarily from $\psi^{-1}(t_i)$. Then again $V_\sigma(\gamma) = V_{\sigma'}(\gamma')$ (since $\gamma'(t'_i) = \gamma(\psi(t'_i) = \gamma(t_i))$. Now we may proceed as above.
(L4) is immediate from the generalized triangle inequality, Lemma 1.2.3. □

From now on any metric space $(X, d)$ will always be equipped with the length structure from Definition 1.2.7, unless explicitly stated otherwise.
The variational length has the following remarkable property.

[4]This argument is only needed for this general class of reparametrisations. If we take homeomorphisms instead, the above argument is symmetric.

**1.2.8 Proposition (Lower semi-continuity of the length).** *Let $(X, C, L_d)$ be a length structure induced by the metric space $(X, d)$. If a sequence of paths $\gamma_k : [a, b] \to M$ converges pointwise to a path $\gamma : [a, b] \to X$ then*

$$L_d(\gamma) \le \liminf L_d(\gamma_k). \tag{1.2.14}$$

*In particular, the length $L_d$ is a lower semi-continuous[5] functional on the space of paths from $[a, b]$ to $X$ w.r.t. uniform convergence.*

**Proof.** Let $\gamma$, $\gamma_k$ be as in the statement. First note that for any partition $\sigma = \{a = t_0, \dots, t_n = b\}$ the variations of the $\gamma_k$ converge, i.e., $V_\sigma(\gamma_k) \to V_\sigma(\gamma)$: Indeed we have

$$\begin{aligned} V_\sigma(\gamma_k) - V_\sigma(\gamma) &= \sum_{i=0}^{n-1} \Big( d\big(\gamma_k(t_i), \gamma_k(t_{i+1})\big) - d\big(\gamma(t_i), \gamma(t_{i+1})\big) \Big) \\ &\le \sum_{i=0}^{n-1} \Big( d\big(\gamma_k(t_i), \gamma(t_i)\big) + d\big(\gamma(t_i), \gamma(t_{i+1})\big) + d\big(\gamma(t_{i+1}), \gamma_k(t_{i+1})\big) - d\big(\gamma(t_i), \gamma(t_{i+1})\big) \Big) \\ &= \sum_{i=0}^{n-1} \Big( d\big(\gamma_k(t_i), \gamma(t_i)\big) + d\big(\gamma(t_{i+1}), \gamma_k(t_{i+1})\big) \Big) \ \to\ 0\,. \end{aligned} \tag{1.2.15}$$

Now suppose $L(\gamma) < \infty$, fix $\varepsilon > 0$ and choose $\sigma$ such that $L(\gamma) - \varepsilon < V_\sigma(\gamma)$. By the above there is $k_0$ such that for all $k \ge k_0$

$$L(\gamma) - \varepsilon < V_\sigma(\gamma_k) \le L(\gamma_k)\,, \tag{1.2.16}$$

and so

$$L(\gamma) - \varepsilon \le \inf_{k \ge k_0} L(\gamma_k) \le \sup_{k_0} \inf_{k \ge k_0} L(\gamma_k) = \liminf_{k \to \infty} L(\gamma_k)\,. \tag{1.2.17}$$

Since $\varepsilon$ was arbitrary, (1.2.14) follows.
In case $\gamma$ is non-rectifiable[6] we replace $L(\gamma) - \varepsilon$ in the above argument by an arbitrary large number $N$ and obtain $\liminf_{k\to\infty} L(\gamma_k) = \infty$.
Finally if $\gamma_k \to \gamma$ uniformly then $\gamma$ is a path and the above arguments apply, in particular the space of paths with the topology of uniform convergence is a metric space, so 1.2.14 implies that the length is lower semi-continuous. □

**1.2.9 Remark (The length is not continuous).** A simple counterexample to the continuity of the length is given by stairs-like paths in $\mathbb{R}^2$, see Figure 1.13. The intuition behind the semi-continuity from below is that one cannot approximate a given path by ones which are uniformly shorter.

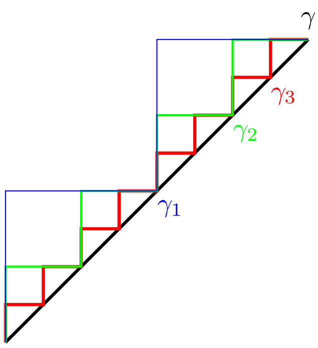



Figure 1.13: The length is not continuous: The $\gamma_k$ have all length 2 and converge uniformly to $\gamma$ which has length $\sqrt{2}$.

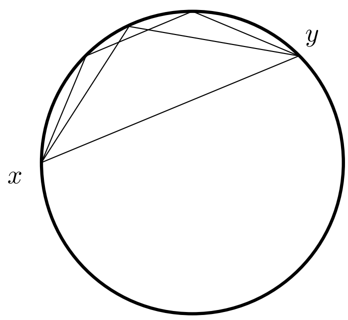



Figure 1.14: The induced intrinsic metric on $\mathbb{S}^1$: The variational length $L_d$ is the length of the arc and so $\hat{d}$ is just the (smaller) angle between two points.

[5] Recall that a map $f : X \to \mathbb{R} \cup \{\infty\}$ on a metric space is called *lower semi-continuous* at $x \in X$ if for any sequence $x_n \to x$ we have $f(x) \le \liminf f(x_n)$, cf. e.g. [Pap14, p. 25f]

[6] Observe that the proof in [BBI01, 2.3.4(iv)] fails in case of infinite length of the limiting path.

### 1.2.3 The induced intrinsic metric

A metric space $(X, d)$ induces a length structure $(X, C, L_d)$ by 1.2.7, with $L_d$ the variational length of 1.2.2. By 1.1.6 this length structure induces a length metric $d_{L_d}$ on $X$, which gives rise to a length space. So we obtain a natural construction of an induced intrinsic metric on $X$

$$(X, d) \quad \rightsquigarrow \quad (X, C, L_d) \quad \rightsquigarrow \quad (X, d_{L_d} = \hat{d})\,. \tag{1.2.18}$$

We first take a look at some key examples.

**1.2.10 Example (Induced intrinsic metrics).**

(i) *The angular metric on the circle.* We start with the unit circle $\mathbb{S}^1$ with the restriction of the Euclidean metric $d$ of $\mathbb{R}^2$. The variational length of a path $\gamma$ on $\mathbb{S}^1$ is the length of the arc, see Figure 1.14. So the induced intrinsic metric $\hat{d}(x, y)$ of points on $\mathbb{S}^1$ is just given by the length of the shorter arc between $x$ and $y$, hence the angle between $x$ and $y$.

(ii) *Union of segments.* Consider the following union of segments in $\mathbb{R}^2$

$$U = \bigcup_{n=1}^{\infty} [(0,1), (\frac{1}{n}, 0)] \cup [(0,1), (0,0)] \tag{1.2.19}$$

(Figure 1.15) with the restriction of the Euclidean metric $d$ of $\mathbb{R}^2$. One can go from one point to any other by traveling along the segments via $(0, 1)$. The induced intrinsic metric clearly is the Euclidean distance traveled along these segments, but its topology dramatically differs from the original metric topology. The sequence of points $(1/n, 0)$ converges to the origin in the Euclidean topology but does not w.r.t. the topology of the induced intrinsic metric. In fact, the distance between these points is constant 2.

(iii) *Cone over a non-rectifiable curve.*[7] We begin with an injective non-rectifiable path in the $xy$-plane in $\mathbb{R}^3$ with its restriction to any nontrivial interval being non-rectifiable. Now construct the cone over it with vertex say in $(0, 0, 1)$, that is, we connect the vertex with any point on the curve with a straight line segment and again we start with the Euclidean metric, see Figure 1.16. As in the previous example the induced intrinsic distance is given by the Euclidean length of the path from one point to the other via the vertex. Now we remove the vertex. Then the space is disconnected (i.e., it is the union of disjoint open sets) in the topology of the induced intrinsic metric by the properties of the 'base curve' but it is still connected in the original topology.

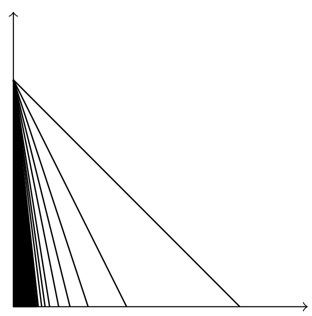


Figure 1.15: The sequence $(0, 1/n)$ converges to $(0, 0)$ in the original topology but is divergent in the topology of the induced intrinsic metric.

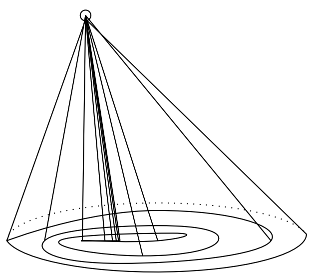


Figure 1.16: A hint at a cone over a non-rectifiable curve with the vertex removed. (Note that on any segment the curve would have to be non-rectifiable, as is hinted here in its center)

[7] The few parts appearing in small print have not been included in the lecture course.

Returning to the construction of the induced intrinsic metric $\hat{d}$ on a metric space $(X,d)$ expressed in (1.2.18), we see that this construction can be iterated. The metric space $(X,\hat{d})$ gives rise to "second stage" induced intrinsic metric $\hat{\hat{d}}$. However, before the reader starts worrying about an emerging hierarchy of induced intrinsic metrics we can give an all-clear signal: the length induced by the induced intrinsic metric $\hat{d}$ is the same as the one induced by the original metric $d$.

**1.2.11 Proposition ($\hat{\hat{d}} = \hat{d}$).** *Let $(X,d)$ be a metric space and let $\hat{d}$ be its induced intrinsic metric, then the following holds:*

*(i) Let $\gamma$ be rectifiable in $(X,d)$ then $L_{\hat{d}}(\gamma) = L_d(\gamma)$.*

*(ii) The intrinsic metric induced by $\hat{d}$ coincides with $\hat{d}$.*

In other words the proposition essentially says that inducing a length metric is an idempotent operation.

**Proof.** (i) Let $\gamma : [a,b] \to (X,d)$ be any rectifiable path.
By the generalized triangle inequality 1.2.3, $L_d(\gamma) \geq d(\gamma(a),\gamma(b))$ hence $\hat{d} \geq d$ and so $L_{\hat{d}}(\gamma) \geq L_d(\gamma)$ since both lengths are variational.
To prove the converse estimate first note that by 1.1.11 $\gamma$ is also a path in $(X,\hat{d})$. Now let $\sigma = \{t_0,\ldots,t_n\}$ be a partition of $[a,b]$ then (since $\hat{d}$ is intrinsic) $\hat{d}(\gamma(t_i),\gamma(t_{i+1})) \leq L_d(\gamma,t_i,t_{i+1})$ which gives

$$V_\sigma^{\hat{d}}(\gamma) = \sum \hat{d}(\gamma(t_i),\gamma(t_{i+1})) \leq L_d(\gamma)\,, \tag{1.2.20}$$

and so $L_{\hat{d}}(\gamma) \leq L_d(\gamma)$.
(ii) To begin with, note that again by the generalized triangle inequality 1.2.3, $L_{\hat{d}}(\gamma) \geq \hat{d}(\gamma(a),\gamma(b))$ for all rectifiable paths $\gamma$ in $(X,\hat{d})$ and so $\hat{\hat{d}} \geq \hat{d}$.
To prove the converse inequality the main issue is to keep track of the respective classes of paths: $\hat{\hat{d}}$ is constructed by using all rectifiable paths $\gamma$ in $(X,\hat{d})$ which by 1.1.11 contains all rectifiable paths in $(X,d)$ for which the respective lengths agree by (i). So we obtain

$$\begin{aligned}
\hat{\hat{d}}(x,y) &= \inf\{L_{\hat{d}}(\gamma) :\ \gamma \text{ rectifiable in } (X,\hat{d}) \text{ from } x \text{ to } y\} \\
&\leq \inf\{L_{\hat{d}}(\gamma) :\ \gamma \text{ rectifiable in } (X,d) \text{ from } x \text{ to } y\} \\
&= \inf\{L_d(\gamma) :\ \gamma \text{ rectifiable in } (X,d) \text{ from } x \text{ to } y\} = \hat{d}(x,y).
\end{aligned} \tag{1.2.21}$$

□

**1.2.12 Remark (Rectifiability is essential).** Observe that the assumption that $\gamma$ is rectifiable is not superfluous. In fact a path in $(X,d)$ may fail to be continuous w.r.t. $\hat{d}$, cf. Lemma 1.1.15. So the set of all paths in $(X,\hat{d})$ in general is a subset of all paths in $(X,d)$ but it contains all rectifiable paths in $(X,d)$ by Lemma 1.1.11.

## 1.3 Characterizing intrinsic metrics

We have called a metric intrinsic, if it can be constructed as the length metric from a length structure, cf. Definition 1.1.8, hence if it can be given by a certain construction. From that definition it is hard to tell whether a given metric is intrinsic or not. The aim of this section is to discuss properties that distinguish intrinsic metrics from general ones and to find criteria which allow to identify intrinsic metrics.

### 1.3.1 An alternative definition of length spaces

To begin with, note that Proposition 1.2.11 gives a constructive criterion in the above sense. A metric is induced by a length structure coming from a metric, iff it induces itself. Indeed the 'only if'-part is just 1.2.11(ii) while the 'if'-statement is immediate since then $d = \hat{d}$ which clearly is induced. In fact the restriction to length structures coming from a metric themselves is not necessary as we show next.

**1.3.1 Proposition ($\hat{d} = d$).** *Let $(X, d)$ be a length space and let $\hat{d}$ be the intrinsic metric induced by d. Then $\hat{d} = d$ and $L_d(\gamma) \le L(\gamma)$ for all paths $\gamma \in A$ of finite L-length, where $(X, A, L)$ is a length structure inducing $(X, d)$.*

**Proof.** Basically we just observe that the respective parts of the proof of 1.2.11 still apply in this situation.
First the generalized triangle inequality 1.2.3 gives $\hat{d} \ge d$.
Conversely any admissible path $\gamma$ in the length structure defining $d$ which is of finite length is also a path in $(X, d)$, again by 1.1.11 and so the analogue of (1.2.20) leads to $L_d(\gamma) \le L(\gamma)$, and so the analogue of (1.2.21) gives $\hat{d} \le d$.

□

Note that (as in the above discussion of the special case) the equality $d = \hat{d}$ implies that $d$ is intrinsic (since $\hat{d}$ is!). This leads us to the following alternative characterisation of intrinsic metrics and length spaces:

$$(X, d) \text{ is a length space, iff } \hat{d} = d \tag{1.3.1}$$

This statement translates into the following criterion which can be seen as an alternative definition of length spaces.

**1.3.2 Corollary (Characterization of (strictly) intrinsic spaces[8]).** *Let $(X, d)$ be a metric space, then we have:*

*(i) $(X, d)$ is a length space, iff for all pairs of points $x, y \in X$ and for all $\varepsilon > 0$ there is a connecting path $\gamma$ with $L_d(\gamma) \le d(x, y) + \varepsilon$.*

*(ii) $(X, d)$ is strictly intrinsic, iff for all pairs of points $x, y \in X$ there is a connecting path $\gamma$ with $L_d(\gamma) = d(x, y)$.*

**Proof.** (i) ⇒: If $(X, d)$ is a length space then by definition there is a connecting path $\gamma$ for any pair of points with $L(\gamma) \le d(x, y) + \varepsilon$ and since by Prop. 1.3.1 $L_d \le L$, the assertion follows.

(ii) ⇒: In this case we even have $L(\gamma) = d(x, y)$ and then, again by Proposition 1.3.1,

$$d(x, y) = \hat{d}(x, y) \le L_d(\gamma) \le L(\gamma) = d(x, y). \tag{1.3.2}$$

(i) ⇐: If such a path $\gamma$ exists we obtain $\hat{d} \le d$ and since $\hat{d} \ge d$ always holds (see the proof of 1.3.1) we are done, thanks to (1.3.1).

(ii) ⇐: In this case we still have $\hat{d} \le d$ and hence $d = \hat{d}$, which is strictly intrinsic. □

[8] This statement was changed after the lecture: (i) is item 1.3.2 of the lecture, while (ii) was added to display the argument occurring at the end of the proof of 1.3.12(i) in a more prominent place. Also the proof has been streamlined.

### 1.3.2 Recovering the length structure

Now we address another natural question: Given a length space $(X, d)$, can we recover the initial length structure $(X, A, L)$?
While this is impossible, since clearly many different length structures may give rise to the same intrinsic metric, there is at least one natural candidate length structure that induces $d$: $(X, C, L_d)$ as suggested by 1.3.1. So we better rephrase our question to: Under which condition is $L = L_d$?
We already know parts of the answer: Since $L_d$ is always lower semi-continuous by 1.2.8, this is a necessary condition for $L = L_d$ to hold. We will see below that it is also sufficient.
Apart from the general issue raised here, it is also practically relevant, to know that $L = L_d$. In case that $L$ is known, one can then use the specific properties of $L$ and combine it with the general properties of the variational length $L_d$.

**1.3.3 Theorem (Semi-continuity of the length is characterizing).** *Let $(X, A, L)$ be a length structure with lower semi-continuous length $L$ (where $A$ is equipped with the topology of uniform convergence). Then $L$ coincides with the length induced by its intrinsic metric $d_L$, i.e., $L(\gamma) = L_{d_L}(\gamma)$ for all $\gamma \in A$ of finite $L$-length.*

**Proof.** Since $L_{d_L}(\gamma) \leq L(\gamma)$ holds for any length structure by 1.3.1 we only have to show the converse inequality. To begin with observe that by (L2) the function $L(t) = L(\gamma|_{[a,t]})$ is uniformly continuous on $[a, b]$ for any $\gamma \in A$ of finite length. Hence for any $\varepsilon > 0$ there is a partition $\sigma = \{t_0, \ldots, t_n\}$ such that

$$d_L\big(\gamma(t_i), \gamma(t_{i+1})\big) \leq |L(t_{i+1}) - L(t_i)| < \varepsilon \qquad \text{for all } 0 \leq i \leq n-1. \tag{1.3.3}$$

By the definition of $d_L$, for all $i$ there is a path $\sigma_i : [t_i, t_{i+1}] \to X$ with endpoints $\sigma(t_i) = \gamma(t_i)$ and $\sigma(t_{i+1}) = \gamma(t_{i+1})$ such that

$$L(\sigma_i) \leq d_L\big(\gamma(t_i), \gamma(t_{i+1})\big) + \frac{\varepsilon}{n}\,. \tag{1.3.4}$$

Denoting by $h_\varepsilon$ the concatenation of all $\sigma_i$ we obtain

$$L(h_\varepsilon) = \sum L(\sigma_i) \leq \sum d_L\big(\gamma(t_i), \gamma(t_{i+1})\big) + \varepsilon \leq L_{d_L}(\gamma) + \varepsilon\,. \tag{1.3.5}$$

Now by the triangle inequality we have for all $t \in [a, b]$ (choosing $t_i$ such that $t_i < t \leq t_{i+1}$)

$$\begin{aligned} d_L\big(\gamma(t), h_\varepsilon(t)\big) &\leq d_L\big(\gamma(t), \gamma(t_i)\big) + d_L\big(\sigma_i(t_i), h_\varepsilon(t)\big) \\ &\leq \varepsilon + L_d(\sigma_i|_{[t_i,t]}) \leq \varepsilon + d\big(\gamma(t_i), \gamma(t_{i+1})\big) + \frac{\varepsilon}{n} \leq 3\varepsilon\,, \end{aligned} \tag{1.3.6}$$

where we have used (1.3.3) twice and (1.3.4) together with $L_{d_L} \leq L$. Since the topology induced by $d_L$ is finer than the original one (by 1.1.15) we obtain that $h_\varepsilon \to \gamma$ uniformly. Now the lower semi-continuity of $L$ (see 1.2.8) gives

$$L(\gamma) \leq \liminf_{\varepsilon \to 0} L(h_\varepsilon) \leq L_{d_L}(\gamma)\,, \tag{1.3.7}$$

where we have used (1.3.5) in the final estimate. □

**1.3.4 Example (Not a semi-continuous length).** On $\mathbb{R}^2$ we consider the Finslerian length of 1.1.18(iv), i.e.,

$$L(\gamma) = \int f(\gamma(t), \gamma'(t))dt\,,$$

where we now choose $f$ to be independent of the first variable. (Intuitively this means that the speed of our travel only depends on the direction but not on the position; think of a sailing boat in a constant wind field.) Moreover we assume that $f(x, (1, 0)) = f(x, (0, 1)) = 1/10$ and $f(x, (1, 1)) = 1$, see Figure 1.17.

This length structure is not lower semi-continuous. To see this consider a sequence of stair like paths of broken lines with its pieces parallel to the axes approaching the segment $[(0,0),(1,1)]$, see Figure 1.18. Now the length of the latter segment is 1 while the stair like paths all have length 1/5. Hence this length structure is not lower semi-continuous and by 1.3.3 does not come from a metric. (The intuitive reason is that traveling along the diagonal is too fast; think of it as the prevailing wind direction.)

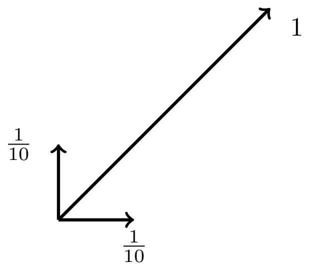



Figure 1.17: The function $f$.

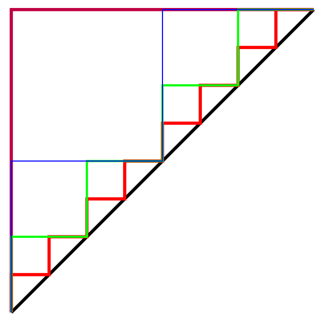


Figure 1.18: The length of the stair like paths is constant 1/5 while the length of the diagonal is 1.

### 1.3.3 Midpoints

The next topic to be discussed is midpoints and their existence. Here we will find another way of characterizing (strictly) intrinsic (complete) metrics. We start by making the key-notion precise.

**1.3.5 Definition (Midpoint).** *A point $y$ in a metric space $(X,d)$ is called a* midpoint *between $x,z \in X$ if*

$$d(x,y) = d(y,z) = \frac{1}{2}d(x,z). \tag{1.3.8}$$

The next lemma should be interpreted as giving a necessary condition for a metric to be strictly intrinsic. In fact, under the assumption of completeness, it is also sufficient, as will see below.

**1.3.6 Lemma (Existence of midpoints).** *If $d$ is strictly intrinsic on $X$ then for any pair of points $x,z \in X$ there is a midpoint $y$.*

**Proof.** Let $\gamma : [a,c] \to X$ be a shortest path between $x$ and $z$. Then $L(\gamma) = d(x,z)$ and $L(t) := L(\gamma_{[a,t]})$ is continuous in $t$ with $L(a) = 0$ and $L(c) = L(\gamma) = d(x,z)$. Hence there is a $b \in [a,c]$ with $L(b) = \frac{1}{2}L(c)$ and we set $y = \gamma(b)$, see Figure 1.19.

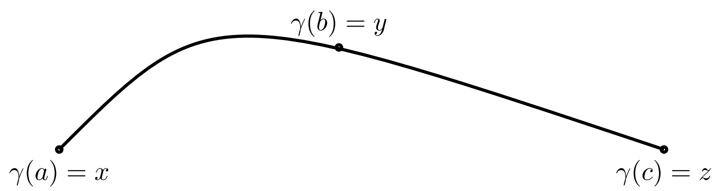



Figure 1.19: The midpoint $y = \gamma(b)$, where $L(b) = \frac{1}{2}L(c)$.

Now we use the fact that the length of a path is not less that the distance between its endpoints to write

$$\begin{aligned} d(x,y) &\le L(\gamma|_{[a,b]}) = L(b) = \frac{1}{2}d(x,z) \\ d(y,z) &\le L(\gamma|_{[b,c]}) = L(c) - L(b) = \frac{1}{2}L(c) = \frac{1}{2}d(x,z). \end{aligned} \tag{1.3.9}$$

Using the latter estimate and the triangle inequality we obtain

$$d(x,y) \geq d(x,z) - d(y,z) \geq d(x,z) - \frac{1}{2}d(x,z) = \frac{1}{2}d(x,z) \tag{1.3.10}$$

and similarly for $d(y,z)$. □

Another way to interpret the assertion of Lemma 1.3.6 is to say that in a strictly intrinsic metric space the closed balls $\bar{B}_{d(x,y)/2}(x)(:= \{z : d(x,z) \leq d(x,y)/2\})$ and $\bar{B}_{d(x,y)/2}(y)$ are not disjoint. Note that the intersection can be much bigger than just one point, e.g. on $S^2$, choosing $x$, $y$ to be the poles, it is the entire equator.

Also observe that it is not essential to deal with *mid*points or with balls of the radius $d(x,z)/2$. In fact, the poof 1.3.6 literally works the same if we choose $d_1, d_2 \in \mathbb{R}$ with $d_1 + d_2 = d(x,z)$ to provide the existence of a point $y$ with $d(x,y) = d_1$ and $d(y,z) = d_2$. Iterating this procedure one further obtains the following statement[9]

**1.3.7 Corollary (Measuring distance by small jumps).** *Let $(X,d)$ be strictly intrinsic. Given any pair of points $x, y \in X$ and any finite sequence of positive numbers $d_1, \ldots, d_k$ with $d_1 + \cdots + d_k = d(x,y)$, there is a sequence of points $x = x_0, x_1, \ldots, x_k = y$ with $d(x_{i-1}, x_i) = d_i$ for all $i = 1, \ldots, k$.*

Note that in this case we have

$$\sum_{i=0}^{k-1} d(x_i, x_{i+1}) = d(x,y). \tag{1.3.11}$$

which (informally) says that in a strictly intrinsic space we can measure lengths (not only by the length of connecting paths but also) by the use of dotted lines or small jumps.

Finally, we mention that there is a notion equivalent to the existence of midpoints called *Menger convexity*. More precisely ([Pap14, Def. 2.6.1]), a metric space $(X,d)$ is called *Menger convex* if for all pairs of points $x, z \in X$ there is a point $y \in X$ *between* $x$ and $z$, i.e., $x \neq y \neq z$ and $d(x,y) + d(y,z) = d(x,y)$. Clearly, the existence of midpoints implies Menger convexity; for a proof of the converse see the proof of (i)⇒(ii) in [Pap14, Prop. 2.6.2, p. 72].

For (merely) intrinsic metrics we have analogous statements based on the notion of $\varepsilon$-midpoints.

**1.3.8 Definition ($\varepsilon$-midpoint).** *A point $y$ in a metric space $(X,d)$ is an $\varepsilon$-midpoint for $x, z \in X$ if*

$$|2d(x,y) - d(x,z)| \leq \varepsilon \quad \text{and} \quad |2d(y,z) - d(x,z)| \leq \varepsilon.$$

**1.3.9 Lemma (Existence of $\varepsilon$-midpoints).** *If $d$ is intrinsic on $X$ then for any positive $\varepsilon$ and for any pair of points $x, z \in X$ there is an $\varepsilon$-midpoint $y$. In other words if $2r > d(x,z)$ then the (open) balls $B_r(x)$ and $B_r(z)$ are not disjoint.*

**Proof.** We may essentially repeat the above proof with a path $\gamma$ from $x$ to $z$ such that $L(\gamma) \leq d(x,z) + \varepsilon$. Then again defining $y = \gamma(b)$, where $L(b) = \frac{1}{2}L(c)$ does the trick. Indeed we obtain

$$d(x,y) \leq L(\gamma|_{[a,b]}) = L(b) = \frac{L(c)}{2} \leq \frac{d(x,z)}{2} + \frac{\varepsilon}{2}, \text{ i.e.,} \quad 2d(x,y) - d(x,z) \leq \varepsilon$$

and similarly $2d(y,z) - d(x,z) \leq \varepsilon$. So by the triangle inequality

$$d(x,y) \geq d(x,z) - d(y,z) \geq d(x,z) - \frac{d(x,z)}{2} - \frac{\varepsilon}{2} \geq \frac{d(x,z)}{2} - \frac{\varepsilon}{2},$$

hence $-2d(x,y) + d(x,z) \leq \varepsilon$ and similarly for $d(y,z)$. □

Tweaking the proof of 1.3.9 in an analogous way we may first replace $d(x,z)/2$ by $d_1 + d_2 = d(x,z)$ and then by iterating obtain the following version of Corollary 1.3.7.

[9] This is actually [Pap14, Cor. 2.6.3] which generalizes [BBI01, Cor. 2.4.12].

**1.3.10 Corollary (Measuring distance by small jumps, 2).** *Let $(X, d)$ be intrinsic. Given any pair of points $x, y \in X$, any $\varepsilon > 0$, and any finite sequence of positive numbers $d_1, \dots, d_k$ with $d_1 + \cdots + d_k \le d(x, y) + \varepsilon$, there is a sequence of points $x = x_0, x_1, \dots, x_k = y$ with $d(x_{i-1}, x_i) = d_i$ for all $i = 1, \dots, k$.*

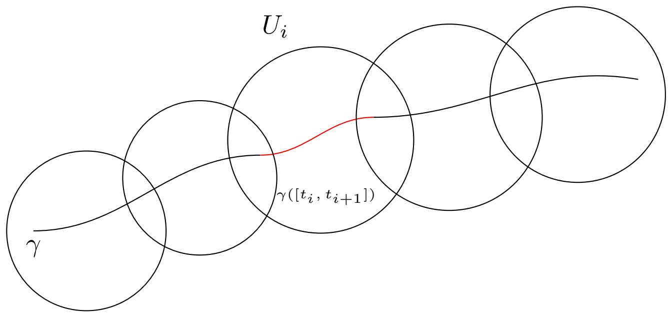
In this case we now have

$$\sum_{i=0}^{k-1} d(x_i, x_{i+1}) \le d(x, y) + \varepsilon. \tag{1.3.12}$$

The essential feature of (1.3.11) and (1.3.12) is that, in many situations, it allows one to turn local properties into global ones.

**1.3.11 Corollary (Global Lipschitz functions on length spaces).** *Let $(X, d)$ be a length space, $Y$ a metric space and let $f : X \to Y$ be a locally Lipschitz map with Lipschitz constant $C$. Then $f$ is globally Lipschitz with Lipschitz constant $C$.*

**Proof.** Let $x, y \in X$, $\varepsilon > 0$ and choose a connecting path $\gamma : [a, b] \to X$ with $L(\gamma) \le d(x, y) + \varepsilon$. We cover $\gamma([a, b])$ by finitely many neighbourhoods $U_i$ $(1 \le i \le k)$ on which the Lipschitz property holds. Now we introduce a partition $a = t_1 < t_2 < \cdots < t_{k+1} = b$ of $[a, b]$ such that $\gamma([t_i, t_{i+1}]) \subseteq U_i$ $(1 \le i \le k)$ and set $x_i = \gamma(t_i)$.




Then we have

$$\sum_{i=1}^{k} d(x_i, x_{i+1}) \le L(\gamma) \le d(x, y) + \varepsilon \tag{1.3.13}$$

which further gives

$$d(f(x), f(y)) \le \sum_{i=1}^{k} d(f(x_i), f(x_{i+1})) \le C \sum_{i=1}^{k} d(x_i, x_{i+1}) \le C(d(x, y) + \varepsilon). \tag{1.3.14}$$

Since $\varepsilon$ was arbitrary we are done. □

In many cases a converse of Lemma 1.3.6 holds true: The existence of midpoints (resp. $\varepsilon$-midpoints) implies that a *complete* metric is strictly intrinsic (resp. intrinsic). In particular, we have a criterion that tells us whether a complete metric space is a length space.

**1.3.12 Theorem (Complete length spaces from midpoints).** *Let $(X, d)$ be a complete metric space. Then we have*

*(i) If for every $x, z \in X$ there is a midpoint, then $d$ is strictly intrinsic.*

*(ii) If for every $x, z \in X$ and every $\varepsilon > 0$ there is an $\varepsilon$-midpoint, then $d$ is intrinsic.*

Before proving the theorem, we note that it has the following immediate consequence which gives another criterion[10] for a complete metric space to be a length space.

[10] Again this is a slight generalization of [BBI01, Cor. 2.4.17].

**1.3.13 Corollary (Complete length spaces from small jumps).** *A complete metric space $(X, d)$ is a length space iff, given any pair of points $x, y \in X$, any $\varepsilon > 0$, and any finite sequence of positive numbers $d_1, \dots, d_k$ with $d_1 + \cdots + d_k \leq d(x, y) + \varepsilon$ there are points $x = x_0, x_1, \dots, x_k = y$ with $d(x_{i-1}, x_i) = d_i$ for all $1 \leq i \leq k$.*

To see the meaning of this statement more clearly set all $d_i = \varepsilon$. Then the complete metric space $(X, d)$ is a length space iff we can reach any $y$ from $x$ by hopping with small jumps of length $\varepsilon$ and the total length of the jumps does not exceed $d(x, y)$ by more than $\varepsilon$.

**Proof of Theorem 1.3.12.** We first prove (i) by showing that $x$ and $y$ can be joined by a path $\gamma$ of length $d(x, y)$. To construct such a path we start by setting $\gamma(0) = x$ and $\gamma(1) = y$. Then we use the existence of a midpoint to obtain some $x_{\frac{1}{2}}$ with

$$d(x, x_{\frac{1}{2}}) = d(x_{\frac{1}{2}}, y) = \frac{1}{2} d(x, y). \tag{1.3.15}$$

We set $\gamma(\frac{1}{2}) := x_{\frac{1}{2}}$ and again by existence of midpoints we obtain $x_{\frac{1}{4}} = \gamma(\frac{1}{4})$, $x_{\frac{3}{4}} =: \gamma(\frac{3}{4})$ such that

$$d\Big(\gamma(\frac{j}{4}), \gamma(\frac{j+1}{4})\Big) = d(x_{\frac{j}{4}}, x_{\frac{j+1}{4}}) = \frac{1}{4} d(x, y) \qquad (0 \leq j \leq 3). \tag{1.3.16}$$

In this way we inductively assign the values of $\gamma$ for all dyadic rational numbers $k/2^m$ for $m \in \mathbb{N}$ and $0 \leq k \leq 2^m$ in a way such that

$$d\Big(\gamma(\frac{j}{2^m}), \gamma(\frac{j+1}{2^m})\Big) = d(x_{\frac{j}{2^m}}, x_{\frac{j+1}{2^m}}) = \frac{1}{2^m} d(x, y) \qquad (0 \leq j \leq m-1). \tag{1.3.17}$$

By construction we hence have for any two dyadics $t, t'$ that

$$d(\gamma(t), \gamma(t')) = |t - t'|\, d(x, y), \tag{1.3.18}$$

implying that the map $\gamma$ defined on the dyadics is Lipschitz, and in particular uniformly continuous. Since the dyadic rationals are dense in $[0, 1]$ and $(X, d)$ is complete, $\gamma$ can be extended uniquely to a (uniformly) continuous map on the entire interval $[0, 1]$ (see e.g. [Wil04, Thm. 39.10]). Hence we obtain a path $\gamma : [0, 1] \to X$ connecting $x$ and $y$. Moreover (1.3.18) extends to all $t, t' \in [0, 1]$ and so

$$L_d(\gamma) = \sup_{\sigma = \{t_0, \dots, t_k\}} \sum d(\gamma(t_i), \gamma(t_{i+1})) = \sum |t_i - t_{i+1}|\, d(x, y) = d(x, y), \tag{1.3.19}$$

and by 1.3.2(ii) we are done.

To prove (ii) we only need a minor modification. Instead of assigning $\gamma(k/2^m)$ to be the respective midpoints we have to use the respective $\varepsilon$-midpoints. Then equations (1.3.15) and (1.3.16) turn into $d(x, x_{\frac{1}{2}}) = \frac{1}{2} d(x, y) \pm \varepsilon$, $d(x_{\frac{1}{2}}, y) = \frac{1}{2} d(x, y) \pm \varepsilon$ and $d(\gamma(\frac{j}{4}), \gamma(\frac{j+1}{4})) = \frac{1}{4} d(x, y) \pm \frac{1}{2}\varepsilon \pm \varepsilon$. Consequently (1.3.17) turns into the estimate

$$\left| d\Big(\gamma(\frac{j}{2^m}), \gamma(\frac{j+1}{2^m})\Big) - \frac{1}{2^m} d(x, y) \right| \leq \sum_{j=0}^{m-1} \frac{1}{2^j} \varepsilon \leq 2\varepsilon, \tag{1.3.20}$$

which still gives uniform continuity of $\gamma$ on the dyadic rationals. Hence we obtain a path $\gamma : [0, 1] \to X$ with $L_d(\gamma) \leq d(x, y) + \varepsilon$ and by 1.3.2(i) we are done. □

Theorem 1.3.12 is a valuable tool to prove that certain constructions involving length spaces are again length spaces, just by showing the existence of $\varepsilon$-midpoints. E.g one can show that the finite product of complete length spaces (equipped with the 2-metric $d_2(x, y) = (\sum_{i=1}^m (d_j(x_j, y_j)^2)^{(1/2)})$ is again a length space (cf. [BBI01, Prop. 3.6.1]) and that the completion of a length space is a length space as we will explicitly show next.

**1.3.14 Proposition (Completion of length spaces).** *The completion $X'$ of a length space $X$ is a length space again.*

**Proof.** Let $x, z \in X'$ and $x_n \to x$, $z_n \to z$ be approximating sequences in $X$. Choose $\varepsilon > 0$ and fix $n$ so large that $d(x_n, x)$, $d(z_n, z) \leq \varepsilon/8$. By Lemma 1.3.9 there is an $\varepsilon/2$-midpoint $y_n$ of $x_n$, $z_n$, i.e.,

$$|2d(x_n, y_n) - d(x_n, z_n)| \leq \frac{\varepsilon}{2}, \quad |2d(y_n, z_n) - d(x_n, z_n)| \leq \frac{\varepsilon}{2}. \tag{1.3.21}$$

We show that $y_n$ is an $\varepsilon$-midpoint for $x$, $z$. By the above estimate we have $d(x_n, y_n) \leq \frac{1}{2}(d(x_n, z_n) + \frac{\varepsilon}{2})$ and so

$$\begin{aligned} d(x, y_n) &\leq d(x, x_n) + d(x_n, y_n) \leq \frac{\varepsilon}{8} + \frac{1}{2} d(x_n, z_n) + \frac{\varepsilon}{4} \\ &\leq \frac{3\varepsilon}{8} + \frac{1}{2}(d(x_n, x) + d(x, z) + d(z, z_n)) \leq \frac{\varepsilon}{2} + \frac{1}{2} d(x, z) \end{aligned} \tag{1.3.22}$$

hence $2d(x, y_n) - d(x, z) \leq \varepsilon$. The other estimates follow similarly. Hence by Theorem 1.3.12(ii) $X'$ is a length space. □

## 1.4 Shortest paths

In this section we will deal with shortest paths, that is paths realizing the metric distance between their endpoints. In particular, we will consider the question of existence of shortest paths between a pair of given points. Moreover, we will state and prove the theorem of Hopf-Rinow-Cohn-Vossen, which generalizes the Hopf-Rinow theorem of Riemannian geometry and relates metric completeness to the possibility of extending shortest paths. We start, however, by introducing the notion of a curve.

### 1.4.1 Curves & natural parametrizations

So far we have always dealt with paths, that is continuous maps $I \to X$. However, geometrically one is more often interested in the image $\gamma(I)$ of a path than in the actual map $\gamma$. To adequately describe the geometric object itself one defines the notion of a curve as an equivalence class of paths. The underlying equivalence relation is given by a suitable notion of reparametrizations, i.e., two paths $\gamma_1 : I_1 \to X$ and $\gamma_2 : I_2 \to X$ are called equivalent if there is a suitable change of parameter $\varphi : I_1 \to I_2$ such that $\gamma_1 = \gamma_2 \circ \varphi$. In analysis, when studying piecewise $C^1$-paths a convenient family of changes of parameter are all $C^1$-diffeomorphims, for continuous paths one often uses homeomorphisms or, more generally, increasing and surjective (hence continuous) maps. Here we will also follow this road, but admit a(n unusually) general class of reparametrisations—as announced already at the beginning of Section 1.2.2. In particular, we want to also include parametrisations where a path is constant on some subinterval. The associated picture one should have in mind is that of a path that stops for a while at one point and then moves on. We formalize this in the following way.

**1.4.1 Definition (Curve).** *A curve is an equivalence class of the minimal equivalence relation satisfying the following: Two paths $\gamma_i : I_i \to X$ $(i = 1, 2)$ are equivalent if there is a nondecreasing surjective map $\varphi : I_1 \to I_2$ such that $\gamma_1 = \gamma_2 \circ \varphi$.*

Paths are hence representatives of curves and also sometimes called *parametrized curves* or *parametrizations* of a curve and *reparametrizations* of each other. Moreover, we call a $\varphi$ as in the definition a (nondecreasing) *change of parameter*. Such a map $\varphi$ automatically is continuous: Suppose $\varphi$ is discontinuous at some $t_0$ then $\lim_{t \nearrow t_0} \varphi(t) < \varphi(t_0)$ and surjectivity fails.

**1.4.2 Remark (Curves as equivalence classes).**

(i) Observe that the existence of a non-strictly monotone change of variables does not lead to an equivalence relation: due to the lack of an inverse change of variables, symmetry of the relation fails.

Now, the minimal equivalence relation generated by a given relation $R$ on a set $B$ (seen as subset $R \subseteq B \times B$) is defined as the intersection of all equivalence relations containing $R$[11]. In the ensuing equivalence relation we have $x \sim y$ if $x = y$ or there are $x = x_1, x_2, \dots, x_k = y$ for some $k \in \mathbb{N}$ such that $(x_i, x_{i+1}) \in R$ or $(x_{i+1}, x_i) \in R$ for all $1 \leq i \leq k-1$.

In our situation this means that two paths $\gamma$, $\tilde{\gamma}$ are equivalent if there is a finite sequence of paths $\gamma = \gamma_1, \dots, \gamma_k = \tilde{\gamma}$ and each $\gamma_i$ and $\gamma_{i+1}$ are related by a nondecreasing change of parameter or vice versa.

(ii) We can, however, simplify this description by introducing the notion of a *never-locally-constant* reparametrization of a curve. More precisely, we call a path $\gamma : I \to X$ never-locally-constant if there is no interval $[a, b] \subseteq I$ such that $a \neq b$ and $\gamma|_{[a,b]}$ is constant.

Now every curve allows for such a parametrization: Let $\gamma : I \to X$ be an arbitrary path and introduce an equivalence relation on $I$ by $s \sim t$ if $\gamma$ is constant on $[s, t]$. Then the quotient $J : I/\sim$ is an interval and there is a unique map $\tilde{\gamma} : J \to X$ such that $\gamma = \tilde{\gamma} \circ \pi$, where $\pi : I \to J$ is the quotient map. By construction $\tilde{\gamma}$ is never-locally-constant.

Then two paths $\gamma_i : I_i \to X$ $(i = 1, 2)$ are equivalent iff both can be reparametrized to the same never-locally-constant path, that is if there is such a path $\gamma : J \to X$ and $\varphi_i : I_i \to J$ with $\gamma_i = \gamma \circ \varphi_i$.

**1.4.3 Remark (Length of a curve).** We have already shown in 1.2.7 (cf. the remark above it) that all (monotonous surjective) parametrizations of a curve have the same length. Hence we can speak of the length of a curve.

In the following we will follow the widely used sloppiness to denote a curve and (one of) its parametrization(s) by the same letter. In particular, we will speak of a curve $\gamma : I \to X$ meaning the curve defined by the class of the path $\gamma$.

Next we make precise the notion of a curve without 'self intersections'.

**1.4.4 Definition (Simple curve).** *A curve $\gamma : [a, b] \to X$ is called simple if the pre-image of every point is an interval.*

The idea is that although we have allowed our paths to stop for a while at one point and then to go on, we now want to exclude them from ever returning to this point again.
Observe that the above definition makes sense because every parametrization of a simple curve is itself simple, since continuous images of connected sets are connected. Moreover, if two simple curves have the same image then they coincide up to a change of variable.

Our next goal is to introduce a parametrization for each curve that is analogous to unit speed parametrization in differential geometry.

**1.4.5 Definition (Natural parametrization).** *A parametrization $\gamma : I \to X$ is natural if $L(\gamma, s, t) = t - s$ for all $s, t \in I$.*

Other names for natural parametrizations are *arc-length parametrizations* or *parametrizations by arc length.*
Observe that a parametrization $\gamma$ is natural iff $L(\gamma, a, t) = t - a$ for some fixed $a \in I$ and all $t \in I$ by additivity of length (L1). Another equivalent condition is that

$$\frac{d}{dt} L(\gamma, a, t) = 1. \tag{1.4.1}$$

[11] It obviously holds that the intersection of any collection of equivalence relations is also an equivalence relation.

[Indeed if $\gamma$ is natural then $L(\gamma, a, t) = t - a$ and hence $L' = 1$ and the converse follows from integrating (1.4.1) and using $L(\gamma, a, a) = 0$.]
Observe that by (1.4.1) a natural parametrization of a curve is unique up to translations $t \mapsto t + c$ for some $c \in \mathbb{R}$. Moreover, motivated by (1.4.1) we call natural parametrizations also *unit speed parametrizations*. More generally we call $\gamma$ a *constant speed parametrization* if

$$L(\gamma, s, t) = v(t - s) \tag{1.4.2}$$

for all $s, t \in I$ and some $0 \leq v \in \mathbb{R}$.

We can always find a unit speed parametrization for any curve.

**1.4.6 Proposition (Existence of natural parametrizations).** *Every rectifiable curve $\gamma : [a, b] \to X$ can be represented in the form*

$$\gamma = \bar{\gamma} \circ \varphi, \text{ where } \bar{\gamma} : [0, L(\gamma)] \to X \text{ is a natural parametrization} \tag{1.4.3}$$

*and $\varphi : [a, b] \to [0, L(\gamma)]$ is a nondecreasing surjective map.*

We will explicitly construct a natural reparametrization for rectifiable paths. The key idea is to define $\bar{\gamma}(t)$ as the point on the image of $\gamma$ for which $L(\gamma, a, t) = t$. We formalize this idea explicitly.

**1.4.7 Lemma (Constructing a natural parametrization).** *Let $\gamma : [a, b] \to X$ be a rectifiable path.*

*(i) For each $l \in [0, L(\gamma)]$ there is a unique point $x \in X$ and some parameter value $t \in [a, b]$ such that*

$$x = \gamma(t) \text{ and } L(\gamma, a, t) = l. \tag{1.4.4}$$

*Moreover, the set of all such values $t$ is a closed subinterval of $[a, b]$ and $\gamma$ is constant on it.*

*(ii) We define the map $\bar{\gamma} : [0, L(\gamma)] \to X$ by $\bar{\gamma}(l) = x = \gamma(t)$, where $x = \gamma(t)$ is the unique point provided by (i). Then $\bar{\gamma}$ is Lipschitz continuous and hence a path. Furthermore, $\bar{\gamma}$ is obtained from $\gamma$ as $\gamma = \bar{\gamma} \circ \varphi$ via the nondecreasing and surjective change of parameter $\varphi : [a, b] \to [0, L(\gamma)]$ defined by $\varphi(t) = L(\gamma, a, t)$.*

**Proof.** (i) By (L2) $L(\gamma, a, .)$ is continuous and so by the mean value theorem for any $l \in [0, L(\gamma)]$ there is $t \in [a, b]$ such that $L(\gamma, a, t) = l$ and we set $x := \gamma(t)$.
Now if $a \leq t < t' \leq b$ are two such parameter values, i.e., $L(\gamma, a, t) = l = L(\gamma, a, t')$ then by (L1) $L(\gamma, t, t') = L(\gamma, a, t') - L(\gamma, a, t) = 0$. By Lemma 1.2.4 $\gamma$ is constant on $[t, t']$, proving that $x$ is unique and that the set of parameter values $t'$ such that $\gamma(t') = x$ is an interval. Finally, again by continuity of $L(\gamma, a, .)$ this interval is closed.

(ii) Let $l_1 \leq l_2 \in [0, L(\gamma)]$ and $t_1, t_2 \in [a, b]$ with $L(\gamma, a, t_i) = l_i$ $(i = 1, 2)$. Then by definition $\bar{\gamma}(l_i) = \gamma(t_i)$ and by the generalized triangle inequality Lemma 1.2.3

$$d(\bar{\gamma}(l_1), \bar{\gamma}(l_2)) = d(\gamma(t_1), \gamma(t_2)) \leq L(\gamma, t_1, t_2) = |l_1 - l_2|, \tag{1.4.5}$$

establishing the Lipschitz property of $\bar{\gamma}$.
Finally, $\varphi$ is obviously nondecreasing and surjective, and by the uniqueness statement in (i) we have $\gamma = \bar{\gamma} \circ \varphi$. □

**Proof of 1.4.6.** Using the construction of 1.4.7(ii) it only remains to check that $\bar{\gamma}$ is a natural parametrization, i.e., that $L(\bar{\gamma}, 0, l) = l$ for all $l \in [0, L(\gamma)]$. Choose such an $l$ and choose $t \in [a, b]$ such that $\varphi(t) = L(\gamma, a, t) = l$. Now, since $\bar{\gamma}|_{[0,l]}$ is just a reparametrization of $\gamma|_{[a,t]}$ we obtain from (L3) that they are of the same length, namely $L(\bar{\gamma}, 0, l) = L(\gamma, a, t) = l$. □

We note that since all arguments employed in the previous proofs only refer to compact subintervals of the domain of $\gamma$, Proposition 1.4.6 remains valid also for half-open or open intervals and non-rectifiable curves (however, the natural parametrization then might be defined on an infinite interval).

Also, from the natural parametrization obtained in this way we may via scaling obtain a parametrization of any prescribed constant speed. Moreover, we note the following geometric consequence of the existence of natural parametrizations.

**1.4.8 Remark (One-dimensional intrinsic geometry is trivial).** Proposition 1.4.6 implies that if a length space is homeomorphic to a segment, then it is already *isometric* to a segment. This implies the fundamental fact that all intrinsic metrics on a line are locally indistinguishable and parallels the fact from Riemannian geometry that all one-dimensional manifolds are flat. However, we will see that already two-dimensional surfaces behave completely different and carry a rich geometric structure.

### 1.4.2 Existence of shortest paths

This section is devoted to proving the following: Every complete and locally compact length space is strictly intrinsic, i.e., there is a shortest path between any two of its points. The proof will be built upon a suitable version of the Arcelà-Ascoli theorem. To begin with, we define the notion of uniform convergence for curves.

**1.4.9 Definition (Uniform convergence of curves).** *A sequence of curves $\gamma_n$ into a metric space is said to converge uniformly to a curve $\gamma$ if all $\gamma_n$ admit parametrizations with the same domain that uniformly converge to a parametrization of $\gamma$.*

Already the next step is to invoke a version of the Arcelà-Ascoli theorem. To formulate it we recall some notions from (functional) analysis. Let $X, Y$ be metric spaces. A sequence of maps $f_n : X \to Y$ is called *uniformly equicontinuous*[12] if

$$\forall \varepsilon > 0\ \exists \eta > 0 :\ \forall x_1, x_2 \in X,\ \forall n : \quad d(x_1, x_2) < \eta \ \Rightarrow\ d(f_n(x_1), f_n(x_2)) < \varepsilon. \tag{1.4.6}$$

The space $X$ is called *proper* if every closed and bounded subset is compact. This property is sometimes also referred to as *finitely compact* or *boundedly compact*[13] or also as the *Heine-Borel property*. The following version of the theorem is taken from [Pap14, Thm. 1.4.9].

**1.4.10 Theorem (Arcelà-Ascoli).** *Let $X, Y$ be metric spaces, where $X$ is separable and $Y$ is proper. Let $f_n : X \to Y$ be a uniformly equicontinuous sequence of maps that is pointwise bounded.*[14] *Then there exists a subsequence of $(f_n)$ that converges uniformly on compact subsets of $X$ to a uniformly continuous map $f : X \to Y$.*

We will use the following result for curves.

**1.4.11 Corollary (Arcelà-Ascoli for curves).** *Let $\gamma_n$ be a sequence of curves into a compact metric space $X$. If the lengths of the $\gamma_n$ are uniformly bounded then they possess a uniformly convergent subsequence.*

**Proof.** For each $\gamma_k$ there is a unique constant speed parametrization on the unit interval $[0,1]$ and by uniform boundedness of their lengths the speeds of these parametrizations are uniformly bounded. But this implies

$$d(\gamma_k(t), \gamma_k(t')) \le L(\gamma_k, t, t') \le C|t - t'| \tag{1.4.7}$$

for all $k$, all $t, t' \in [0,1]$, and some $C > 0$. So the $\gamma_k$ are uniformly equicontinuous. Furthermore, by compactness of $X$ the sequences $(\gamma_k(t))_k$ are bounded and $X$ is proper, so Theorem 1.4.10 applies to give a uniformly convergent subsequence. □

Although we have already informally used the term shortest path we now give an official definition of this key notion.

[12] Sometimes this property is just called equicontinuity. We prefer, however, to reserve the term equicontinuity at some point for the version of condition1.4.6 where one point is fixed.

[13] E.g., in [BBI01].

[14] That is, $(f_n(x))_n$ is bounded in $Y$ for each $x \in X$.

**1.4.12 Definition (Shortest path).** *A curve $\gamma : [a,b] \to X$ is called a shortest path if $L(\gamma) \leq L(\tilde{\gamma})$ for any curve $\tilde{\gamma}$ connecting $\gamma(a)$ to $\gamma(b)$.*

It is a trivial but noteworthy fact that the restriction of a shortest path to any subinterval is still a shortest path. In length spaces we have the following reformulation of Definition 1.4.12: $\gamma : [a,b] \to X$ is a shortest path if and only if

$$L(\gamma) = d(\gamma(a), \gamma(b)). \tag{1.4.8}$$

Hence one often refers to shortest paths in length spaces as *distance minimizers* or *distance realizing* paths. They enjoy some properties which do *not* hold for shortest paths in general metric spaces, e.g. the following one.

**1.4.13 Proposition (Convergence of distance minimizers).** *If a sequence of distance minimizers $\gamma_n$ in a length space $(X,d)$ converges pointwise to a path $\gamma$, then $\gamma$ is also a distance minimizer.*

**Proof.** Since the endpoints of the $\gamma_n$ converge to the endpoints $x, y$ of $\gamma$ and all $\gamma_n$ are shortest paths we have $L(\gamma_n) \to d(x,y)$. Now by the lower semicontinuity of length 1.2.8 we have

$$L(\gamma) \leq \lim_n L(\gamma_n) = d(x,y) \tag{1.4.9}$$

and we are done. □

**1.4.14 Example (Counterexample to 1.4.13 in metric spaces).** Consider the 'fan', which is not a length space (1.1.16(iii)) and the curves connecting the endpoints of two consecutive segments in the 'fan', see Figure 1.6. Then the limiting curve is twice the segment $[(0,0),(1,0)]$ with $L = 2$ but the endpoints converge both to $(0,1)$ and hence their distance converges to zero.

We now approach the question of existence of shortest paths. We have already discussed below Example 1.1.16 that in a length space there need not be any shortest path between a given pair of points, e.g. $\mathbb{R}^2 \setminus \{0\}$. Also if there is a shortest path, it need not be unique. Just consider antipodal points on a sphere.
Given a pair of points $x$, $y$ in a metric space we will denote a shortest path connecting them by $[x,y]$. This notion, of course, is only convenient in a situation where shortest paths are unique or where the actual choice of an specific shortest path is not relevant. Also this notation fits with the one applied to segments in Euclidean space since there segments are (unique) shortest paths. We now have the following basic existence result:

**1.4.15 Proposition (Existence of shortest paths, 1).** *Let $(X,d)$ be a proper metric space. Then for any two points $x$, $y \in X$ which can be connected by a rectifiable path at all, there is also a shortest path connecting them.*

**Proof.** Let $L$ be the length of some rectifiable path connecting $x$ and $y$. This path and any shorter path are hence contained in the closed ball $\bar{B}_L(x)$, which is compact by properness of $X$. We now exclusively work in the compact metric space $(\bar{B}_L(x), d|_{\bar{B}_L(x)})$.
Let $L'$ be the infimum of the lengths of rectifiable paths connecting $x$ and $y$. Then there is a sequence $\gamma_k$ of connecting paths with $L(\gamma_k) \to L'$. Hence their lengths are uniformly bounded and Corollary 1.4.11 applies to provide a convergent subsequence, again denoted by $\gamma_k$, and a limit path $\gamma$. Clearly $\gamma$ connects $x$ and $y$ and by the lower semi-continuity of the length (Proposition 1.2.8) we obtain

$$L(\gamma) \leq \lim L(\gamma_k) = L', \tag{1.4.10}$$

hence the result $L' = L(\gamma)$. □

The final building block in establishing the main result of this section is again about (topological) properties of length spaces that do not hold in general. Recall that a metric space is called *locally compact* if every point possesses a compact neighbourhood.

**1.4.16 Proposition (The Heine-Borel property in length spaces).** *Every complete and locally compact length space is proper.*

This proposition fails to hold e.g. in discrete metric spaces (i.e., the distance of all pairs of distinct points equals 1): Such a space is locally compact and complete but the closed unit ball is not compact unless the space is finite.

Next we collect some basic facts on balls in length spaces, which will be used repeatedly in the following. Recall that while in $\mathbb{R}^n$ the boundary of open balls are the spheres, i.e., $\partial B_r(x) = S_r(x) := \{y : d(x, y) = r\}$ and the closures of the open balls are the closed ones, i.e. $\overline{B_r(x)} = \bar{B}_r(x) := \{y : d(x, y) \leq r\}$, this is not true in general. Indeed in the discrete space we have $B_1(x) = \{x\}$ and $\overline{B_1(x)} = \{x\} \subsetneq \bar{B}_1(x) = X$ and $\partial B_1(x) = \emptyset \subsetneq S_1(x) = X \setminus \{x\}$. However, these phenomena do not occur in length spaces, as we prove next.

**1.4.17 Lemma (Balls in length spaces).** *In a length space $(X, d)$ we have for all $x \in X$, $r, R > 0$:*

*(i)* $\overline{B_r(x)} = \bar{B}_r(x)$ *and* $\partial B_r(x) = S_r(x)$

*(ii)* $B_r(\bar{B}_R(x)) = B_{R+r}(x)$.

*(iii)* $\bar{B}_r(\bar{B}_R(x)) = \bar{B}_{R+r}(x)$.

**Proof.** (i): $\subseteq$ in both cases holds true in any metric space: For $y \in \overline{B_r(x)}$ there is $y_n \to y$ with $d(x, y_n) < r$ and so $d(x, y) \leq r$. Moreover, if $y \in \partial B_r(x)$ then $d(x, y) \geq r$ and so $d(x, y) = r$.

$\supseteq$: Let $y \in \bar{B}_r(x)$ and suppose w.l.o.g. that $d(x, y) = r$. We construct a sequence of points $y_n \in B_r(x)$ with $y_n \to y$: Since $X$ is a length space we can choose paths $\gamma_n : [0, 1] \to X$ connecting $x$ and $y$ with $L(\gamma_n) =: r_n \leq r + 1/(2n)$. Choose $t_n \in [0, 1]$ such that $L(\gamma_n|_{[0,t_n]}) = r_n - 1/n$ and set $y_n = \gamma_n(t_n)$. Then we have

$$d(x, y_n) \leq L(\gamma_n|_{[0,t_n]}) = r_n - \frac{1}{n} \leq r + \frac{1}{2n} - \frac{1}{n} < r \tag{1.4.11}$$

and so $y_n \in B_r(x)$. Moreover,

$$d(y_n, y) \leq L(\gamma_n|_{[t_n,1]}) = r_n - L(\gamma_n|_{[0,t_n]}) = r_n - \Big(r_n - \frac{1}{n}\Big) = \frac{1}{n}, \tag{1.4.12}$$

and so $y_n \to y$ and $y \in \overline{B_r(x)}$. Finally, $\partial B_r(x) = \overline{B_r(x)} \setminus B_r(x) = \bar{B}_r(x) \setminus B_r(x) = S_r(x)$.

(ii) $\subseteq$ holds in any metric space. Let $z \in B_r(\bar{B}_R(x))$, then by definition there exists some $y \in \bar{B}_R(x)$ with $d(z, y) < r$ and so $d(x, z) \leq d(x, y) + d(y, z) < R + r$.

$\supseteq$: Let $z \in B_{R+r}(x)$ and assume without loss of generality that $z \notin \bar{B}_r(x)$. There is a curve $\gamma$ from $x$ to $z$ with $L(\gamma) < R + r$. Now call $\gamma_1$ the part of $\gamma$ starting at $x$ up to its first intersection with $\partial B_R(x)$, see Figure 1.20. Then $L(\gamma_1) \geq R$. Writing $\gamma_2$ for the remainder of $\gamma$ we have

$$d(z, \bar{B}_R(x)) \leq L(\gamma_2) = L(\gamma) - L(\gamma_1) < R + r - R = r. \tag{1.4.13}$$

(iii) Analogous to (ii), only choosing $\gamma$ from $x$ to $z$ with $L(\gamma) < R + r + \eta$ for any given $\eta > 0$. □

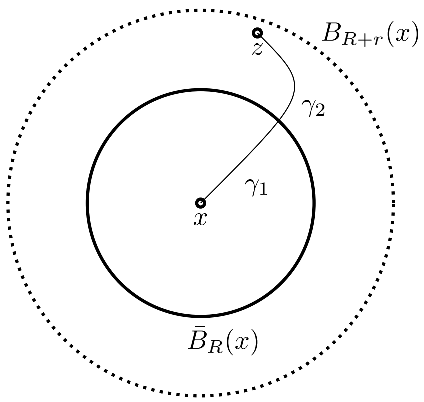



Figure 1.20: The construction used in the proof of Lemma 1.4.17(ii)

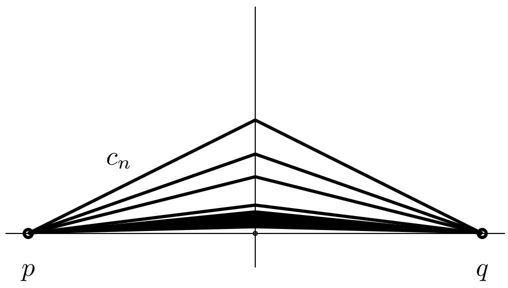



Figure 1.21: The space of Example 1.4.19.

To prepare the proof of Proposition 1.4.16, recall from topology (e.g. [Wil04, Thm. 39.9]) that a metric space $X$ is compact iff it is complete and totally bounded. The latter property (often called precompactness) means that for any $\varepsilon > 0$ there are finitely many open balls of radius $\varepsilon$ that cover the space. More formally, for any $\varepsilon > 0$ there is a finite set of points called an *$\varepsilon$-mesh* (or $\varepsilon$-net) $S = \{x_1, \ldots, x_k\}$ such that $\bigcup_i B_\varepsilon(x_i) = X$.

**Proof of Proposition 1.4.16.** First observe that it suffices to prove that every closed ball is compact and furthermore that if the closed ball $\overline{B}_r(x)$ is compact, so is every $\overline{B}_{r'}(x)$ with $r' \leq r$. To begin with let $x \in X$ be arbitrary and define

$$R = \sup\{r > 0 : \ \overline{B}_r(x) \quad \text{is compact}\}. \tag{1.4.14}$$

Since $x$ has a compact neighbourhood we have $R > 0$ and we will prove the proposition by showing that $R = \infty$. We assume by contradiction that $R < \infty$ and proceed in two steps.

(1) We prove that $B := \overline{B}_R(x)$ is compact.
Since $B$ is closed it suffices to show that for any $\varepsilon > 0$ there is an $\varepsilon$-mesh. Assuming w.l.o.g. that $\varepsilon < R$ we set $B' := \overline{B}_{R-\varepsilon/3}(x)$. Since $B'$ is compact there is an $\varepsilon/3$-mesh $S$ for $B'$.
Next we apply Lemma 1.4.17(iii) to obtain

$$B = \bar{B}_R(x) = \bar{B}_{\varepsilon/3}(\bar{B}_{R-\varepsilon/3}(x)) = \bar{B}_{\varepsilon/3}(B'), \tag{1.4.15}$$

and so for any $y \in B$ we have $d(y, B') \leq \varepsilon/3$.
Hence there is a point $y' \in B'$ with $d(y, y') < \varepsilon/2$. Also $d(y', S) < \varepsilon/2$ and so $d(y, S) < \varepsilon$ and we have established compactness of $B$.

(2) We derive a contradiction to the finiteness of $R$.
Since every $x \in X$ has a compact neighbourhood $U_x$ we can cover $B$ by a finite subcover of $\bigcup_{x\in B} U_x$ providing us with a compact neighbourhood $U$ of $B$. Since for compact sets $A$ the $B_\varepsilon(A)$ form a basis of neighbourhoods ([Die69, 3.17.11]) there is some $\varepsilon$ with $B_\varepsilon(B) \subseteq U$. Now again applying Lemma 1.4.17(ii) we have $B_\varepsilon(B) = B_{R+\varepsilon}(x)$, and so $\overline{B}_{R+\varepsilon}(x)$ is a closed subset of the compact set $U$ and hence compact itself. This contradicts the definition of $R$. □

Now we combine the Propositions 1.4.15 and 1.4.16 to immediately obtain the main result.

**1.4.18 Theorem (Existence of shortest paths, 2).** *Let $(X, d)$ be a complete and locally compact length space. Then for any two points $x$, $y \in X$ that can be connected by a rectifiable path at all, there is also a shortest path connecting them.*

The example of $\mathbb{R}^2 \setminus \{0\}$ shows that completeness is essential in the above theorem. So is local compactness, as we show now, ending this section.

**1.4.19 Example (A complete but not strictly intrinsic length space).** We consider the subspace $X \subseteq \mathbb{R}^2$ consisting of $p = (-1, 0)$ and $q = (1, 0)$ and all connecting broken straight lines $c_n$ that go via the point $(0, 1/n)$ $(n \in \mathbb{N})$ (see Figure 1.21) with its intrinsic metric coming from the usual length structure of $\mathbb{R}^2$. Then this space is complete but not locally compact, since no neighbourhood of $p$ or $q$ is compact. Indeed such a neighbourhood can be covered by small pieces of each $c_n$, which possess no finite subcover. Clearly there is no path from $p$ to $q$ realizing the distance, which is $2 = \inf L(c_n)$.

### 1.4.3 The Hopf-Rinow theorem

In this section we will state and prove a generalization of the well-known Hopf-Rinow theorem of Riemannian geometry ([KS20, Th. 2.4.2] or [O'N83, 5 Thm. 21]) to length spaces. Roughly speaking this theorem connects completeness of a Riemannian manifold as a metric space (w.r.t. Riemannian distance) to geodesic completeness.

To formulate its generalization to our present setting we first introduce some more notation. To begin with we extend the notion of shortest paths to also include non-closed intervals as domains.

**1.4.20 Definition (Geodesics).** *Let $I$ be an arbitrary interval and $X$ a metric space and let $\gamma : I \to X$ be a path.*

*(i) $\gamma$ is called a* shortest path *(or sometimes a* minimal geodesic*), if its restriction to any subinterval $[a, b] \subseteq I$ is a shortest path (in the sense of Definition 1.4.12).*

*(ii) $\gamma$ is called a* geodesic*, if each point $t \in I$ is contained in a relatively open subinterval $J \subseteq I$ such that $\gamma|_J$ is a shortest path.*

Hence we may say that a geodesic is a curve that locally is a distance minimizer. Geodesics need not be shortest paths. Indeed great circle segments on the sphere are geodesics, but if they are longer that $\pi$ they are *not* minimizing. The example of the sphere also shows that shortest paths need not be unique: all great circles joining antipodal points are shortest paths. On the other hand shortest paths on the sphere are locally unique and this is a general feature in Riemannian geometry, see e.g. [O'N83, 5, Prop 16(2)]. However, this is not true in general length spaces.

**1.4.21 Example (Non-unique shortest paths on the cube).** We consider the surface of a cube in $\mathbb{R}^3$ with its intrinsic metric induced by the Euclidean metric of the ambient space. This is a complete and locally compact length space. Let $A$ and $B$ be points at a distance $\eta$ from the vertex $E$ along the diagonal of the top face and the right front edge respectively, see Figure 1.22. Then the path $\gamma$ from $A$ along the diagonal to the vertex $E$ and down the edge to $B$ has length $L(\gamma) = 2\eta$.

However, as we shall see, it is shorter to go via the edge and the front face. To determine the shortest such path $\alpha$ we first consider a path $\alpha_\varepsilon$ passing the edge at a point $P_\varepsilon$ at a distance $\varepsilon$ from the corner $E$. Using the law of cosines we calculate the length of $\alpha_\varepsilon$ to be

$$L(\alpha_\varepsilon) = \sqrt{\varepsilon^2 + \eta^2 - \sqrt{2}\varepsilon\eta} + \sqrt{\varepsilon^2 + \eta^2}, \tag{1.4.16}$$

which becomes minimal for $\varepsilon_0 = (\sqrt{2} - 1)\eta$. We write $\alpha := \alpha_{\varepsilon_0}$ and $P := P_{\varepsilon_0}$ and calculate the length of the shortest path connecting $A$ and $B$ via the front edge and face to be

$$L(\alpha) = \eta(1 + \sqrt{2})\sqrt{2 - \sqrt{2}} = \eta\sqrt{2 + \sqrt{2}}. \tag{1.4.17}$$

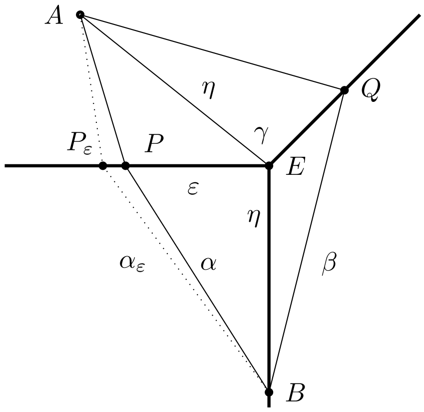



Figure 1.22: The two shortest paths $\alpha$ and $\beta$ connecting a point $A$ on the diagonal of the top face with a point $B$ on the front right edge via $P$ resp. $Q$.

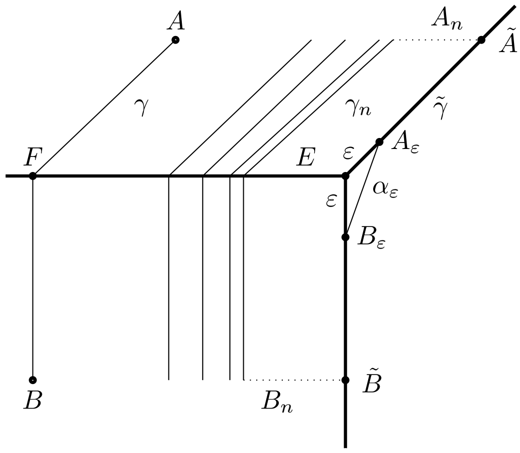



Figure 1.23: The geodesics $\gamma_n$ converge to the curve $\tilde{\gamma}$ which is not a geodesic. In fact, the vertex $E$ has no neighbourhood where $\tilde{\gamma}$ is a shortest curve— the curve $\alpha_\varepsilon$ on the right face is always shorter between any two points $A_\varepsilon$, $B_\varepsilon$ on $\tilde{\gamma}$.

By symmetry there is also a path $\beta$ connecting $A$ and $B$ via the point $Q$ at a distance $\varepsilon_0$ from $E$ on the right upper edge, which is of the same length as $\alpha$. Hence we obtain

$$L(\alpha) = L(\beta) = \eta\sqrt{2+\sqrt{2}} < 2\eta = L(\gamma), \tag{1.4.18}$$

and so we have found two shortest paths connecting $A$ and $B$. Moreover, since we clearly can make $\eta$ as small as we wish, we have established that the vertex $E$ has no neighbourhood where shortest paths are unique.

Another natural conjecture about shortest paths turns out to be wrong on the cube as well: The limit of geodesics in general fails to be a geodesic—contrary to the situation for shortest paths, cf. Proposition 1.4.13.

**1.4.22 Example (Limit of geodesics on the cube).** We again consider the cube of Example 1.4.21 and fix the side length to be 1. We now look at the straight path $\gamma$ connecting the midpoint $A$ of the top face to the midpoint $B$ of the front face, see Figure 1.23. This is a geodesic of length $L(\gamma) = 1$. Indeed any point on the curve that lies on either of the two faces obviously has a neighbourhood where $\gamma$ is minimizing. Moreover, the same holds true for the point $F$ on the edge. Now consider the sequence of points $A_n$ and $B_n$ on the top and front face respectively that approach the right edge of the cube at a distance of $1/2$ from the front edge. Denote the corresponding connecting curves by $\gamma_n$, again see Figure 1.23. Then the $\gamma_n$ are again geodesics of length $L(\gamma_n) = 1$. They converge to the curve $\tilde{\gamma}$ connecting the midpoints of the edges $\tilde{A}$ and $\tilde{B}$ along the edges via the vertex $E$. However, this curve is *not a geodesic* since $E$ possesses no neighbourhood where $\tilde{\gamma}$ is a shortest path. Indeed, the points $A_\varepsilon$ and $B_\varepsilon$ on $\tilde{\gamma}$ at distance $\varepsilon$ from $E$ can be joined by a path $\alpha_\varepsilon$ on the right face, which is of length $L(\alpha_\varepsilon) = \sqrt{2}\varepsilon$. Hence $\alpha_\varepsilon$ clearly is shorter than $\tilde{\gamma}$ between $A(\varepsilon)$ and $B(\varepsilon)$, which is of length $2\varepsilon$.
Now you might start worrying that this even is a counterexample to Proposition 1.4.13. This is, however, *not* the case: The curves $\gamma_n$ are not distance minimizing for $n$ large enough. More precisely, let us consider points $A_{\eta,\delta}$ and $B_{\eta,\delta}$ which have $x$-distance $\delta$ from the right face and distance $\eta$ from the top edge, see Figure 1.24. Then we can connect $A_{\eta,\delta}$ and $B_{\eta,\delta}$ via the geodesic $\gamma_{\eta,\delta}$ of length $L(\gamma_{\eta,\delta}) = 2\eta$ (clearly independent of $\delta$).

But we can also connect them by a curve 'around the corner', that is, around the vertex $E$ via the right face of the cube. We first want to detect the shortest such path and then compute its length. To this end, we consider the points $A_\varepsilon$ and $B_\varepsilon$ at distance $\varepsilon$ from the vertex $E$ along the edges and the path $\alpha^\varepsilon_{\eta,\delta}$ from $A_{\eta,\delta}$ to $B_{\eta,\delta}$ via $A_\varepsilon$, $B_\varepsilon$. We find its length to be

$$L(\alpha^\varepsilon_{\eta,\delta}) = 2\sqrt{\delta^2 + (\eta-\varepsilon)^2} + \sqrt{2}\varepsilon. \tag{1.4.19}$$

Now $L(\alpha^\varepsilon_{\eta,\delta})$ takes its minimum for $\varepsilon_0 = \eta - \delta$ and we set $A' := A_{\varepsilon_0}$, $B' := B_{\varepsilon_0}$, and $\alpha_{\eta,\delta} := \alpha^{\varepsilon_0}_{\eta,\delta}$. Also we easily find that $L(\alpha_{\eta,\delta}) = \sqrt{2}(\eta+\delta)$.

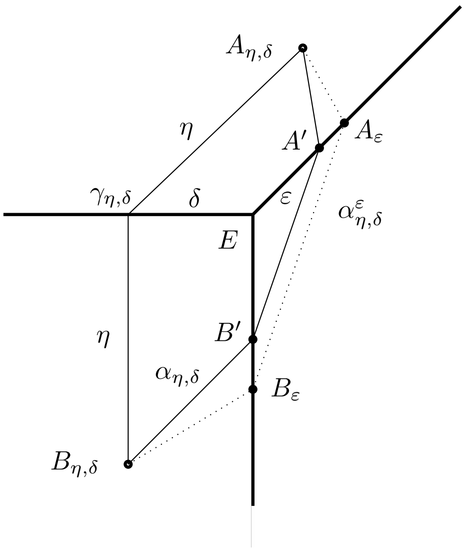



Figure 1.24: The curve $\alpha_{\eta,\delta}$ 'around the corner' is shorter than that the curve $\gamma_{\eta,\delta}$, if $\eta > (1+\sqrt{2})\delta$.

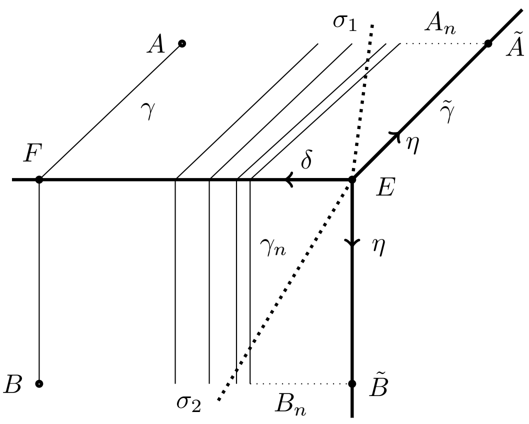



Figure 1.25: The lines $\sigma_1$, $\sigma_2$ bound the region where the distance $\eta$ from the top edge is smaller than $(1+\sqrt{2})\delta$, with $\delta$ the distance from the right edge. Inside that region the geodesics $\gamma_n$ are also shortest curves.

Now we can solve the following issue: Given a distance $\delta$ from the right face of the cube we determine the distance $\eta$ from the front edge for which the curve $\gamma_{\eta,\delta}$ ceases to be a shortest path, that is $L(\gamma_{\eta,\delta}) > L(\alpha_{\eta,\delta})$. Indeed we have

$$L(\gamma_{\eta,\delta}) = 2\eta > L(\alpha_{\eta,\delta}) = \sqrt{2}(\eta+\delta) \;\Leftrightarrow\; \eta > (1+\sqrt{2})\delta. \tag{1.4.20}$$

That is, there is a neighbourhood of the front edge of the cube bounded by the line $\sigma_1$ on the top face and the line $\sigma_2$ on the front face, where the curves $\gamma_{\eta,\delta}$ are local minimizers, see Figure 1.25. Outside of that neighbourhood the curves $\alpha_{\eta,\delta}$ passing around the corner $E$ on the right face are shorter.
Finally we see that the curves $\gamma_n$ from $A_n$ to $B_n$ from the beginning of this example, while being geodesics and shortest curves for small $n$, cease to be shortest curves for large $n$.

Finally, we motivate the main result of this section, respectively its formulation. Intuitively a space $X$ fails to be complete if a point is 'missing', as e.g. in $\mathbb{R}^2 \setminus \{0\}$. However, intuitively the removal of a point would be noticed by geodesics passing through that point. Indeed such a geodesic would be defined on a maximal open interval $\neq \mathbb{R}$ and could not be extended further. The precise formulation of this idea is as follows.

**1.4.23 Theorem (Hopf-Rinow-Cohn-Vossen).** *For a locally compact length space $X$ the following four conditions are equivalent:*

*(i)* *X is complete.*

*(ii)* *X is proper.*

*(iii)* *Every rectifiable geodesic* $\gamma : [0, a) \to X$ *can be extended to a continuous path* $\bar{\gamma} : [0, a] \to X$.

*(iv)* *There is a point* $x \in X$ *such that every rectifiable geodesic* $\gamma : [0, a) \to X$ *with* $\gamma(0) = x$ *can be extended to a continuous path* $\bar{\gamma} : [0, a] \to X$.

**Proof.** (ii) $\Rightarrow$ (i) is a standard argument valid in any metric space: The point set of a Cauchy sequence in $X$ is bounded and by (ii) its closure is compact. Hence the sequence possesses a convergent subsequence, and being Cauchy, converges itself.

(i) $\Rightarrow$ (iii) By the remark following the proof of Proposition 1.4.6 we may assume that $\gamma$ is naturally parametrized. Since $\gamma$ is rectifiable, $a = L(\gamma) < \infty$ and we pick any sequence $a_n \nearrow a$. Then $\gamma(a_n)$ is a Cauchy sequence and converges by completeness. Moreover the limit is unique as is seen from mixing two sequences $a_n$, $b_n \nearrow a$.

(iii) $\Rightarrow$ (iv) is trivial, and we are left with proving

(iv) $\Rightarrow$ (ii): The general lay out of the argument is similar to the one used in the proof of Proposition 1.4.16. However, the details are more complicated, since instead of using completeness we have to use condition (iv).
Let $x$ be the point in condition (iv). Since $X$ is locally compact, small closed balls $\bar{B}_r(p)$ are compact and we define

$$R := \sup\{r : \ \bar{B}_r(p) \quad \text{is compact}\}. \tag{1.4.21}$$

We indirectly assume $R$ to be finite and derive at a contradiction. Observe that this indeed proves (ii): For any arbitrary $y \in X$ the closed balls $\bar{B}_r(y)$ are contained in a large closed and compact ball centered at $x$ and hence are compact themselves.

Again proceed in two steps:

(1) We prove that the closed ball $\bar{B}_R(p)$ is compact. To do so we will prove:

$$\text{any sequence } x_n \in B_R(p) \text{ has a subsequence converging in } X \text{ to some } x \in \bar{B}_R(p) \tag{1.4.22}$$

This really suffices: Let $y_n$ be a sequence in $\bar{B}_R(p)$. Then for all $n$ there exists $x_n \in B_R(p)$ with $d(x_n, y_n) < 1/n$. By (1.4.22) $x_n$ has a converging subsequence $x_{n_k} \to x \in \bar{B}_R(p)$ and so the subsequence $y_{n_k} \to x$ as well, and $\bar{B}_R(p)$ is compact.

Now we prove (1.4.22): Set $r_n = d(p, x_n)$. We may assume that $r_n \nearrow R$ since otherwise a subsequence of $x_n$ is contained in a smaller closed ball $\bar{B}_r(p)$ with $r < R$, which is compact by assumption and hence $x_n$ possesses a convergence subsequence (even in $B_R(p)$). Moreover we may assume w.l.o.g. that $r_n$ is increasing.

By Proposition 1.4.15 there are shortest paths $\gamma_n : [0, r_n] \to X$ connecting $p$ to $x_n$. Indeed, since $x_n$ lies in some $\bar{B}_r(p)$ for $r < R$ which is compact and since $X$ is a length space, the proposition applies. Also we may assume the $\gamma_n$ to be naturally parametrized.
Next we restrict the $\gamma_n$ to $[0, r_1]$ and apply Corollary 1.4.11 to extract a converging subsequence. Restricting this subsequence to $[0, r_2]$ we may, again by Corollary 1.4.11 extract a converging subsequence. Now we proceed iteratively and apply the Cantor diagonal method. More precisely we pick the $n$-th element form the $n$-th subsequence to obtain a sequence $\gamma_{n_k}$ with the following property: for any $t \in [0, R)$ the point $\gamma_{n_k}(t)$ is defined for large enough $k$ and the sequence $\gamma_{n_k}(t)$ converges in $X$.
Define $\gamma(t) := \lim_k \gamma_{n_k}(t)$ $(0 \leq t < R)$. Then $\gamma : [0, R) \to X$ is Lipschitz and hence a path, since the $\gamma_{n_k}$ are naturally parametrized. Indeed, for all $\varepsilon > 0$ there is $k$ such that the estimate

$$d(\gamma(t), \gamma(s)) \leq d(\gamma(t), \gamma_{n_k}(t)) + d(\gamma_{n_k}(t), \gamma_{n_k}(s)) + d(\gamma_{n_k}(s), \gamma(s)) \leq 2\varepsilon + |t - s| \tag{1.4.23}$$

holds. Moreover, applying Proposition 1.4.13 on any $[0, r]$ with $r < R$ it follows that the length of $\gamma$ is less than $R$ on any such subinterval, hence so is its total length. The same result also implies that $\gamma$ is a geodesic.
Consequently, condition (iv) applies to $\gamma$ and there is a continuous extension $\bar{\gamma} : [0, R] \to X$ of $\gamma$. Finally, the points $x_{n_k}$ that are the endpoints $\gamma_{n_k}(r_{n_k})$ of the converging curves $\gamma_{n_k}$ converge to $\bar{\gamma}(R)$ by the following argument: For any $\varepsilon > 0$ choose $r < R$ such that $R - r < \varepsilon/3$ and $d(\bar{\gamma}(r), \bar{\gamma}(R)) < \varepsilon/3$. Then choose $k_0$ such that for $k \geq k_0$ we have $r_{n_k} \in (r, R)$ and $d(\gamma_{n_k}(r), \bar{\gamma}(r)) < \varepsilon/3$. Then

$$d(\gamma_{n_k}(r_{n_k}), \gamma_{n_k}(r)) \leq |r - r_{n_k}| \leq R - r \leq \frac{\varepsilon}{3}, \tag{1.4.24}$$

and so

$$\begin{aligned} &d(\gamma_{n_k}(r_{n_k}), \bar{\gamma}(R)) \\ &\quad \leq d(\gamma_{n_k}(r_{n_k}), \gamma_{n_k}(r)) + d(\gamma_{n_k}(r), \bar{\gamma}(r)) + d(\bar{\gamma}(r), \bar{\gamma}(R)) \leq \frac{\varepsilon}{3} + \frac{\varepsilon}{3} + \frac{\varepsilon}{3} = \varepsilon. \end{aligned} \tag{1.4.25}$$

(2) We derive a contradiction just as in step (2) of the proof of Proposition 1.4.16: Since $X$ is locally compact, every $x \in X$ has a compact neighbourhood $U_x$ and we cover the compact closed ball $\bar{B}_R(p)$ by a finite collection of these to obtain a compact neighbourhood $U$ of $\bar{B}_R(p)$. $U$ in turn contains a $B_\varepsilon(B) = B_{R+\varepsilon}(x)$. So $\bar{B}_{R+\varepsilon}(x)$ is a closed subset of $U$, hence compact. This contradicts the definition of $R$. □

## 1.5 Length and Hausdorff measure

Since the length of a curve in a metric space $(X, d)$ is independent of its parametrization one may expect that it should be feasible to determine it merely from knowledge of the image of the curve as a subset of $X$. In this section we show that this is indeed the case. In fact, the length of the curve equals the (one dimensional) Hausdorff measure of its image. We therefore start out by recalling some basic facts about Hausdorff measures (cf., e.g., [Els05]).

**1.5.1 Definition.** *Let $(X, d)$ be a metric space and let $r \geq 0$ be a real number. For any countable cover $(S_i)_{i \in I}$ of $X$, we define its $r$-weight $w_r((S_i))$ by*

$$w_r((S_i)) := \sum_i (\text{diam} S_i)^r,$$

*where we set $0^0 := 1$. Then for any $\varepsilon > 0$ set*

$$\mu_{r,\varepsilon}(X) := \inf\{w_r((S_i)) : \text{diam}(S_i) < \varepsilon \ \forall i \in I\},$$

*where the infimum is taken over all countable coverings of $X$ by sets of diameter smaller than $\varepsilon$. If no such covering exists then the infimum is set to be $+\infty$. Finally, the $r$-dimensional Hausdorff measure of $X$ is defined as*

$$\mu_r(X) := C(r) \lim_{\varepsilon \to 0} \mu_{r,\varepsilon}(X). \tag{1.5.1}$$

*Here, the normalization constant $C(r)$ is chosen in such a way that for integer $r$ the $r$-dimensional Hausdorff measure of the unit cube in $\mathbb{R}^r$ equals 1. Moreover, we set $\mu_r(\emptyset) = 0$ for each $r \geq 0$.*

Note that $\mu_{r,\varepsilon_1}(X) \geq \mu_{r,\varepsilon_2}(X)$ for $\varepsilon_1 < \varepsilon_2$, so the limit in Definition 1.5.1 always exist (in $[0, \infty]$). Also, for integer $r$, $\mu_r$ is precisely the Lebesgue measure on $\mathbb{R}^r$.

**1.5.2 Lemma.** *For any connected metric space $X$, $\mu_1(X) \geq \text{diam}(X)$.*

**Proof.** We first note that, in the definition of $\mu_1(X)$, it suffices to only consider open coverings $(S_i)$ of $X$. In fact, let $(S_i)$ be any covering and consider the open covering $(S_i')$, where

$$S_i' := U_{\delta/2^i} := \{x \in X : \operatorname{dist}(x, S_i) < \delta/2^i\}$$

for some $\delta > 0$. Then $\operatorname{diam}(S_i') \leq \operatorname{diam}(S_i) + 2\delta/2^i$ and therefore $w_1((S_i')) \leq w_1((S_i)) + 2\delta$. Since $\delta > 0$ can be chosen arbitrarily small, the claim follows.

Thus let $(S_i)$ be an open covering of $X$. Since $X$ is connected, for any $x, y \in X$ there exists a finite sequence $S_{i_1}, \dots, S_{i_n}$ such that $x \in S_{i_1}$, $y \in S_{i_n}$, and $S_{i_k} \cap S_{i_{k+1}} \neq \emptyset$ for all $1 \leq k \leq n-1$: In fact, fix any $x \in X$ and denote by $Y$ the set of all $y \in X$ for which such a chain of sets in $(S_i)$ exists. Then for any $S_i$, either $S_i \subseteq Y$ or $S_i \subseteq X \setminus Y$, so both $Y$ and $X \setminus Y$ are open. Consequently, $X = Y$.

Now let $(S_i)$ be an open covering of $X$, $x, y \in X$ and $S_{i_1}, \dots, S_{i_n}$ a connecting chain as above. For $k = 1, \dots, n-1$, let $x_k \in S_{i_k} \cap S_{i_{k+1}}$, and set $x_0 := x$, $x_n := y$. Then $d(x_{k-1}, x_k) \leq \operatorname{diam}(S_{i_k})$ for all $k = 1, \dots, n-1$. Therefore,

$$\sum \operatorname{diam}(S_i) \geq \sum_{k=1}^{n} \operatorname{diam}(S_{i_k}) \geq \sum_{k=1}^{n} d(x_{k-1}, x_k) \geq d(x_0, x_n) = d(x, y).$$

It follows that, first, $w_1((S_i)) \geq d(x, y)$, and a fortiori $\mu_1(X) \geq d(x, y)$. Since $x, y$ were arbitrary, this gives the claim. □

The desired result now is as follows:

**1.5.3 Theorem.** *Let $X$ be a metric space and let $\gamma : [a, b] \to X$ be a rectifiable simple curve. Then $L(\gamma) = \mu_1(\gamma([a, b]))$.*

**Proof.** Let $S := \gamma([a, b])$ and $L := L(\gamma)$. We may assume that $\gamma$ is parametrized by arclength, so $\gamma : [0, L] \to X$. For any $N \in \mathbb{N}$, $S$ is covered by the sets $\gamma([i\frac{L}{N}, (i+1)\frac{L}{N}])$, $i = 0, \dots, N-1$. Each of these sets has diameter bounded above by the length of $\gamma|_{[i\frac{L}{N}, (i+1)\frac{L}{N}]}$, which is $L/N$. Thus the sum of these diameters is $\leq L$. Since the diameter of each covering set goes to 0 as $N \to \infty$, it follows from (1.5.1) that $\mu_1(S) \leq L(\gamma)$.

Conversely, let $a = t_0 \leq t_1 \leq \cdots \leq t_n = b$ be any partition of $[a, b]$ and let $S_i := \gamma([t_i, t_{i+1}])$ $(0 \leq i \leq n-1)$. Since $\gamma$ was supposed to be simple, the sets $S_i$ are disjoint up to the finitely many points $\gamma(t_i)$. Also the Hausdorff measure of a single point is 0, so $\mu_1(S) = \sum_{i=0}^{n-1} \mu_1(S_i)$. Now Lemma 1.5.2 implies $\mu_1(S_i) \geq \operatorname{diam}(S_i) \geq d(\gamma(t_i), \gamma(t_{i+1}))$. As the partition was chosen arbitrarily, this gives $\mu_1(S) \geq L(\gamma)$. □

**1.5.4 Remark.** If $\gamma$ is not supposed to be simple, the above proof still implies $L(\gamma) \geq \mu_1(\gamma([a, b]))$.

## 1.6 Length and Lipschitz speed

It is a well-known fact from analysis that the variational length of a Lipschitz curve $\gamma : [a, b] \to \mathbb{R}^n$ can be calculated as the integral of its speed:

$$L(\gamma) = \int_a^b \|\dot{\gamma}(t)\| \, dt$$

and we will in fact establish this in more generality on Riemannian manifolds in Chapter 4 below. The aim of the present section is to show that a similar description is true even in general metric spaces. The velocity of a curve in this context is measured as follows:

**1.6.1 Definition.** *Let $(X, d)$ be a metric space and let $\gamma : I \to X$ be a curve. The speed of $\gamma$ at $t \in I$ is defined by its* metric derivative

$$v_\gamma(t) := \lim_{\varepsilon \to 0} \frac{d(\gamma(t+\varepsilon), \gamma(t))}{|\varepsilon|},$$

*if the limit exists.*

Below we shall prove that for a Lipschitz curve the metric derivative in fact exists almost everywhere. To do this we will require the following result from measure theory:

**1.6.2 Theorem.** *(Vitali's covering theorem) Let $X$ be a bounded subset of $\mathbb{R}^n$ and let $\mathcal{B}$ be a family of closed balls in $\mathbb{R}^n$ such that, for every $\varepsilon > 0$ and every $x \in X$, there exists a ball $B \in \mathcal{B}$ with $x \in B$ and $\mathrm{diam}(B) < \varepsilon$. Then $\mathcal{B}$ contains an at most countable sub-family $\{B_i\}$ of disjoint balls that covers $X$ up to a set of (Lebesgue) measure zero: $B_i \cap B_j = \emptyset$ for all $i \neq j$ and $\mu_n(X \setminus \bigcup_i B_i) = 0$.*

**Proof.** Without loss of generality we may assume that each $B \in \mathcal{B}$ is non-empty and of radius $\leq 1$. All such balls are contained in the 2-ball $U_2(X)$ around $X$, which has finite measure. Picking any $B_1 \in \mathcal{B}$, we proceed by induction. So suppose that (pairwise disjoint) $B_1, \dots, B_m$ have already been constructed and let $\mathcal{B}_m$ denote the set of all $B \in \mathcal{B}$ that do not intersect $B_1, \dots, B_m$. If $\mathcal{B}_m$ is empty then $X \subseteq \bigcup_{k=1}^m B_k$ and we are done: in fact, suppose that there existed some $x \in X \setminus \bigcup_{k=1}^m B_k$. As the right hand side is relatively open in $X$, there would then exist some ball $B_r(x)$ that does not intersect $\bigcup_{k=1}^m B_k$. But by assumption, $x$ is contained in some $B' \in \mathcal{B}$ with $\mathrm{diam}(B) < r$, so that

$$B' \subseteq B_r(x) \subseteq X \setminus \bigcup_{k=1}^{m} B_k,$$

contradicting the assumption that $\mathcal{B}_m$ is empty.
So let us suppose that $\mathcal{B}_m \neq \emptyset$. Then we can choose some $B_{m+1} \in \mathcal{B}_m$ such that

$$\mathrm{diam}(B_{m+1}) > \frac{1}{2} \sup\{\mathrm{diam}(B) : B \in \mathcal{B}_m\}, \tag{1.6.1}$$

in this way arriving at a countable collection of disjoint sets $B_i$ ($i \in \mathbb{N}$). It remains to show that they cover $X$ up to a set of zero measure. Note first that $\sum_{i=1}^\infty \mu_n(B_i) \leq \mu_n(U_2(X)) < \infty$. Fixing any $\varepsilon > 0$, we may therefore pick some $m \in \mathbb{N}$ such that $\sum_{i=m+1}^\infty \mu_n(B_i) < \varepsilon$. If $\bigcup_i B_i = X$ we are done. Otherwise, let $x \in X \setminus \bigcup_i B_i \subseteq X \setminus \bigcup_{i=1}^m B_i$ and, as above (and noting that only the last set here is relatively open in $X$), pick a ball $B \in \mathcal{B}$ that contains $x$ and does not intersect $B_1, \dots, B_m$. Then $B$ must intersect $\bigcup_{i=m+1}^\infty B_i$: otherwise it would follow that $B \in \mathcal{B}_m$ for all $m$, contradicting (via (1.6.1)) the fact that $\mu_n(B_i) \to 0$ for $i \to \infty$. Let $k$ be the minimal index with $B \cap B_k \neq \emptyset$ (so necessarily $k > m$).
Then $B \in \mathcal{B}_{k-1}$, so $\mathrm{diam}(B_k) > \frac{1}{2}\mathrm{diam}(B)$ by (1.6.1). Hence the distance from $x$ to the center of $B_k$ is less or equal than 5 times the radius of $B_k$ (note that $B$ and $B_k$ might only intersect at their boundaries). It follows that $x \in 5B_k$, the ball with the same center and 5 times the radius of $B_k$. Summing up, any $x \in X \setminus \bigcup_i B_i$ belongs to $5B_k$ for some $k > m$, i.e., $X \setminus \bigcup_i B_i \subseteq \bigcup_{i=m+1}^\infty (5B_i)$, so

$$\mu_n\Big(X \setminus \bigcup_i B_i\Big) \leq \sum_{i=m+1}^\infty \mu_n(5B_i) = 5^n \sum_{i=m+1}^\infty \mu_n(B_i) < 5^n \varepsilon.$$

As $\varepsilon$ was arbitrary, the claim follows. □

Now we can prove:

**1.6.3 Theorem.** *Let $(X, d)$ be a metric space, and let $\gamma : [a, b] \to X$ be a rectifiable curve. Then for almost every $t \in [a, b]$ either*

$$\liminf_{\varepsilon,\varepsilon' \to 0+} \frac{L(\gamma|_{[t-\varepsilon,t+\varepsilon']})}{\varepsilon + \varepsilon'} = 0, \tag{1.6.2}$$

*or*

$$\lim_{\varepsilon,\varepsilon' \to 0+} \frac{d(\gamma(t-\varepsilon), \gamma(t+\varepsilon'))}{L(\gamma|_{[t-\varepsilon,t+\varepsilon']})} = 1. \tag{1.6.3}$$

Note that each individual term in (1.6.3) is $\leq 1$, so (1.6.3) holds if and only if it holds with lim replaced by lim inf.

**Proof.** Suppose that the theorem is false and let, for any $\alpha > 0$, $Z_\alpha$ be the set of all $t \in [a,b]$ such that

$$\liminf_{\varepsilon,\varepsilon'\to 0+} \frac{L(\gamma|_{[t-\varepsilon,t+\varepsilon']})}{\varepsilon+\varepsilon'} > \alpha,$$

and

$$\liminf_{\varepsilon,\varepsilon'\to 0+} \frac{d(\gamma(t-\varepsilon),\gamma(t+\varepsilon'))}{L(\gamma|_{[t-\varepsilon,t+\varepsilon']})} < 1-\alpha.$$

Also, let $Z_0$ be the set of all $t \in [a,b]$ such that either (1.6.2) or (1.6.3) hold true. Then $Z_0 = [a,b] \setminus \bigcup_{\alpha>0} Z_\alpha = [a,b] \setminus \bigcup_{n\in\mathbb{N}} Z_{1/n}$, so our assumption requires that some $Z_\alpha$ (and thereby also any $Z_{\alpha'}$ with $\alpha' < \alpha$) must have nonzero Lebesgue measure: $\mu_1(Z_\alpha) > 0$. Fix any such $\alpha > 0$ and abbreviate $Z_\alpha$ by $Z$ and $\mu_1(Z)$ by $\mu$. By Lemma 1.2.5, we may choose $\varepsilon_0 > 0$ such that for any partition $a = y_0 < \cdots < y_N = b$ of $[a,b]$ of modulus less than $\varepsilon_0$ we have

$$L(\gamma) - \sum_{i=1}^{N} d(\gamma(y_{i-1}),\gamma(y_i)) < \frac{\mu\alpha^2}{2}. \tag{1.6.4}$$

Next, denote by $\mathcal{B}$ the family of all intervals of the form $[t-\varepsilon, t+\varepsilon']$ such that $t \in Z$, $\varepsilon+\varepsilon' < \varepsilon_0$,

$$L(\gamma|_{[t-\varepsilon,t+\varepsilon']}) > \alpha(\varepsilon+\varepsilon'), \tag{1.6.5}$$

and

$$d(\gamma(t-\varepsilon),\gamma(t+\varepsilon')) < (1-\alpha)L(\gamma|_{[t-\varepsilon,t+\varepsilon']}). \tag{1.6.6}$$

Then by definition of $Z_\alpha$, any point $t \in Z$ is contained in arbitrarily short elements of $\mathcal{B}$. Hence Vitali's covering theorem 1.6.2 shows that we can extract a countable sub-family $\{[t_i-\varepsilon_i, t_i+\varepsilon_i']\}_{i=1}^\infty$ of disjoint intervals that cover $Z$ up to a set of zero measure. Thus

$$\sum_{i=1}^{\infty}(\varepsilon_i+\varepsilon_i') = \mu_1\Big(\bigcup_i [t_i-\varepsilon_i, t_i+\varepsilon_i']\Big) \geq \mu_1(Z) = \mu.$$

We may therefore pick $M$ so large that

$$\sum_{i=1}^{M}(\varepsilon_i+\varepsilon_i') > \frac{\mu}{2}.$$

Since the intervals $\{[t_i-\varepsilon_i, t_i+\varepsilon_i']\}_{i=1}^M$ are disjoint, we can include their endpoints in a partition $\{y_j\}_{j=1}^N$ of modulus less than $\varepsilon_0$, by possibly adding extra points that do not lie in any of the $[t_i-\varepsilon_i, t_i+\varepsilon_i']$. Then setting $L_j := L(\gamma|_{[y_{j-1},y_j]})$ and $d_j := d(\gamma(y_{j-1}),\gamma(y_j))$, (1.6.4) gives

$$\sum_{j=1}^{N}(L_j-d_j) = L(\gamma) - \sum_{j=1}^{N} d_j < \frac{\mu\alpha^2}{2}. \tag{1.6.7}$$

All summands on the left hand side are non-negative, and whenever $[y_{j-1},y_j] \in \mathcal{B}$ (i.e., $y_{j-1} = t_i - \varepsilon_i$ and $y_j = t_i + \varepsilon_i'$ for some $i$) we have, using (1.6.5) and (1.6.6):

$$L_j - d_j > \alpha L_j > \alpha^2(y_j - y_{j-1}) = \alpha^2(\varepsilon_i+\varepsilon_i').$$

Consequently,

$$\sum_{j=1}^{N}(L_j-d_j) \geq \alpha^2 \sum_{i=1}^{M}(\varepsilon_i+\varepsilon_i') > \frac{\mu\alpha^2}{2},$$

contradicting (1.6.7). □

Inspection of the above proof shows that it works just as well if either $\varepsilon$ or $\varepsilon'$ is fixed to 0 throughout, so we obtain:

**1.6.4 Corollary.** *Let $(X,d)$ be a metric space, and let $\gamma : [a,b] \to X$ be a rectifiable curve. Then for almost every $t \in [a,b]$ either*

$$\liminf_{\varepsilon \to 0} \frac{L(\gamma|_{[t,t+\varepsilon]})}{|\varepsilon|} = 0,$$

*or*

$$\lim_{\varepsilon \to 0} \frac{d(\gamma(t), \gamma(t+\varepsilon))}{L(\gamma|_{[t,t+\varepsilon]})} = 1.$$

*(where the interval $[t, t+\varepsilon]$ should mean $[t+\varepsilon, t]$ for $\varepsilon < 0$).*

**1.6.5 Theorem.** *Let $(X,d)$ be a metric space and let $\gamma : [a,b] \to X$ be a Lipschitz curve. Then the speed $v_\gamma(t)$ of $\gamma$ exists for almost every $t \in [a,b]$, is integrable, and*

$$L(\gamma) = \int_a^b v_\gamma(t)\, dt.$$

**Proof.** By [Coh13, Cor. 6.3.8], any Lipschitz function $f : [a,b] \to \mathbb{R}$ is differentiable almost everywhere and satisfies the fundamental theorem of calculus, see Proposition 1.6.7 (i) and (iv) below. We wish to apply this result to the function $f : [a,b] \to \mathbb{R}$, $f(t) := L(\gamma|_{[a,t]})$. Since

$$\sum d(\gamma(t_i), \gamma(t_{i+1})) \leq \mathrm{Lip}(\gamma) \sum (t_{i+1} - t_i) = \mathrm{Lip}(\gamma)(s - r)$$

on any sub-interval $[r,s]$ of $[a,b]$, the very definition of curve length implies that $|f(r) - f(s)| \leq \mathrm{Lip}(\gamma)|r - s|$, i.e., $f$ is indeed Lipschitz, hence in particular differentiable almost everywhere. It also follows that $\gamma$ is rectifiable. We have

$$f'(t) = \lim_{\varepsilon \to 0} \frac{L(\gamma|_{[t,t+\varepsilon]})}{|\varepsilon|} = \lim_{\varepsilon \to 0} \frac{L(\gamma|_{[t,t+\varepsilon]})}{d(\gamma(t), \gamma(t+\varepsilon))} \cdot \frac{d(\gamma(t), \gamma(t+\varepsilon))}{|\varepsilon|}. \tag{1.6.8}$$

By Corollary 1.6.4, for almost all $t \in [a,b]$ either $f'(t) = 0$ or the first factor in the last product goes to 1 as $\varepsilon \to 0$. In the first case,

$$v_\gamma(t) = \lim_{\varepsilon \to 0} \frac{d(\gamma(t), \gamma(t+\varepsilon))}{|\varepsilon|} \leq \lim_{\varepsilon \to 0} \frac{L(\gamma|_{[t,t+\varepsilon]})}{|\varepsilon|} = f'(t) = 0,$$

so $v_\gamma(t) = 0$. In the second case,

$$v_\gamma(t) = \lim_{\varepsilon \to 0} \frac{d(\gamma(t), \gamma(t+\varepsilon))}{|\varepsilon|} = f'(t).$$

Consequently, for almost every $t \in [a,b]$, $v_\gamma(t)$ exists and equals $f'(t)$. The fundamental theorem of calculus cited above therefore implies

$$L(\gamma) = f(b) - f(a) = \int_a^b f'(t)\, dt = \int_a^b v_\gamma(t)\, dt.$$

□

Finally, we are going to extend the validity of Theorem 1.6.5 to an even bigger class of curves. We first recall:

**1.6.6 Definition.** *A curve $\gamma : [a,b] \to X$ into a metric space $(X,d)$ is called* absolutely continuous *if for any $\varepsilon > 0$ there exists some $\delta > 0$ such that, for any finite sequence of pairwise disjoint subintervals $\{(a_i, b_i) : 1 \leq i \leq m\}$ of $[a,b]$ with $\sum_{i=1}^m (b_i - a_i) < \delta$ we have $\sum_{i=1}^m d(\gamma(a_i), \gamma(b_i)) < \varepsilon$.*

Some useful properties of absolutely continuous maps are collected in the following result:

**1.6.7 Proposition.** *Let $\gamma : [a,b] \to (X,d)$.*

*(i) If $\gamma$ is Lipschitz, then it is absolutely continuous.*

*(ii) If $\gamma$ is absolutely continuous, then it is uniformly continuous.*

*(iii) If $\gamma$ is absolutely continuous and $F : (X, d) \to (Y, d_Y)$ is Lipschitz, then $F \circ \gamma$ is absolutely continuous.*

*(iv) A function $f : [a, b] \to \mathbb{R}$ is absolutely continuous if and only if it is differentiable Lebesgue-almost everywhere, $f'$ is integrable, and $f$ satisfies the fundamental theorem of calculus:*

$$f(x) = f(a) + \int_a^x f'(x)\,dx \quad (x \in [a, b]).$$

*(v) If $\gamma$ is absolutely continuous, it is rectifiable.*

**Proof.** (i)–(iii) are easy consequences of the definition of absolute continuity.
(iv) is [Coh13, Cor. 6.3.8].
(v) Pick $\delta > 0$ such that, for any finite sequence of pairwise disjoint subintervals $\{(a_i, b_i) : 1 \leq i \leq m\}$ of $[a, b]$ with $\sum_{i=1}^m (b_i - a_i) < \delta$ we have $\sum_{i=1}^m d(\gamma(a_i), \gamma(b_i)) < 1$. Let $M \in \mathbb{N}$ such that $\frac{b-a}{M} < \delta$ and set $a_n := a + n\frac{b-a}{M}$ for $n = 0, 1, \ldots, M$. Then $a = a_0 < a_1 < \cdots < a_m = b$ is a partition of $[a, b]$, and

$$L(\gamma) = \sum_{i=1}^{M} L(\gamma|_{[a_{i-1},a_i]}).$$

Now let $a_{i-1} = y_0 < y_1 < \cdots < y_k = a_i$ be any partition of $[a_{i-1}, a_i]$. Then

$$\sum_{j=0}^{k-1} (y_{j+1} - y_j) = a_i - a_{i-1} = \frac{b-a}{M} < \delta,$$

so $\sum_{j=0}^{k-1} d(\gamma(y_j), \gamma(y_{j+1})) < 1$. Taking the supremum over all such partitions it follows that $L(\gamma|_{[a_{i-1},a_i]}) \leq 1$, so

$$L(\gamma) = \sum_{i=1}^{M} L(\gamma|_{[a_{i-1},a_i]}) \leq M.$$

□

Now we are ready to extend the validity of Theorem 1.6.5:

**1.6.8 Theorem.** *Let $(X, d)$ be a metric space and let $\gamma : [a, b] \to X$ be an absolutely continuous curve. Then the speed $v_\gamma(t)$ of $\gamma$ exists for almost every $t \in [a, b]$, is integrable, and*

$$L(\gamma) = \int_a^b v_\gamma(t)\,dt.$$

**Proof.** If we can show that $f : [a, b] \to \mathbb{R}$, $f(t) := L(\gamma|_{[a,t]})$ is absolutely continuous, then using Proposition 1.6.7 (iv), the same proof as in Theorem 1.6.5 will give the result (the proof uses Theorem 1.6.3, which requires $\gamma$ to be rectifiable – this holds by Proposition 1.6.7 (v)). So let $\varepsilon > 0$ and pick $\delta > 0$ such that for any finite sequence of pairwise disjoint subintervals $\{(a_i, b_i) : 1 \leq i \leq m\}$ of $[a, b]$ with $\sum_{j=1}^m (b_j - a_j) < \delta$ we have $\sum_{j=1}^m d(\gamma(a_j), \gamma(b_j)) < \varepsilon$.
Now for each $j \in \{1, \ldots, m\}$ pick any partition $\{t_i^{(j)} : 0 \leq i \leq n_j\}$ of the interval $[a_j, b_j]$. Then $\sum_{i,j} |t_{i+1}^{(j)} - t_i^{(j)}| = \sum_j (b_j - a_j) < \delta$, so

$$\sum_j \sum_i d(\gamma(t_i^{(j)}), \gamma(t_{i+1}^{(j)})) < \varepsilon.$$

Taking the supremum over all partitions, it follows that

$$\sum_j |f(b_j) - f(a_j)| = \sum_j L(\gamma|_{[a_j,b_j]}) \leq \varepsilon,$$

yielding the claim. □

In the remainder of this section we follow [Sch18].

**1.6.9 Proposition.** *Let $\gamma : [a, b] \to X$ be a curve into a metric space $(X, d)$. Then the following are equivalent:*

*(i) $\gamma$ is absolutely continuous.*

*(ii) There exists some $l \in L^1([a, b])$ such that for all $s_1 \leq s_2 \in [a, b]$*

$$d(\gamma(s_1), \gamma(s_2)) \leq \int_{s_1}^{s_2} l(t)\, dt. \tag{1.6.9}$$

**Proof.** (i)⇒(ii): If $\gamma$ is absolutely continuous then by Theorem 1.6.8 we may use $l := v_\gamma \in L^1([a, b])$ to verify (1.6.9).
(ii)⇒(i): With $l$ as in (ii), the function $F : [a, b] \to \mathbb{R}$, $F(t) := \int_a^t l(s)\, ds$ is absolutely continuous (cf. [Coh13, Prop. 4.4.6]). Moreover, for any collection of pairwise disjoint subintervals $(a_i, b_i)$, $i = 1, \ldots, m$ we have

$$\sum_{i=1}^m d(\gamma(a_i), \gamma(b_i)) \leq \sum_{i=1}^m \int_{[a_i,b_i]} l(t)\, dt \leq \sum_{i=1}^m |F(b_i) - F(a_i)|,$$

so the claim follows. □

In fact, the metric derivative of an absolutely continuous curve $\gamma$ is the minimal $L^1$-function satisfying (ii) in Proposition 1.6.9:

**1.6.10 Theorem.** *Let $(X, d)$ be a metric space and let $\gamma : [a, b] \to X$ be an absolutely continuous path. Then the metric derivative $v_\gamma$ is the minimal $L^1([a, b])$-function such that*

$$d(\gamma(s), \gamma(t)) \leq \int_s^t v_\gamma(t)\, dt, \qquad \text{for all } s, t \in [a, b],\ s \leq t, \tag{1.6.10}$$

*i.e., $v_\gamma(t) \leq l(t)$ almost everywhere, for any $l \in L^1([a, b])$ such that (1.6.9) holds.*

**Proof.** $\gamma([a, b])$ is compact and metrizable, hence separable (cf. [Kun16, Prop. 11.2.10]). Let $\{x_n : n \in \mathbb{N}\}$ be a countable dense subset of $\gamma([a, b])$. The map $p \mapsto d(p, q)$ is 1-Lipschitz for any fixed $q \in M$, hence by Proposition 1.6.7 (iii) the functions

$$\phi_n : [a, b] \to \mathbb{R},\ t \mapsto d(\gamma(t), x_n)$$

are absolutely continuous. By Proposition 1.6.7 (iv) each $\phi_n$ is differentiable almost everywhere. Since countable unions of null-sets are null, at almost every point all $\phi_n$ are simultaneously differentiable, so we can define

$$\phi(t) := \sup_{n\in\mathbb{N}} |\phi_n'(t)| \text{ for almost every } t \in [a, b]. \tag{1.6.11}$$

We will now show that $\phi$ is integrable and that $\phi(t) = v_\gamma(t)$ almost everywhere. Note first that

$$|\phi_n(t + \varepsilon) - \phi_n(t)| = |d(\gamma(t + \varepsilon), x_n) - d(\gamma(t), x_n)| \leq d(\gamma(t + \varepsilon), \gamma(t)), \tag{1.6.12}$$

so

$$\liminf_{\varepsilon\to 0} \frac{d\big(\gamma(t + \varepsilon), \gamma(t)\big)}{|\varepsilon|} \geq \liminf_{\varepsilon\to 0} \frac{|\phi_n(t + \varepsilon) - \phi_n(t)|}{|\varepsilon|} = |\phi_n'(t)|$$

for almost every $t \in I$ and all $n \in \mathbb{N}$. This implies

$$\liminf_{\varepsilon \to 0} \frac{d\big(\gamma(t+\varepsilon), \gamma(t)\big)}{|\varepsilon|} \geq \phi(t) \text{ for a.e. } t \in [a,b]. \tag{1.6.13}$$

Next we show that, for any $t \in [a,b]$ as above, $d(\gamma(t+\varepsilon), \gamma(t)) = \sup_{n \in \mathbb{N}} |\phi_n(t+\varepsilon) - \phi_n(t)|$. Indeed, by (1.6.12), for all $n \in \mathbb{N}$ we have $d(\gamma(t+\varepsilon), \gamma(t)) \geq \sup_{n \in \mathbb{N}} |\phi_n(t+\varepsilon) - \phi_n(t)|$ . Since $(x_n)_n$ is dense in $[a,b]$, there is a subsequence $(x_{n_k})_k$ such that $\lim_{k \to \infty} x_{n_k} = \gamma(t)$. Therefore,

$$\begin{aligned} \sup_{n \in \mathbb{N}} |\phi_n(t+\varepsilon) - \phi_n(t)| &\geq \lim_{k \to \infty} |\phi_{n_k}(t+\varepsilon) - \phi_{n_k}(t)| \\ = \lim_{k \to \infty} |d(\gamma(t+\varepsilon), x_{n_k}) - d(\gamma(t), x_{n_k})| &= d(\gamma(t+\varepsilon), \gamma(t)). \end{aligned}$$

By Proposition 1.6.9 there exists $l \in L^1([a,b])$ such that (1.6.9) holds. Now for every Lebesgue point $t \in [a,b]$ of $l$ that satisfies (1.6.13) we get

$$0 \leq \phi(t) \leq \liminf_{\varepsilon \to 0} \frac{1}{|\varepsilon|} d(\gamma(t+\varepsilon), \gamma(t)) \leq \liminf_{\varepsilon \to 0} \frac{1}{|\varepsilon|} \int_{[t,t+\varepsilon]} l(r)\, dr = l(t), \tag{1.6.14}$$

where $[t, t+\varepsilon]$ is to be read as $[t+\varepsilon, t]$ in case $\varepsilon < 0$. Since almost every point in $[a,b]$ is a Lebesgue point of $l$ (cf. [Coh13, Prop. 6.3.10]), we conclude that $\phi \in L^1([a,b])$. Hence we obtain

$$\begin{aligned} d(\gamma(t+\varepsilon), \gamma(t)) &= \sup_{n \in \mathbb{N}} |\phi_n(t+\varepsilon) - \phi_n(t)| \\ = \sup_{n \in \mathbb{N}} \left| \int_{[t,t+\varepsilon]} \phi_n'(r)\, dr \right| &\leq \int_{[t,t+\varepsilon]} \sup_{n \in \mathbb{N}} |\phi_n'(r)|\, dr = \int_{[t,t+\varepsilon]} \phi(r)\, dr, \end{aligned} \tag{1.6.15}$$

Furthermore, (1.6.15) implies

$$\limsup_{\varepsilon \to 0} \frac{1}{|\varepsilon|} d(\gamma(t+\varepsilon), \gamma(t)) \leq \phi(t),$$

and together with (1.6.14) we get

$$\limsup_{\varepsilon \to 0} \frac{1}{|\varepsilon|} d(\gamma(t+\varepsilon), \gamma(t)) \leq \phi(t) \leq \liminf_{\varepsilon \to 0} \frac{1}{|\varepsilon|} d(\gamma(t+\varepsilon), \gamma(t)),$$

for almost every $t \in [a,b]$. Hence $v_\gamma = \phi$ in $L^1([a,b])$. Finally, (1.6.10) holds by (1.6.15), and the minimality of $v_\gamma$ follows from (1.6.14). □

# Chapter 2

# Constructions

In this chapter we introduce a number of basic techniques for producing new length spaces from given ones, as well as for generating interesting examples for the concepts to be studied later on.

## 2.1 Locality, gluing and maximal metrics

### 2.1.1 Locality

**2.1.1 Lemma.** *Let $(X_\alpha)_{\alpha\in A}$ be an open covering of a topological space $X$ such that each $X_\alpha$ carries a length structure $L_\alpha$. Suppose that the $L_\alpha$ are compatible, in the sense that a curve $\gamma$ whose image lies in $X_\alpha \cap X_\beta$ is admissible for $L_\alpha$ if and only if it is admissible for $L_\beta$ and $L_\alpha(\gamma) = L_\beta(\gamma)$. Then there exists a unique length structure $L$ on $X$ whose restriction to $X_\alpha$ equals $L_\alpha$ for each $\alpha \in A$. If $X$ is connected and all the intrinsic metrics induced by $L_\alpha$ on $X_\alpha$ are finite, then so is $L$.*

**Proof.** Given any curve $\gamma : [a,b] \to X$, compactness of $\gamma([a,b])$ implies that there are finitely many $X_\alpha$ which cover this set. Thus we may form a partition $a = t_0 < t_1 < \cdots < t_n = b$ such that each $\gamma([t_i,t_{i+1}])$ is contained in some $X_{\alpha_i}$. Consequently, the length of $\gamma|_{[t_i,t_{i+1}]}$ is unequivocally given by $L_{\alpha_i}(\gamma)$ and we define $L(\gamma)$ as the sum over the $L_{\alpha_i}$-lengths of $\gamma|_{[t_i,t_{i+1}]}$. If $\gamma([a,b])$ is also covered by some other finite collection of certain $X_{\beta_j}$, then using a partition such that each piece of $\gamma$ is contained in an intersection of some $X_{\alpha_i}$ and some $X_{\beta_j}$ shows that $L(\gamma)$ is well-defined. Moreover, the requirement of additivity on $L$ shows uniqueness. All the properties of a length structure for $L$ now follow readily from those of the $L_\alpha$.
Finally, suppose that $X$ is connected and all the intrinsic metrics induced by $L_\alpha$ on $X_\alpha$ are finite. Fixing any $x \in X$, let

$$Y := \{y \in X : d_L(x,y) < \infty\}.$$

Then any $X_\alpha$ is either contained in $Y$ or in $X \setminus Y$, showing that both sets are open, and so connectedness of $X$ implies $X = Y$. □

**2.1.2 Corollary.** *Let $d_1$, $d_2$ be two intrinsic metrics on a set $X$ that induce the same topology on $X$. Suppose that for each $x \in X$ there exists a neighborhood $U_x$ of $x$ such that $d_1|_{U_x\times U_x} = d_2|_{U_x\times U_x}$. Then $d_1 = d_2$.*

**Proof.** Let $x$, $y \in X$ and fix any $\varepsilon > 0$. Then either $d_1(x,y) = \infty \ge d_2(x,y)$ or there exists a curve $\gamma : [a,b] \to X$ connecting $x$ to $y$ such that $d_1(x,y) \ge L_{d_1}(x,y) - \varepsilon$. Covering $\gamma([a,b])$ by neighborhoods as in the assumption we may pick a partition $a = t_0 < t_1 < \cdots < t_n = b$ with $\gamma([t_i,t_{i+1}]) \subseteq U_{x_i}$ for suitable $x_i \in X$. Then by definition of $L_{d_1}$,

$$d_1(x,y) \ge L_{d_1}(x,y) - \varepsilon \ge \sum_{i=0}^{n-1} d_1(\gamma(t_i),\gamma(t_{i+1})) - \varepsilon = \sum_{i=0}^{n-1} d_2(\gamma(t_i),\gamma(t_{i+1})) - \varepsilon \ge d_2(x,y) - \varepsilon.$$

As $\varepsilon$ was arbitrary, $d_1(x, y) \geq d_2(x, y)$. If $d_1(x, y) = \infty$, then so is $d_2(x, y)$, otherwise the above calculation with the roles of $d_1$ and $d_2$ reversed would yield $d_2(x, y) \geq d_1(x, y) = \infty$.
In case, both are finite $d_2(x, y) \geq d_1(x, y)$ follows as above by symmetry.

□

That the conclusion of Corollary 2.1.2 fails if the metrics are not required to be intrinsic follows from the following result, which shows that for a non-intrinsic, complete metric there exists another metric, coinciding locally with it, but not globally.

**2.1.3 Proposition.** *Let $d$ be a complete metric on a set $X$ and suppose that $d$ is not intrinsic. Then there exists another metric $d_1 \neq d$ such that every point in $X$ has a neighborhood on which $d$ and $d_1$ coincide.*

**Proof.** For any $\varepsilon > 0$, define a metric $d_\varepsilon$ on $X$ by

$$d_\varepsilon(x, y) := \inf \sum_{i=0}^{k} d(p_i, p_{i+1}),$$

with the infimum being taken over all sequences of points $x = p_0, p_1, \ldots, p_{k+1} = y$ such that $d(p_i, p_{i+1}) \leq \varepsilon$ for all $i$. Then clearly $d_\varepsilon(x, y) = d(x, y)$ whenever $d(x, y) \leq \varepsilon$, so $d$ and $d_\varepsilon$ coincide on any ball of radius $\leq \varepsilon/2$. On the other hand, there must be some $\varepsilon > 0$ such that $d \neq d_\varepsilon$ because otherwise we could appeal to Corollary 1.3.13 to conclude that $d$ is intrinsic. □

### 2.1.2 Gluing

In this section we want to clarify how to glue two or more length spaces in such a way that the intrinsic distances in the original spaces remain the same. To begin with, we look at some examples.

**2.1.4 Example.** Consider the strip $\mathbb{R} \times [0, 1]$ in $\mathbb{R}^2$, and for every $x \in \mathbb{R}$ identify $(x, 1)$ with $(x + 100, 0)$ (which topologically is a cylinder). What then should the distance between the points $(0, 1/2)$ and $(1000, 1/2)$ be? To find this out, let us measure the length of the following connecting path:

$$\begin{aligned} |(0,1/2) \to (0,1)| = 1/2,\ &|(0,1) \to (100,0)| = 0 \\ |(100,0) \to (100,1)| = 1,\ &|(100,1) \to (200,0)| = 0 \\ \vdots \qquad\qquad & \qquad\qquad \vdots \\ |(900,0) \to (900,1)| = 1,\ &|(900,1) \to (1000,0)| = 0 \\ &|(1000,0) \to (1000,1/2)| = 1/2 \end{aligned}$$

In fact, this path is of minimal length, so the distance $(0, 1/2)$ and $(1000, 1/2)$ should be defined to be 10, see figure 2.1.

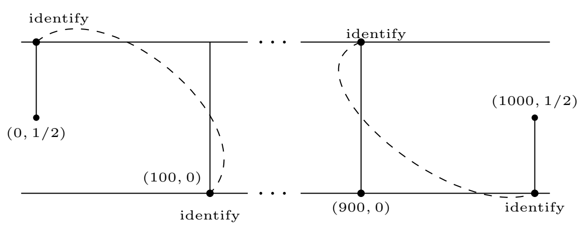



Figure 2.1: The figure depicts the shortest path (vertical lines) between the points $(0, 1/2)$ and $(1000, 1/2)$

**2.1.5 Example.** This time, start out with $\mathbb{R}^2$ and identify any $(x, y)$ with $(-y, 2x)$. Then the distance of any point to the origin should be set to zero, because there is an arbitrarily short path from $(x, y)$ to $(0, 0)$, namely

$$(x, y) \to (y/2, -x) \to (-x/2, -y/2) \to (-y/4, x/2) \to (x/4, -y/4) \to \dots$$

Consequently, the distance between any two points should be set to zero.

These examples suggest a strategy for defining a metric on a space that results from identifying certain points: consider finite sequences of paths by gluing the endpoint of one path to the starting point of the next (in case they do not already agree anyways). Define the distance of two points as the infimum of the total length of such finite sequences of paths connecting them. More precisely:

**2.1.6 Definition.** *Let $(X, d)$ be a metric space and let $R$ be an equivalence relation on $X$. Then the quotient semi-metric $d_R$ on $X$ is defined as*

$$d_R(x, y) := \inf \left\{ \sum_{i=1}^{k} d(p_i, q_i) : p_1 = x, q_k = y, \ k \in \mathbb{N} \right\}, \tag{2.1.1}$$

*where the infimum is taken over all choices of $\{p_i\}$ and $\{q_i\}$ such that $q_i \sim_R p_{i+1}$ for all $i = 1, \dots, k-1$.*
*Identifying points of vanishing $d_R$-distance we obtain a metric space, the so-called* quotient metric space $(X/d_R, d_R)$*, resulting from gluing $(X, d)$ along the relation $R$.*

**2.1.7 Remark.** Let us briefly verify that $d_R$ is indeed a semi-metric: non-negativity and symmetry are obvious from the definition, so it only remains to establish the triangle inequality. Let $x, y, z \in X$ (all different) and let $\{p_i\}$, $\{q_i\}$ $(1 \le i \le k)$ be as in Definition 2.1.6 for $d_R(x, z)$. Adding $y$ if necessary we may assume that for some $1 < l < k$ we have $p_l = q_l = y$. For any such choice we have

$$d_R(x, z) \le \sum_{i=1}^{l} d(p_i, q_i) + \sum_{i=l+1}^{k} d(p_i, q_i).$$

Given any $\varepsilon > 0$, by a suitable choices of $\{p_i\}$, $\{q_i\}$ the right hand side can be made smaller than $d_R(x, y) + d_R(y, z) + \varepsilon$, giving the claim.

Henceforth we will only consider quotients of length spaces. For these we have:

**2.1.8 Lemma.** *If $(X, d)$ is a length space, then so is $(X/d_R, d_R)$.*

**Proof.** Setting $k = 1$ in (2.1.1) it follows that $d_R \le d$. Consequently, any $d$-continuous curve is also $d_R$-continuous and its $d_R$-length is less or equal to its $d$-length. Now let $\varepsilon > 0$ and pick $\{p_i\}$ and $\{q_i\}$ as in 2.1.6 such that $\sum_{i=1}^{k} d(p_i, q_i) < d_R(x, y) + \varepsilon$. Since $(X, d)$ is a length space, for each $1 \le i \le k$ there exists a path in $X$ from $p_i$ to $q_i$ whose length is smaller than $d(p_i, q_i) + \frac{\varepsilon}{k}$. Concatenating these paths we obtain a curve $\gamma$ in $X/d_R$ that is continuous with respect to $d_R$ because $p_{i+1}$ is identified with $q_i$ for each $i$. Moreover,

$$L_{d_R}(\gamma) \le L_d(\gamma) \le \sum_{i=1}^{k} d(p_i, q_i) + \varepsilon < d_R(x, y) + 2\varepsilon,$$

which shows that indeed $(X/d_R, d_R)$ is a length space. $\square$

**2.1.9 Remark.** It may happen that the equivalence relation induced on $X$ by first identifying via $R$ and then factoring out the points of vanishing $d_R$-distance is strictly stronger than $R$. Therefore, $(X/d_R, d_R)$ need not be homeomorphic to the topological quotient $X/R$. As an example, consider on $[0, 1]$ the equivalence relation that identifies all rational points. This space is not even $T_1$ (cf. [Kun16, 3.2.4]), whereas $(X/d_R, d_R)$ is a metric space.

**2.1.10 Definition.** *Let* $(X_\alpha, d_\alpha)_{\alpha\in A}$ *be a family of length spaces. On the disjoint union* $X := \coprod_{\alpha\in A} X_\alpha$ *we introduce a length metric* $d$ *by setting*

$$d(x,y) := \begin{cases} d_\alpha(x,y) & \text{if } x,y \in X_\alpha \text{ for some } \alpha \\ \infty & \text{otherwise} \end{cases}$$

*This metric is called the* length metric of the disjoint union*.*

As a general procedure, if $(X_\alpha, d_\alpha)_{\alpha\in A}$ is a family of length spaces and $R$ is an equivalence relation on their disjoint union, then in order to glue the $X_\alpha$ along $R$ one first endows the disjoint union with the length metric from definition 2.1.10 and then forms the metric quotient according to Definition 2.1.6.

**2.1.11 Example.** On $I = [0, 2\pi]$ with the standard metric, consider the equivalence relation identifying 0 with $2\pi$ (and leaving the other points untouched). Then $(I/d_R, d_R)$ can be identified with the unit circle.

**2.1.12 Example.** Consider a countable family of disjoint compact intervals $I_i$ ($i \in \mathbb{N}$) and identify all their left ends. We want to find out whether the resulting metric quotient $X$ is compact. It turns out that this depends on the lengths of the $I_i$: If each $I_i$ has length 1 then any two right end points have distance 2. This gives a sequence of points that does not possess a convergent subsequence, so the answer is negative. However, if the length of $I_i$ is $1/i$ then any sequence in $X$ either has a subsequence that eventually remains in one fixed $I_j$, or that moves to ever higher $I_i$. In both cases we can extract a further subsequence that converges. Hence in this case $X$ is compact.

**2.1.13 Example.** Let $G$ be some group of isometries (rigid motions) of $\mathbb{R}^2$. We define an equivalence relation $R$ on $\mathbb{R}^2$ by letting $xRy$ if there exists some $g \in G$ with $x = gy$. The equivalence classes then are precisely the group orbits. Our aim is to show that the corresponding quotient metric is given by

$$d_R(x,y) = \inf_{g\in G} d(gx, y),$$

with $d(x,y) = |x-y|$ the Euclidean norm. In fact, let $\{p_i\}$ and $\{q_i\}$ be as in Definition 2.1.6. Then there exist $g_i \in G$ such that $p_{i+1} = g_i q_i$ for $1 \le i \le k-1$. Now

$$d(x,q_1) + d(g_1q_1, q_2) = d(x,q_1) + d(q_1, g_1^{-1}q_2) \ge d(x, g_1^{-1}q_2).$$

Iterating this process, we obtain

$$\begin{aligned} d_R(x,y) &= \inf(d(x,q_1) + d(g_1q_1,q_2) + \cdots + d(g_{k-1}q_{k-1}, y)) \\ &\ge \inf(d(x, g_1^{-1}q_2) + d(g_2q_2, q_3) + \cdots + d(g_{k-1}q_{k-1}, y)) \\ &= \inf(d(g_1x, q_2) + d(q_2, g_2^{-1}q_3) + \cdots + d(g_{k-1}q_{k-1}, y)) \\ &\ge \inf(d(g_2g_1x, q_3) + d(g_3q_3, q_4) + \cdots + d(g_{k-1}q_{k-1}, y)) \\ &\cdots \ge \inf d(g_{k-1}\ldots g_1x, y) = \inf d(gx, y). \end{aligned}$$

The converse inequality is obvious, so the claim follows.
Specifically, if we take for $G$ all translations by vectors with integer coordinates then the resulting metric quotient is precisely the two-dimensional flat torus. If we let $G$ be the group of all rotations around the origin, the resulting space is isometric to the half-ray $[0, \infty)$. Many other examples can be created along these lines.

Next we are going to develop an alternative perspective on gluing length spaces. When identifying points via gluing, we set their distances equal to zero, while we want other distances to remain the same. Simply doing this, however, might lead to a violation of the triangle inequality. Hence, in general, also the distances of points that are not identified will have to be changed. Nevertheless,

we expect that gluing will at least not increase any distances. Thus we are led to searching for a metric that, while smaller than the original one, should give zero on points that are identified. In fact, we will see that choosing the maximal metric satisfying these constraints will precisely give the quotient metric introduced above.
By a semi-metric we mean a function $d : X \times X \to \mathbb{R}_+ \cup \{\infty\}$, such that $d(x, x) = 0$, $d$ is symmetric and the triangle inequality holds.
We first show that a reasonable notion of 'maximal metric' does indeed exist:

**2.1.14 Lemma.** *Let $X$ be a set, let $b : X \times X \to \mathbb{R}_+ \cup \{\infty\}$ be any map and denote by $D$ the set of all semi-metrics $d$ on $X$ with $d \leq b$ (i.e., $d(x, y) \leq b(x, y)$ for all $x, y \in X$). Then there is a unique $d_m \in D$ such that $d_m \geq d$ for all $d \in D$.*

**Proof.** Given $x, y \in X$, simply set $d_m(x, y) := \sup\{d(x, y) : d \in D\}$. Then it only remains to show that $d_m$ satisfies the triangle inequality. Thus let $x, y, z \in X$. Then

$$d_m(x, y) = \sup_{d\in D} d(x, y) \leq \sup_{d\in D}(d(x, z) + d(z, y)) \leq \sup_{d\in D} d(x, z) + \sup_{d\in D} d(z, y) = d_m(x, z) + d_m(z, y).$$

Further since still $\sup_{d\in D} d(x, y) \leq b(x, y)$, we have $d_m \in D$ and clearly $d_m \geq d$ for all $d \in D$. □

**2.1.15 Corollary.** *Let $X$ be a set and suppose that $X = \bigcup_{\alpha\in A} X_\alpha$, where each $X_\alpha$ carries a semi-metric $d_\alpha$. Let $D$ be the set of all semi-metrics (possibly taking infinite values) $d$ on $X$ such that $d|_{X_\alpha\times X_\alpha} \leq d_\alpha$ for all $\alpha \in A$. Then there is a unique maximal semi-metric $d_m$ in $D$ (i.e., $d_m \geq d$ for each $d \in D$). If all $d_\alpha$ are intrinsic then so is $d_m$.*

**Proof.** By setting $d_\alpha(x, y) := \infty$ whenever $x \notin X_\alpha$ or $y \notin X_\alpha$, we may without loss of generality assume that each $d_\alpha$ is in fact defined on all of $X \times X$. Then existence and uniqueness of $d_m$ follows from 2.1.14 upon setting $b(x, y) := \inf_{\alpha\in A} d_\alpha(x, y)$.
Assume now that each $d_\alpha$ is intrinsic and let $\widehat{d_m}$ be the intrinsic metric induced by $d_m$ (proceed as in the case $d_m$ were a metric). Then for each $\alpha$, $d_m \leq d_\alpha$ implies $\widehat{d_m} \leq \widehat{d_\alpha} = d_\alpha$, implying $\widehat{d_m} \in D$. By construction, then, $\widehat{d_m} \leq d_m$. But the converse inequality always holds (cf. Lemma 1.2.3), so $\widehat{d_m} = d_m$, which means that $d_m$ is intrinsic. □

**2.1.16 Theorem.** *Let $(X, d)$ be a metric space and let $R$ be an equivalence relation on $X$. Let $b_R : X \times X \to \mathbb{R}$,*

$$b_R(x, y) := \begin{cases} 0 & \text{if } x \text{ is } R\text{-equivalent to } y \\ d(x, y) & \text{otherwise.} \end{cases}$$

*Then the maximal semi-metric not exceeding $b_R$ is precisely the quotient semi-metric $d_R$ from Definition 2.1.1.*

**Proof.** Let $D := \{d' \text{ semi-metric on } X : d' \leq b_R\}$. By definition, $d_R \in D$, so it remains to show that $d_R \geq d'$ for every $d' \in D$. To see this, let $\{p_i\}$ and $\{q_i\}$ be as in Definition 2.1.1, and set $p_{k+1} := y$. Then for any $d' \in D$ we have

$$\begin{aligned} d'(x, y) &\leq \sum_{i=1}^{k} d'(p_i, p_{i+1}) \leq \sum_{i=1}^{k} d'(p_i, q_i) + \sum_{i=1}^{k} d'(q_i, p_{i+1}) \leq \sum_{i=1}^{k} b_R(p_i, q_i) + \sum_{i=1}^{k} b_R(q_i, p_{i+1}) \\ &\leq \sum_{i=1}^{k} b_R(p_i, q_i) + 0 \leq \sum_{i=1}^{k} d(p_i, q_i). \end{aligned}$$

Consequently, $d' \leq d_R$. □

**2.1.17 Remark.** If $d$ is intrinsic we can use Corollary 2.1.15 to provide an alternative proof that $d_R$ is intrinsic in this case (cf. Lemma 2.1.8). Indeed we have for any semi-metric $\tilde{d}$ exceeded by $b_R$, that $\tilde{d} \leq b_R \leq d$ and so Corollary 2.1.15 shows that the $d_R$ is intrinsic.

## 2.2 Polyhedral Spaces

A large class of examples for length spaces obtained via gluing is constituted by polyhedral spaces. Basically they are constructed by gluing convex polyhedra along certain faces. Inductively, a 0-dimensional polyhedral space $P_0$ is a metric space consisting of finitely many points (the *vertices* of $P_0$) whose pairwise distances are all infinite. Given such a $P_0$ and a finite collection of length spaces $E = \{E_i\}$, where each $E_i$ is isometric to some compact interval, we may construct a 1-dimensional polyhedral space by fixing injective maps $e_i$ that send the endpoints of $E_i$ to $P_0$. Then $P_1$ results from metrically gluing $P_0$ along $\{e_i\}$. We can regard the segments of $E$ as lying in $P_1$. One-dimensional polyhedral spaces are also called *metric graphs*.
Next, given a 1-dimensional polyhedral space $P_1$ and a finite collection of polygons $F = \{F_i\}$, we may construct a 2-dimensional polyhedral space by fixing, for each $F_i$, an injective map from the boundary of $F_i$ to $P_1$ such that $f_i$ maps each side of $F_i$ isometrically onto an edge from $E$. Now gluing $F$ to $P_1$ along $\{f_i\}$ produces a 2-dimensional polyhedral space $P_2$. The copies of $F_i$ lying in $P_2$ are called the *faces* of $P_2$.

**2.2.1 Example.** The surface of a convex polyhedron in $\mathbb{R}^3$ is a 2-dimensional polyhedral space.

More generally, one has:

**2.2.2 Definition.** *Let $(P, d)$ be a metric space covered by a family of length spaces $(P_\alpha, d_\alpha)$, each isometric to a convex polyhedron. Suppose that for any $\alpha \neq \beta$, $P_\alpha \cap P_\beta$ is a face of both $P_\alpha$ and $P_\beta$ and that the metrics induced by $d_\alpha$ and $d_\beta$ coincide on it. Let $d$ be the maximal metric majorized by all $d_\alpha$ according to Corollary 2.1.15. Then $(P, d)$ is called a (Euclidean)* polyhedral space. *The polyhedra $P_\alpha$ are called the faces of $P$. 0- resp. 1-dimensional faces are called* vertices *and* edges*, respectively.*

Let us now return to the important special case of metric graphs: Given a disjoint family of segments $\{E_i\}$ and points $\{v_j\}$, consider first the disjoint union of these length spaces according to Definition 2.1.10. By definition, we want to metrically glue the edges $\{E_i\}$ along an equivalence relation $R$ defined on the union of $\{v_j\}$ and the endpoints of the $E_j$.
To obtain the corresponding metric graph, consider the family of all metrics on the individual segments, supplemented by the semi-metric $d_R(x, y)$ that is zero if $xRy$ and infinite otherwise. Then form the maximal semi-metric bounded by this family of (semi-)metrics and factor out the points of $d_R$-distance zero. The $E_i$, viewed as subsets of the metric graph, are called edges, and the $v_j$ the vertices. The cardinality of the equivalence class representing a vertex is called its *degree*.
An equivalent and very natural way of constructing a metric graph consists in starting out from a set $V$ of vertices, then specifying which pairs of vertices should be connected by edges and, finally, specifying lengths for these edges. Re-translating this construction into the above picture, one takes a collection of segments $\{E_i\}$ with the desired lengths and if $E_i$ is supposed to connect vertices $v$ and $w$, identify one endpoint of $E_i$ with $v$ and the other with $w$. In this way one determines the equivalence relation $R$, and the process described previously gives precisely the desired graph. In particular this implies that any topological graph can be turned into a metric graph by assigning lengths to its edges. Note, however, that this procedure may change the topology of the graph. Also note that, as a general effect of gluing, it may happen that points in a metric graph can get identified despite the fact that they are not $R$-equivalent, cf. Example 2.1.5.

**2.2.3 Example.** *Any* metric space can be realized as the set of vertices of a (sufficiently monstrous) metric graph: Let $(X, d)$ be a metric space, and form the disjoint union of the segments $I_{x,y} := [0, d(x, y)]$ for any pair $(x, y) \in X^2$. Now identify the left ends of two segments $I_{x,y}$ and $I_{x',y'}$ if $x = x'$, and their right ends if $y = y'$. Then $X$ can be naturally identified with the set of vertices of the resulting metric graph, and the intrinsic metric of the graph is precisely the original metric $d$ on $X$.

## 2.3 Products and cones

Given two length spaces $(X, d_X)$, $(Y, d_Y)$, we may define a metric on the cartesian product $Z = X \times Y$ of $X$ and $Y$ by

$$d((x_1, y_1), (x_2, y_2)) := \sqrt{d_X(x_1, x_2)^2 + d_Y(y_1, y_2)^2}.$$

This metric is called the *product metric* and $(Z, d)$ is called the metric product of $X$ and $Y$.

**2.3.1 Theorem.** *Let $(X, d_X)$ and $(Y, d_Y)$ be metric spaces and let $(Z, d)$ be their metric product.*

*(i) The projections $\mathrm{pr}_X : Z \to X$ and $\mathrm{pr}_Y : Z \to Y$ are distance- and length-decreasing. They are isometries on each fiber $X \times \{y\}$ resp. $\{x\} \times Y$.*

*(ii) For any path $\gamma : [a, b] \to Z$, $\gamma(t) = (\gamma_X(t), \gamma_Y(t))$ we have $L(\gamma_X)^2 + L(\gamma_Y)^2 \le L(\gamma)^2$.*

*(iii) If $\gamma : [a, b] \to Z$, $\gamma(t) = (\gamma_X(t), \gamma_Y(t))$ is a distance realizing path, then so are $\gamma_X$ and $\gamma_Y$. The converse is true if both $\gamma_X$ and $\gamma_Y$ are constant speed distance realizing paths.*

*(iv) $d$ is intrinsic if and only if both $d_X$ and $d_Y$ are. In other words, $(Z, d)$ is a length space if and only if both $(X, d_X)$ and $(Y, d_Y)$ are.*

*(v) $d$ is strictly intrinsic if and only if both $d_X$ and $d_Y$ are.*

**Proof.** (i) It suffices to consider the case $\mathrm{pr}_X$. By definition of $d$, $\mathrm{pr}_X$ is distance-decreasing and an isometry when restricted to any fiber $X \times \{y\}$. It is then immediate from Definition 1.2.2 that $\mathrm{pr}_X$ is also length-decreasing.
(ii) Generally, as is easily verified, if $a_i, b_i \ge 0$, then

$$\Big(\sum_i a_i\Big)^2 + \Big(\sum_i b_i\Big)^2 \le \Big(\sum_i \sqrt{a_i^2 + b_i^2}\Big)^2.$$

Consequently, if $a = t_0 < \cdots < t_N = b$ is any partition, then

$$\Big(\sum_i d_X(\gamma_X(t_i), \gamma_X(t_{i+1}))\Big)^2 + \Big(\sum_i d_Y(\gamma_Y(t_i), \gamma_Y(t_{i+1}))\Big)^2 \le \Big(\sum_i d(\gamma(t_i), \gamma(t_{i+1}))\Big)^2.$$

Letting the modulus of the partition go to zero, the claim follows from Lemma 1.2.5.
(iii) Let $\gamma$ be distance realizing from $z_1 = (x_1, y_1)$ to $z_2 = (x_2, y_2)$ and suppose, to the contrary, that $\gamma_X$ is not distance realizing from $x_1$ to $x_2$. Then (ii) implies

$$d(z_1, z_2) = \sqrt{d_X(x_1, x_2)^2 + d_Y(y_1, y_2)^2} < \sqrt{L(\gamma_X)^2 + L(\gamma_Y)^2} \le L(\gamma),$$

a contradiction. Conversely, suppose that both $\gamma_X$ and $\gamma_Y$ are distance realizing paths and are parametrized proportional to arclength, say $a = 0$ and $L_0^t(\gamma_X) = c_X \cdot t$, $L_0^t(\gamma_Y) = c_Y \cdot t$. Then $d_X(\gamma_X(t), \gamma_X(t')) = c_X|t - t'|$, $d_Y(\gamma_Y(t), \gamma_Y(t')) = c_Y|t - t'|$, and so $d(\gamma(t), \gamma(t')) = \sqrt{c_X^2 + c_Y^2}|t - t'|$. This implies

$$L(\gamma) = \sup \left\{ \sum_i d(\gamma(t_i), \gamma(t_{i+1})) \right\} = \sqrt{c_X^2 + c_Y^2} \cdot b = d(\gamma(0), \gamma(b)).$$

(iv) Suppose first that $(Z, d)$ is a length space, and let $x_1, x_2 \in X$ and $\varepsilon > 0$. For any fixed $y \in Y$, there exists a path $\gamma = (\gamma_X, \gamma_Y)$ in $Z$ from $z_1 = (x_1, y)$ to $z_2 = (x_2, y)$ such that

$$L(\gamma_X) \le L(\gamma) < d((x_1, y), (x_2, y)) + \varepsilon = d_X(x_1, x_2) + \varepsilon,$$

so $(X, d_X)$ is a length space as well. Now suppose that both $d_X$ and $d_Y$ are intrinsic and let $z_1 = (x_1, y_1)$, $z_2 = (x_2, y_2)$ and $\varepsilon > 0$. Then there exist paths $\gamma_X : [0, 1] \to X$ from $x_1$ to

$x_2$ and $\gamma_Y : [0,1] \to Y$ from $y_1$ to $y_2$, both parametrized proportional to arclength such that $L(\gamma_X)^2 < d_X(x_1,x_2)^2 + \varepsilon^2/2$ and $L(\gamma_Y)^2 < d_Y(y_1,y_2)^2 + \varepsilon^2/2$. Then $\gamma = (\gamma_X, \gamma_Y)$ joins $z_1$ to $z_2$ and, by Lemma 1.2.5 has length $\sup_n \sum_{i=0}^{n-1} d(\gamma(i/n), \gamma((i+1)/n))$. Since $\gamma_X$ and $\gamma_Y$ are parametrized proportional to arclength, their total length is $n$ times the length of any of their restrictions to an interval $[i/n, (i+1)/n]$. Thus for any $i \in \{0, \dots, n-1\}$ we have:

$$\begin{aligned} n^2 d(\gamma(i/n), \gamma((i+1)/n))^2 &= n^2 d_X(\gamma_X(i/n), \gamma_X((i+1)/n))^2 + n^2 d_Y(\gamma_Y(i/n), \gamma_Y((i+1)/n))^2 \\ &\le L(\gamma_X)^2 + L(\gamma_Y)^2 < d_X(x_1,x_2)^2 + \varepsilon^2/2 + d_Y(y_1,y_2)^2 + \varepsilon^2/2 \\ &= d(z_1,z_2)^2 + \varepsilon^2. \end{aligned}$$

Consequently, $d(\gamma(i/n), \gamma((i+1)/n)) < \frac{1}{n}(d(z_1,z_2) + \varepsilon)$ for any such $i$, and so

$$\sum_{i=0}^{n-1} d(\gamma(i/n), \gamma((i+1)/n)) < d(z_1,z_2) + \varepsilon$$

Letting $n \to \infty$ gives $L(\gamma) \le d(z_1,z_2) + \varepsilon$. This shows that $(X,d)$ is a length space.
Finally, (v) is immediate from (iii). □

**2.3.2 Remark.** (i) Note that the assumption of constant speed in the converse direction of (iii) is essential: Indeed, in $\mathbb{R}^2 = \mathbb{R} \times \mathbb{R}$, *any* path $(\gamma_1, \gamma_2)$ with increasing coordinates $\gamma_1$ and $\gamma_2$ is the product of two distance realizing paths.
(ii) The isometry group of $(Z,d)$ is at least as rich as the product of the isometry groups of $X$ and $Y$. To see this, note that if $I_X$ is an isometry of $X$ and $I_Y$ is one of $Y$, then $I_X \times I_Y : (x,y) \mapsto (I_X(x), I_Y(y))$ is an isometry of $Z$, so the isometry group of $Z$ contains isomorphic copies of the isometry groups of $X$ and $Y$. Already the case of $\mathbb{R} \times \mathbb{R}$ shows that in general there are many more isometries of the product space than in the product of the individual groups.

**2.3.3 Definition.** *A subset $A$ in a metric space $(X,d)$ is called* convex *if the restriction of $d$ to $A$ is strictly intrinsic.*

The reason for calling such subsets convex becomes apparent via the following result:

**2.3.4 Lemma.** *Let $(X,d)$ be a strictly intrinsic metric space and let $A \subseteq X$. Then the following are equivalent:*

*(i) $A$ is convex.*

*(ii) For any points $x,y \in A$ there is a shortest (in $X$) path from $x$ to $y$ which is contained in $A$.*

**Proof.** (i)⇒(ii): Since $d|_{A\times A}$ is strictly intrinsic, for any $x,y \in A$ there exists a path in $A$ from $x$ to $y$ of length $d(x,y)$.
(ii)⇒(i): Let $x,y \in A$ and take a shortest path as provided by (ii). This path of course automatically realizes the distance $d|_{A\times A}$, so $(A, d|_{A\times A})$ is strictly intrinsic. □

**2.3.5 Definition.** *Let $(X,d)$ be a strictly intrinsic metric space and let $A \subseteq X$. Then $A$ is called* locally convex *if every point in $A$ has a neighborhood $U$ in $A$ such that for any two points $y,z \in U$ there exists a shortest path from $y$ to $z$ that is contained in $A$.*

**2.3.6 Remark.** Any totally geodesic submanifold $A$ of a Riemannian manifold $M$ is locally convex in this sense. Indeed, let $U$ be a convex neighborhood in $A$ that is contained in a convex neighborhood $V$ in $M$. Then any two points of $U$ are connected by a minimizing geodesic in $A$ that is also a minimizing geodesic in $M$.

**2.3.7 Lemma.** *Let $(X,d)$ be a strictly intrinsic metric space and let $F : X \to Y$ ($Y$ a metric space) be a distance-preserving map. Then $F(X)$ is convex in $Y$.*

**Proof.** It suffices to observe that $X$ is convex in itself and that $F : X \to (F(X), d_Y|_{F(X)\times F(X)})$ is an isometry. □

**2.3.8 Proposition.** *Let $X$ and $Y$ be length spaces and let $\alpha : [a,b] \to X$, $\beta : [c,d] \to Y$ be shortest paths. Then $R := \alpha([a,b]) \times \beta([c,d])$ is convex in $X \times Y$ and isometric to a Euclidean rectangle.*

**Proof.** Let both $\alpha$ and $\beta$ be parametrized by arclength. Then $F : [a,b] \times [c,d] \to Z$, $F(t,s) := (\alpha(t), \beta(s))$ is an isometry, because

$$d^2(F(t,s), F(t',s')) = d_X^2(\alpha(t), \alpha(t')) + d_Y^2(\beta(t), \beta(t')) = (t-t')^2 + (s-s')^2.$$

The claim now follows from Lemma 2.3.7. □

Next we turn to an important class of examples, namely cones over metric spaces. To begin with, the cone $\mathrm{Con}(X)$ over a topological space $X$ is defined to be the quotient of the product $X \times [0,\infty)$ that results from gluing all the points in the fiber $X \times \{0\}$. The point resulting from this identification is called the origin or apex of the cone.
To get an idea of how to endow a cone over a metric space $(X,d)$ with a suitable metric, consider first the model case of $X$ being a subset of the unit sphere $S^2 = \{(x,y,z) : x^2+y^2+z^2 = 1\} \subseteq \mathbb{R}^3$. The cone over $X$ then has its apex in the origin $O$ and consists of the half-rays emanating from $O$ and containing a point of $X$. This suggests to use spherical coordinates, assigning to any point $a$ in the cone the pair $(x,t)$, with $x \in X$ and $t = |aO|$, so $a = tx$. To calculate the Euclidean distance between two points $a = (x,t)$ and $b = (y,s)$ in the cone, consider the triangle $\Delta Oab$. Then $|Oa| = t$, $|Ob| = s$ and the angle $\measuredangle aOb$ is the angular distance $d(x,y)$ in $X$. The law of cosines therefore gives:

$$|ab| = \sqrt{t^2 + s^2 - 2ts\cos(d(x,y))}.$$

We take this as a clue for our definition also in the general case:

**2.3.9 Definition.** *Let $X$ be a metric space with $\mathrm{diam}(X) \le \pi$. The cone metric $d_c$ on $\mathrm{Con}(X)$ is defined by*

$$d_c(p,q) := \sqrt{t^2 + s^2 - 2ts\cos(d(x,y))}, \tag{2.3.1}$$

*for $p, q \in \mathrm{Con}(X)$, $p = (x,t), q = (y,s)$.*

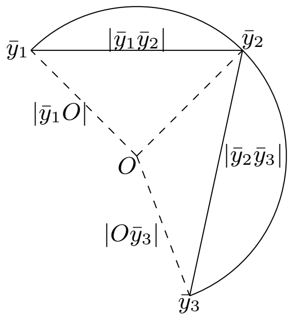
**2.3.10 Proposition.** *Let $(X,d)$ be a metric space with $\mathrm{diam}(X) \le \pi$. Then $d_c$ is a metric on $\mathrm{Con}(X)$.*

**Proof.** Clearly $d_c$ is non-negative and symmetric, so it remains to verify the triangle inequality. Let $y_1 = (x_1, r_1)$, $y_2 = (x_2, r_2)$, $y_3 = (x_3, r_3) \in \mathrm{Con}(X)$, and set $\alpha := d(x_1, x_2)$, $\beta := d(x_2, x_3)$. Now let $\bar{y}_1, \bar{y}_2, \bar{y}_3$ be points in $\mathbb{R}^2$ whose distances to the origin are $r_1, r_2, r_3$ and such that $\measuredangle \bar{y}_1 O \bar{y}_2 = \alpha$, $\measuredangle \bar{y}_2 O \bar{y}_3 = \beta$, and the rays $O\bar{y}_1$ and $O\bar{y}_2$ lie in different half-planes with respect to $O\bar{y}_2$. Then (again by the law of cosines), for the Euclidean distances we have $|\bar{y}_1\bar{y}_2| = d_c(y_1, y_2)$ and $|\bar{y}_2\bar{y}_3| = d_c(y_2, y_3)$. We distinguish two cases:
$\alpha + \beta \le \pi$: Then $\measuredangle \bar{y}_1 O \bar{y}_3 = \alpha + \beta = d(x_1,x_2) + d(x_2,x_3) \ge d(x_1,x_3)$, and so the law of cosines implies

$$\begin{aligned} |\bar{y}_1\bar{y}_3| &= \sqrt{r_1^2 + r_3^2 - 2r_1 r_3 \cos(d(x_1,x_2) + d(x_2,x_3))} \\ &\ge \sqrt{r_1^2 + r_3^2 - 2r_1 r_3 \cos(d(x_1,x_3))} = d_c(y_1,y_3). \end{aligned} \tag{2.3.2}$$

Hence

$$d_c(y_1,y_2) + d_c(y_2,y_3) = |\bar{y}_1\bar{y}_2| + |\bar{y}_2\bar{y}_3| \ge |\bar{y}_1\bar{y}_3| \ge d_c(y_1,y_3). \tag{2.3.3}$$




The other case is $\alpha+\beta \geq \pi$: In this case, since the broken line $\bar{y}_1\bar{y}_2\bar{y}_3$ lies outside the sector $\bar{y}_1O\bar{y}_3$, it follows that $|\bar{y}_1\bar{y}_2| + |\bar{y}_2\bar{y}_3| \geq |\bar{y}_1O| + |O\bar{y}_3|$. Therefore,

$$d_c(y_1, y_2) + d_c(y_2, y_3) = |\bar{y}_1\bar{y}_2| + |\bar{y}_2\bar{y}_3| \geq |\bar{y}_1O| + |\bar{O}\bar{y}_3| = r_1 + r_3. \tag{2.3.4}$$

Finally, note that

$$d_c(y_1, y_3) = \sqrt{r_1^2 + r_3^2 - 2r_1r_3\cos(d(x_1, x_3))} \leq \sqrt{r_1^2 + r_3^2 + 2r_1r_3} = r_1 + r_3.$$

□

**2.3.11 Proposition.** *Let $\tilde{\gamma} : [a, b] \to \mathrm{Con}(X)$ be a curve into the metric cone over $(X, d)$, $\tilde{\gamma}(t) = (\gamma(t), r(t))$.*

*(i) If $L(\gamma) \leq \pi$, then $L(\tilde{\gamma}) \geq \sqrt{r(a)^2 + r(b)^2 - 2r(a)r(b)\cos(L(\gamma))}$.*

*(ii) If $L(\gamma) \geq \pi$, then $L(\tilde{\gamma}) \geq r(a) + r(b)$.*

**Proof.** (i) Let $a = t_0 < \cdots < t_n = b$ be any partition, then $L(\tilde{\gamma}) \geq \sum_{i=0}^{n-1} d_c(\tilde{\gamma}(t_i), \tilde{\gamma}(t_{i+1}))$. As in the discussion preceding (2.3.3) we now pick points $\bar{y}_i \in \mathbb{R}^2$ such that their distance from the origin is $r(t_i)$ and $\measuredangle \bar{y}_iO\bar{y}_{i+1} = d(\gamma(t_i), \gamma(t_{i+1}))$ (always proceeding, say, clockwise on $S^1$ with these angles). Then the angle between $\bar{y}_0$ and $\bar{y}_n$ is $\sum_{i=0}^{n-1} d(\gamma(t_i), \gamma(t_{i+1}))$. Because this sum is $\leq L(\gamma) \leq \pi$, by the law of cosines we can calculate $|\bar{y}_0\bar{y}_n|$ via this angle, to obtain

$$\begin{aligned} L(\tilde{\gamma}) &\geq \sum_{i=0}^{n-1} d_c(\tilde{\gamma}(t_i), \tilde{\gamma}(t_{i+1})) = \sum_{i=0}^{n-1} |\bar{y}_i\bar{y}_{i+1}| \\ &\geq |\bar{y}_0\bar{y}_n| = \Big(r(a)^2 + r(b)^2 - 2r(a)r(b)\cos\Big(\sum_i d(\gamma(t_i), \gamma(t_{i+1}))\Big)\Big)^{\frac{1}{2}}. \end{aligned}$$

Taking the supremum over all partitions we arrive at our claim:

$$L(\tilde{\gamma}) \geq \sqrt{r(a)^2 + r(b)^2 - 2r(a)r(b)\cos(L(\gamma))}.$$

(ii) Due to (i) we only need to consider the case where $L(\gamma) > \pi$, Then we may pick a partition $a = t_0 < \cdots < t_n = b$ such that $\sum_{i=0}^{n-1} d(\gamma(t_i), \gamma(t_{i+1})) > \pi$. In the geometric representation employed in the proof of Proposition 2.3.10, the $d(\gamma(t_i), \gamma(t_{i+1}))$ correspond to the angles between consecutive points $\bar{y}_i$. Thus letting $y_i := \tilde{\gamma}(t_i)$ $(0 \leq i \leq n)$, we have $\sum_{i=0}^{n-1} \measuredangle \bar{y}_iO\bar{y}_{i+1} > \pi$, so by the analogue of (2.3.4) for finitely many points we obtain

$$L(\tilde{\gamma}) \geq \sum_{i=0}^{n-1} d_c(y_i, y_{i+1}) = \sum_{i=0}^{n-1} |\bar{y}_i\bar{y}_{i+1}| \geq r_0 + r_n = r(a) + r(b).$$

□

Our next aim is to show that $d$ is (strictly) intrinsic if and only if $d_c$ is. This will require some preparations. We first look at an instructive example:

**2.3.12 Example.** Let $X = S^2$, so $\mathrm{Con}(X) = \mathbb{R}^3$. A shortest path $\gamma : [0, a] \to S^2$ in $X$ then is an arc contained in a great circle of $X$, and so the cone over $\gamma$ is a planar sector. Any point in this sector has cone coordinates $(\gamma(\tau), t)$. Now if $\gamma$ is parametrized by arc length, $\tau$ and $t$ are precisely the polar coordinates in this planar sector, where the angle coordinate is measured with respect to the ray $[O, \gamma(0)]$.

More generally, we have:

**2.3.13 Lemma.** *Let $(X, d)$ be a length space with $\mathrm{diam}(X) \le \pi$, and let $\gamma : [0, L] \to X$ be a shortest path in $X$, parametrized by arclength. Then the cone over the image of $\gamma$ is (isometric to) a convex flat surface in $\mathrm{Con}(X)$.*

**Proof.** Denote by $(r, \varphi)$ polar coordinates in the plane and let $Q$ be the set of points in the plane with $\varphi$-coordinate between $0$ and $L$. Let $F : Q \to \mathrm{Con}(X)$, $F(r, \varphi) := (\gamma(\varphi), r)$. Then the cone over $\gamma$ is precisely the image of $F$. Moreover, $F$ is distance-preserving since

$$\begin{aligned} d_c^2(F(r,\varphi), F(r',\varphi')) &= r^2 + r'^2 - 2rr' \cos d(\gamma(\varphi), \gamma(\varphi')) = r^2 + r'^2 - 2rr' \cos(\varphi - \varphi') \\ &= d_{\mathbb{R}^2}((r,\varphi),(r,\varphi')). \end{aligned}$$

Thus Lemma 2.3.7 shows that $F(Q)$ is convex. Also, $F$ is an isometry and $Q$ is flat, so the claim follows. □

**2.3.14 Remark.** Based on the above results, we obtain a complete description of distance realizing paths in $\mathrm{Con}(X)$ for any length space with $\mathrm{diam}(X) \le \pi$: On the one hand, if $\gamma : [0, L] \to X$ is a shortest path in $X$ and $r, r' \ge 0$, then let $\alpha : [0, L] \to Q$ be a shortest path (i.e., a straight line) from $(r, 0)$ to $(r', L)$. Then $F \circ \alpha : [0, L] \to (\mathrm{Con}(X), d_c)$ is a distance realizing path from $(\gamma(0), r)$ to $(\gamma(L), r')$: In fact, by the proof of Lemma 2.3.13 we have

$$d_c(F \circ \alpha(0), F \circ \alpha(L)) = d_{\mathbb{R}^2}(\alpha(0), \alpha(L)) = L(\alpha) = L(F \circ \alpha),$$

where the last equality holds since $F$ is distance-preserving.
Conversely, let $\bar\gamma : [a, b] \to \mathrm{Con}(X)$ be a distance realizing path that doesn't pass through the origin. $\bar\gamma$ can be written in the form $\bar\gamma(t) = (\gamma(t), r(t))$, and we claim that the *projection* $\gamma : [a, b] \to X$ is a shortest path. Using the same notation as in Proposition 2.3.10, let $y_i := \bar\gamma(t_i)$, $i = 1, 2, 3$, $t_1 < t_2 < t_3$. Since $\bar\gamma$ is distance realizing,

$$d_c(y_1, y_3) = d_c(y_1, y_2) + d_c(y_2, y_3) = |\bar y_1 \bar y_2| + |\bar y_2 \bar y_3|, \tag{2.3.5}$$

and it suffices to show that, with $y_i = (x_i, r_i)$, we have $d(x_1, x_3) = d(x_1, x_2) + d(x_2, x_3)$. Again we distinguish two cases:
$\alpha + \beta < \pi$: Here, (2.3.3) and (2.3.5) imply $|\bar y_1 \bar y_2| + |\bar y_2 \bar y_3| = |\bar y_1 \bar y_3|$. Thus $\bar y_1$, $\bar y_2$ and $\bar y_3$ lie on one straight line. Further by definition (see proof of 2.3.10) the triangle $\Delta \bar y_1 O \bar y_3$ has angle $\alpha + \beta$ at $O$ and has sidelengths $r_1$, $r_3$ and $|\bar y_1 \bar y_2| + |\bar y_2 \bar y_3| = d_c(y_1, y_3)$. By the very definition of $d_c$, the angle of this triangle at $O$ must be $d(x_1, x_3)$. Consequently, $d(x_1, x_3) = \alpha + \beta = d(x_1, x_2) + d(x_2, x_3)$.
$\alpha + \beta = \pi$: this case cannot occur: Confer with the sketch in the proof of 2.3.10. Indeed, (2.3.5) implies that in (2.3.4) we must have equalities everywhere. But this is only possible if $r_1 = |\bar y_1 \bar y_2|$ and $r_3 = |\bar y_2 \bar y_3|$, meaning that $r_2 = 0$. This, however, contradicts the assumption that $\bar\gamma$ does not pass through the origin.
Finally, we turn to the case of paths that do pass through the origin. Clearly, every point $(x, r) \in \mathrm{Con}(X)$ is connected to $O$ by the unique shortest path $\{(x, t) : t \in [0, r]\}$. The concatenation of two such segments with endpoints $(x_1, r_1)$ and $(x_2, r_2)$ is a shortest path if and only if $d(x_1, x_2) = \pi$.

As a next step we drop the assumption that $\mathrm{diam}(X) \le \pi$. In this case we can no longer use (2.3.1) to define a metric on $\mathrm{Con}(X)$ because the triangle inequality will no longer hold in general. We do, however, still want (2.3.1) to hold for small distances and we still want $\mathrm{Con}(X)$ to be a length space. As we shall see below, the following definition satisfies these criteria (and is unique by Corollary 2.1.2):

**2.3.15 Definition.** *Let $(X, d)$ be a metric space. Then the cone distance $d_c(a, b)$ between two points $a = (x, t)$ and $b = (y, s)$ in* $\mathrm{Con}(X)$ *is defined by*

$$d_c(a,b) := \begin{cases} \sqrt{t^2 + s^2 - 2ts\cos(d(x,y))}, & d(x,y) \leq \pi \\ t + s, & d(x,y) \geq \pi. \end{cases}$$

Defining a new metric on $X$ by $\bar{d}(x, y) := \min(d(x, y), \pi)$, $d_c$ is the cone metric corresponding to $(X, \bar{d})$, so Proposition 2.3.10 implies that $d_c$ is a metric.
To gain some intuition we consider cones over a circle.

**2.3.16 Example (Cones over a circle).** Let $C$ be a circle in $\mathbb{R}^2$ of circumference $L$ and hence of radius $r = L/2\pi$. As a set the cone $K$ over $C$ consists of pairs $(x, t)$ with $x \in C$ and $t \in [0, \infty)$ with all points $(x, 0)$ identified. The idea is now that angles of sectors with vertex at 0 get scaled by the factor $L/2\pi$, that is they get shrunk if $L < 2\pi$, and the get enlarged if $L > 2\pi$.
More precisely let $a \in \mathbb{R}^2$ have polar coordinates $(r, \varphi)$. Then by definition of $d_c$ we have $d_c(0, a) = r$, i.e., radial distances are preserved. However, for points on different rays the distance is given by the cosine-formula, where the angle is replaced by the length of the corresponding section on $C$ between the two rays. Now if the circumference of the circle is $2\pi$ then that length is exactly the angle (in $\mathbb{R}^2$) between the rays, but if the circumference differs the lengths in the cone over $C$ are scaled. For an easy example take $a = (1, 0)$ and $b = (0, 1)$, see figures 2.2 and 2.3. Then we have

$$d_c(a,b) = \sqrt{1 + 1 - 2\cos(L/4)} \begin{cases} < \sqrt{2} & \text{if } L < 2\pi \\ > \sqrt{2} & \text{if } L > 2\pi. \end{cases} \tag{2.3.6}$$

More generally we have $d_c(a, b) < d_{\text{eucl.}}(a, b)$ for $L < 2\pi$ and $d_c(a, b) > d_{\text{eucl.}}(a, b)$ for $L > 2\pi$.

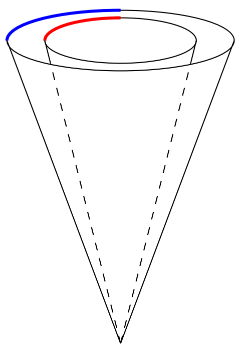


Figure 2.2: The distance between corresponding points on the cone over a larger circle (blue) is bigger than the one in a smaller circle (red).

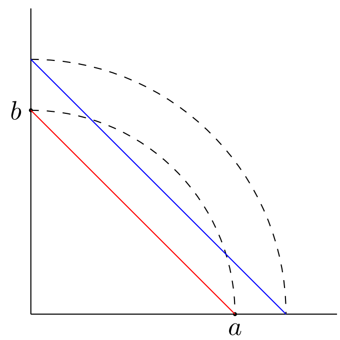



Figure 2.3: The blue line depicts the distance $d_c$ of the points in a cone over a cricle of circumference $> 2\pi$ compared to their distance (red) in one of circumference $2\pi$.

The scaling of the angles can be made even more explicit: Let $L > 2\pi$ and consider (with $a$, and $b$ as above) the triangle $\triangle a0b$ and a corresponding triangle $\triangle \bar{a}0\bar{b}$ with side lengths 1, 1 and $d_c(a, b) > \sqrt{2}$, see figures 2.4 and 2.5. Now its angle $\triangle \bar{a}0\bar{b}$ at the origin 0 satisfies

$$\measuredangle \bar{a}0\bar{b} > \frac{\pi}{2} = \measuredangle a0b. \tag{2.3.7}$$

With other words the angles in $K$ are fractional parts of $L$ while Euclidean angles are fractional parts of $2\pi$ and they are related by the scaling, i.e.,

$$\measuredangle \bar{a}0\bar{b} = \frac{2\pi}{L} \measuredangle a0b. \tag{2.3.8}$$

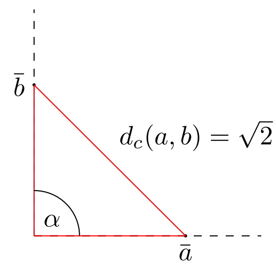



Figure 2.4: The angle of the triangle $\Delta\bar{a}0\bar{b}$ for the cone with circumference $2\pi$.

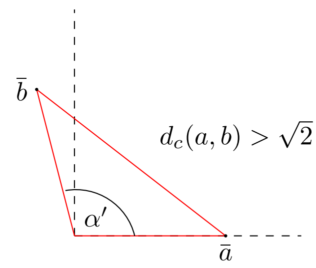



Figure 2.5: The angle of the triangle $\Delta\bar{a}0\bar{b}$ for the cone with circumference $> 2\pi$.

Next we consider (strictly) intrinsic metrics on cones. The fundamental result is:

**2.3.17 Theorem.** *The metric $d_c$ is intrinsic (resp. strictly intrinsic) if and only if $d$ is intrinsic (resp. strictly intrinsic) at distances less than $\pi$ (by which we mean that, whenever $d(x,y) < \pi$ there is a curve in $X$ from $x$ to $y$ whose length is arbitrarily close (resp. equal to) $d(x,y)$).*

**Proof.** Suppose first that $d_c$ is strictly intrinsic and let $x, y \in X$ with $d(x,y) < \pi$. Let $\tilde{\gamma}$ be a shortest path in $\mathrm{Con}(X)$ from $a := (x,1)$ to $b := (y,1)$. Then

$$L(\tilde{\gamma}) = d_c(a,b) = \sqrt{1^2 + 1^2 - 2\cos(d(x,y))} < 2, \tag{2.3.9}$$

so $\tilde{\gamma}$ does not pass through $O$. It therefore has a well-defined and continuous projection $\gamma$. Proposition 2.3.11 (ii) with $r(a) = r(b) = 1$, together with (2.3.9) implies that $L(\gamma) \leq \pi$. Hence Proposition 2.3.11 (i) gives ♣ `check estimates` ♣

$$\sqrt{1^2 + 1^2 - 2\cos(d(x,y))} = d_c(a,b) = L(\tilde{\gamma}) \geq \sqrt{2 - 2\cos(L(\gamma))}.$$

Thus $(d(x,y) \leq) L(\gamma) \leq d(x,y)$, i.e., $\gamma$ is minimizing.
Now suppose that $(\mathrm{Con}(X), d_c)$ is (merely) intrinsic, let $x, y \in X$ with $d(x,y) < \pi$ and let $\varepsilon > 0$. By assumption, there exists a curve $\tilde{\gamma}$ in $\mathrm{Con}(X)$ connecting $a = (x,1)$ to $b = (y,1)$ such that $L(\tilde{\gamma}) < d_c(a,b) + \varepsilon$. Hence ♣ `check estimates` ♣

$$d_c(a,b) = \sqrt{2 - 2\cos(d(x,y))} \geq L(\tilde{\gamma}) - \varepsilon \geq \sqrt{2 - 2\cos(L(\gamma))} - \varepsilon,$$

where in the last step we used Proposition 2.3.11 (i). Since we always have $L(\gamma) \geq d(x,y)$, setting $f(\eta) := \sqrt{2 - 2\cos(\eta)}$, we have shown that for any $\varepsilon > 0$ there exists some $\gamma$ with $f(d(x,y)) \leq f(L(\gamma)) \leq f(d(x,y)) + \varepsilon$. Since $f$ is continuous and strictly monotonically increasing, this implies that for any $\varepsilon > 0$ there exists some $\gamma$ with $L(\gamma) \in [d(x,y), d(x,y) + \varepsilon]$, i.e., $d$ is intrinsic at distances less than $\pi$.
Conversely, let $d$ be strictly intrinsic at distances less than $\pi$, let $a = (x,t), b = (y,s) \in \mathrm{Con}(X)$. If $d(x,y) < \pi$, then by Remark 2.3.14, from a distance realizing path connecting $x$ and $y$ in $X$ we obtain a distance realizing path from $a$ to $b$ in $\mathrm{Con}(X)$. If $d(x,y) \geq \pi$, then $d_c(a,b) = t + s$ and the concatenation of the straight segments from $a$ to $O$ and from $O$ to $b$ has precisely this length.
Finally, let $d$ be intrinsic at distances less than $\pi$, let $\varepsilon > 0$ and let $p = (x_p, r_p), q = (x_q, r_q) \in \mathrm{Con}(X)$. As shown in the previous case we may suppose that $d(x_p, x_q) < \pi$, since otherwise we already know there is a distance-realizing curve. By assumption there exists a continuous curve $\gamma : [0,1] \to X$ with $L(\gamma) < d(x_p, x_q) + \varepsilon$. Thus there exists some $\eta > 0$ such that for any partition $0 = t_0 < \cdots < t_n = 1$ of modulus smaller than $\eta$, we have (with $x_0 = x_p$, $x_1 = \gamma(t_1)$, $\ldots$, $x_n = \gamma(1) = x_q$):

$$\sum_{i=0}^{n-1} d(x_i, x_{i+1}) < d(x_p, x_q) + \varepsilon. \tag{2.3.10}$$

Now draw a triangle in the plane with lower vertex in the origin, left sidelength $r_p$, right sidelength $r_q$, and angle $\sum_{i=0}^{n-1} d(x_i, x_{i+1})$ at $O$. Let $r : [0,1] \to \mathbb{R}^2$ be a parametrization of the upper side of the triangle, and let $\tilde{\gamma} : [0,1] \to \mathrm{Con}(X)$, $t \mapsto (\gamma(t), \|r(t)\|)$. Let $\tilde{x}_i := \tilde{\gamma}(t_i)$, $r_i := \|r(t_i)\|$. Then the length of the upper side is $\sum_{i=0}^{n-1} d_c(\tilde{x}_i, \tilde{x}_{i+1})$, and if we insert rays from zero of angle $d(x_0, x_1)$ to $r_p$, then of angle $d(x_1, x_2)$ to the first ray, etc., we obtain a subdivision of the triangle with radial sidelengths $r_i$ and such that the corresponding segments on the upper line have lengths $d_c(\tilde{x}_0, \tilde{x}_1)$, $d_c(\tilde{x}_1, \tilde{x}_2)$, etc. Now compare this to the triangle with lower vertex in the origin, left sidelength $r_p$, right sidelength $r_q$, and angle $d(x_p, x_q)$ at $O$ (which, thereby, has upper sidelength $d_c(p,q)$), see figures 2.6 and 2.7. Then (2.3.10) implies that, given any $\tilde{\varepsilon} > 0$, for modulus $\eta$ sufficiently small,

$$\sum_{i=0}^{n} d_c(\tilde{x}_i, \tilde{x}_{i+1}) < d_c(p,q) + \tilde{\varepsilon}.$$

Therefore, $L(\tilde{\gamma}) \le d_c(p,q) + \tilde{\varepsilon}$, so $d_c$ is intrinsic.

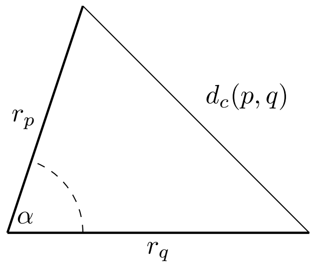



Figure 2.6: Here $\alpha = d(x_p, x_q)$

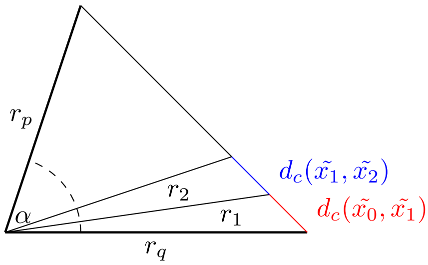



Figure 2.7: Here $\alpha = \sum d(x_i, x_{i+1})$ and the summands are the angles between the $r_i$

□

## 2.4 Angles in metric spaces

Our goal in this section is to define a notion of angle between paths in a metric space. To see how to go about this, let us first re-visit the familiar setting of Euclidean geometry. Let $\alpha, \beta : [0, \infty) \to \mathbb{R}^2$ be two rays in $\mathbb{R}^2$ with the same initial point $a = \alpha(0) = \beta(0)$. Given $t, s > 0$, by the law of cosines the angle $\measuredangle \alpha(t) a \beta(s)$ is given by

$$\arccos \frac{|a\alpha(t)|^2 + |a\beta(s)|^2 - |\alpha(t)\beta(s)|^2}{2|a\alpha(t)||a\beta(s)|}$$

(where $|pq| = \|p - q\|$ is the Euclidean distance). If we replace the rays by more general paths, the above expression will cease to be independent of $s$ and $t$. However, we may still reasonably expect that the limit $s, t \to 0$ will exist and will equal the angle between the paths (i.e., between their tangent vectors at $a$).

**2.4.1 Definition.** *Let $x, y, z$ be distinct points in a metric space $(X, d)$. The comparison angle $\tilde{\measuredangle} xyz$ (or $\tilde{\measuredangle}(x,y,z)$) at $y$ is*

$$\tilde{\measuredangle} xyz := \arccos \frac{d(x,y)^2 + d(y,z)^2 - d(x,z)^2}{2d(x,y)d(y,z)}$$

Geometrically, $\tilde{\measuredangle} xyz$ is the angle at $\bar{y}$ of a triangle $\bar{x}\bar{y}\bar{z}$ in $\mathbb{R}^2$ whose sidelengths satisfy $|\bar{x}\bar{y}| = d(x,y)$, $|\bar{y}\bar{z}| = d(y,z)$, and $|\bar{x}\bar{z}| = d(x,z)$: $\tilde{\measuredangle} xyz = \measuredangle \bar{x}\bar{y}\bar{z}$. The comparison triangle $\bar{x}\bar{y}\bar{z}$ is unique up to rigid motion.

**2.4.2 Definition.** *Let $\alpha, \beta : [0, \varepsilon) \to X$ be two paths in a length space $X$ with the same initial point $p = \alpha(0) = \beta(0)$. The* angle *between $\alpha$ and $\beta$ is defined as*

$$\measuredangle(\alpha, \beta) := \lim_{s,t \to 0} \tilde{\measuredangle}(\alpha(s), p, \beta(t))$$

*if this limit exists.*

Fixing $\alpha$ and $\beta$ as above, let

$$\theta(s,t) := \widetilde{\measuredangle}(\alpha(s), p, \beta(t)), \tag{2.4.1}$$

so $\measuredangle(\alpha,\beta) = \lim_{s,t\to 0}\theta(s,t)$. Now suppose in addition that $\alpha$ and $\beta$ are distance realizing paths parametrized by arclength. Then $d(p,\alpha(s)) = s$, $d(p,\beta(t)) = t$, and $\theta(s,t)$ is determined solely by $d(\alpha(s),\beta(t))$:

$$\theta(s,t) = \arccos\frac{s^2+t^2-d(\alpha(s),\beta(t))^2}{2st}. \tag{2.4.2}$$

Thus the angle $\measuredangle(\alpha,\beta)$ exists and equals $\theta_0 \in [0,\pi]$ if and only if

$$d(\alpha(s),\beta(t))^2 = s^2+t^2-2st\cos\theta_0 + o(st),$$

as $s,t\to 0$, where $o$ is the Landau symbol.

**2.4.3 Remark.** It can be shown (cf. [BH99, 1A.7 Cor.]) that for intersecting geodesics in Riemannian manifolds the angle always exists and equals the angle between the tangent vectors at the intersection point. For the special case of $\mathbb{R}^2$ this follows directly from the above definitions since geodesics then are just straight lines: $\alpha(s) = p+sv$, $\beta(t) = p+tw$, and we may assume that $v$ and $w$ are unit vectors. Then $\|\alpha(s)-\beta(t)\|^2 = \langle sv-tw, sv-tw\rangle = s^2+t^2-2st\langle v,w\rangle$, so that $\theta(s,t) = \arccos\langle v,w\rangle$ for all $s,t$.

**2.4.4 Example.** Even between shortest paths $\alpha,\beta$ in a metric space, the angle $\measuredangle(\alpha,\beta)$ need not exist in general. As an Example, consider two rays $\alpha,\beta$ in $\mathbb{R}^2$ emanating from the origin $O$ at an angle $\varphi > 0$, equipped with the induced metric from $\mathbb{R}^2$. Now for each $n\in\mathbb{N}$, identify the two points on the rays that have distance $2^{-n}$ from $O$. Then by (2.4.2), $\theta(2^{-n},2^{-n}) = \arccos(1) = 0$, whereas if $s_n,t_n\to 0$ in such a way that neither $s_n$ nor $t_n$ equal $2^{-k}$ for any $k,n$, then $\theta(s_n,t_n)\to\varphi$. Consequently, $\lim_{s,t\to 0}\theta(s,t)$ does not exist.

**2.4.5 Proposition.** *Let $(X,d)$ be a length space.*

*(i) Every shortest path in $X$ has zero angle with itself at every point.*

*(ii) If the concatenation $[abc]$ of two shortest segments $[a,b]$ and $[b,c]$ in $X$ is itself a shortest path, then the angle between $[b,a]$ and $[b,c]$ is $\pi$.*

**Proof.** (i) Setting $\alpha=\beta$ in (2.4.2), and noting that $d(\alpha(s),\beta(t)) = |s-t|$ because $\alpha$ is parametrized by arclength, we obtain

$$\theta(s,t) = \arccos\frac{s^2+t^2-(s-t)^2}{2st} = \arccos(1) = 0.$$

(ii) This time, $\alpha(s) = \beta(-s)$, and $d(\alpha(s),\beta(t)) = s+t$, so $\theta(s,t) = \arccos(-1) = \pi$. □

While the angle between paths may not exist (cf. Example 2.4.4), the *upper angle* always does:

**2.4.6 Definition.** *Let $\alpha,\beta : [0,\varepsilon)\to X$ be two paths in a length space $X$ with the same initial point $p=\alpha(0)=\beta(0)$. The* upper angle *between $\alpha$ and $\beta$ is defined as*

$$\measuredangle_U(\alpha,\beta) := \limsup_{s,t\to 0}\widetilde{\measuredangle}(\alpha(s),p,\beta(t)).$$

**2.4.7 Theorem (Triangle inequality for angles).** *Let $\gamma_1,\gamma_2,\gamma_3$ be curves in a length space $X$ emanating from the same point $p$. Assume that the angles $\alpha_1 := \measuredangle(\gamma_2,\gamma_3)$, $\alpha_2 := \measuredangle(\gamma_1,\gamma_3)$, and $\alpha_3 := \measuredangle(\gamma_1,\gamma_2)$ exist. Then*

$$\alpha_3 \le \alpha_1+\alpha_2.$$

**Proof.** We may assume that $\alpha_1 + \alpha_2 < \pi$ since otherwise there is nothing to prove. According to the definition of angle, each $\alpha_i$ is a limit of an appropriate function $\theta$ as in (2.4.1). Hence given any $\varepsilon > 0$, for sufficiently small $s, t, r$ we have

$$|\alpha_1 - \theta(b,c)| \le \varepsilon, \quad |\alpha_2 - \theta(a,c)| \le \varepsilon, \quad |\alpha_3 - \theta(a,b)| \le \varepsilon, \tag{2.4.3}$$

where $a = a(s) = \gamma_1(s)$, $b = b(t) = \gamma_2(t)$, and $c = c(r) = \gamma_3(r)$ (where we use an ad-hoc notation to label the functions to the appropriate paths).
The proof rests on comparison with certain model triangles $\Delta\bar{p}\bar{a}\bar{c}$ and $\Delta\bar{p}\bar{b}\bar{c}$ in the Euclidean plane. Thus we pick four points $\bar{p}, \bar{a}, \bar{b}, \bar{c} \in \mathbb{R}^2$ such that their Euclidean distances satisfy

$$|\bar{p}\bar{a}| = d(p,a), \;\; |\bar{p}\bar{b}| = d(p,b), \;\; |\bar{p}\bar{c}| = d(p,c), \;\; |\bar{c}\bar{b}| = d(c,b), |\bar{a}\bar{c}| = d(a,c)$$

and such that $\bar{a}$ and $\bar{b}$ lie on opposite sides of the line $(\bar{p}\bar{c})$.

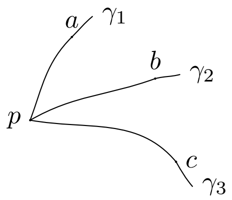


Figure 2.8

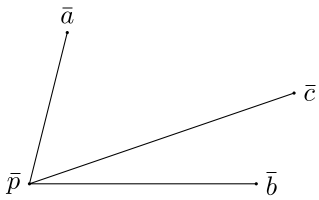


Figure 2.9

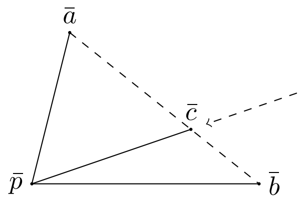


Figure 2.10

Fixing $a$ and $b$ and moving $c$ towards $p$ (i.e., fixing $s$ and $t$ and decreasing $r$), for $c$ close to $p$ the points $\bar{p}$ and $\bar{c}$ lie on the same side of the line $(\bar{a}\bar{b})$. On the other hand, if we fix $r$ and make $s$ and $t$ sufficiently small, then $\bar{p}$ and $\bar{c}$ lie on opposite sides of $(\bar{a}\bar{b})$. By continuity, therefore, we can pick values for $s, t, r$ such that $\bar{c}$ lies exactly on the segment $[\bar{a}, \bar{b}]$. More precisely, picking $\bar{p} = (0,0)$, and $\bar{c} = (d(p,c), 0)$, the choices of $\bar{a}$ and $\bar{b}$ are uniquely determined and depend continuously on the distances $d(p,a), d(p,b), d(c,a), d(c,b)$, and $d(p,c)$. Hence the points move continuously as we vary $s, t, r$, see figures 2.8 2.9 2.10.
Having reached this configuration $\bar{c} \in [\bar{a}, \bar{b}]$, we obtain

$$|\bar{a}\bar{b}| = |\bar{a}\bar{c}| + |\bar{c}\bar{b}| = d(a,c) + d(c,b) \ge d(a,b).$$

We now add a point $\tilde{b}$ to the plane such that

$$|\bar{p}\tilde{b}| = |\bar{b}\bar{p}| = d(p,b), \;\; |\bar{a}\tilde{b}| = d(a,b),$$

and such that $\tilde{b}$ lies on the same side of the line $(\bar{p}\bar{a})$ as $\bar{b}$, see 2.11.
By definition, $\theta(a,b)$ is the angle $\measuredangle\bar{a}\bar{p}\tilde{b}$ in the planar triangle $\Delta\tilde{b}\bar{p}\bar{a}$ at $\bar{p}$. Also, $\theta(a,c) = \measuredangle\bar{a}\bar{p}\bar{c}$, and $\theta(b,c) = \measuredangle\bar{b}\bar{p}\bar{c}$. It follows that

$$\theta(a,c) + \theta(b,c) = \measuredangle\bar{a}\bar{p}\bar{b}.$$

Now comparing the triangles $\Delta\bar{b}\bar{p}\bar{a}$ and $\Delta\tilde{b}\bar{p}\bar{a}$, we see that they have two sides of equal lengths, while for the third sides we have $|\bar{a}\bar{b}| \ge |\bar{a}\tilde{b}|$. Therefore their angles must satisfy $\measuredangle\bar{a}\bar{p}\bar{b} \ge \measuredangle\bar{a}\bar{p}\tilde{b}$. In terms of the $\theta$-functions, this means

$$\theta(a,c) + \theta(b,c) \ge \theta(a,b).$$

Combining this with (2.4.3), it follows that $\alpha_3 \le \alpha_1 + \alpha_2 + 3\varepsilon$. Since $\varepsilon > 0$ was arbitrary, this concludes the proof of the triangle inequality.
Finally, we note that the above configuration (with $\gamma_3$ 'running between $\gamma_1$ and $\gamma_2$') is the only non-trivial one to consider: if, e.g., $\gamma_2$ were 'between' $\gamma_1$ and $\gamma_3$, then the above would immediately give $\alpha_3 \le \alpha_2 \le \alpha_1 + \alpha_2$.

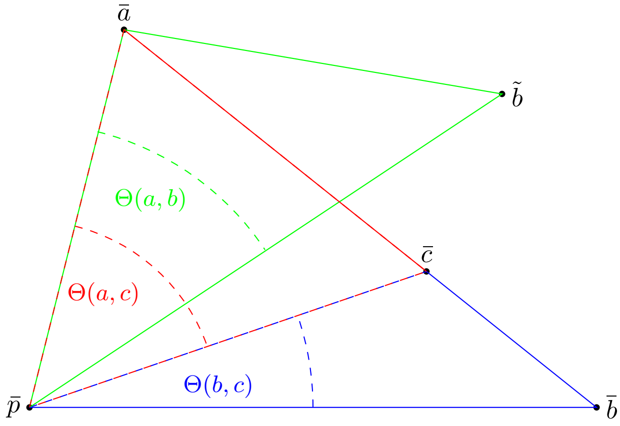



Figure 2.11: The possible situation of the angles after adding $\tilde{b}$

$\square$

# Chapter 3

# Spaces of Bounded Curvature

As we have seen in earlier chapters, general length spaces can be rather nasty objects and most results require additional properties. In this chapter we will introduce the most important such property, which is essentially geometric in nature: *curvature bounds*. Loosely speaking, curvature bounds from above and below correspond to a certain degree of convexity and concavity of the distance function, respectively.
We will first concentrate on the two most important curvature bounds, namely spaces of non-negative and spaces of non-positive curvature. While these two classes of spaces enjoy very distinct and different properties the metric space machinery to deal with them can be developed mainly in parallel. Later on we will also be concerned with the general case of spaces of curvature bounded below or above by non-vanishing constants. Such length spaces which obey a curvature bound either from below of above are called *Alexandrov spaces*.

## 3.1 Basic definitions

In this section we will introduce the main definitions for spaces of non-negative and non-positive curvature respectively. These definitions will be built upon comparison with Euclidean space and will result in spaces which neither topologically or metrically 'look like' Euclidean space but will share certain features with them. E.g. it is possible to define angles in Alexandrov spaces which is *not* possible in general normed vector spaces, see Section 2.4.
We will give several equivalent definitions of non-positively (resp. non-negatively) curved spaces formalizing the following ideas

- Distance functions are not less convex (resp. concave) than for the Euclidean plane.
- Geodesics emanating from a point diverge at least as fast as (resp. not faster than) in the Euclidean plane.
- Triangles are not 'thicker' (resp. not 'thinner') than Euclidean triangles with the same side lengths.

The sphere is the key example of a space of non-negative curvature; all the above properties can easily be seen to hold: spherical geodesic triangles look 'fat' as compared to their Euclidean counterparts, spherical geodesics emanating from a point are focused rather than diverge linearly as their Euclidean counterparts do.
More generally every convex surface in $\mathbb{R}^3$ (i.e., any boundary of a convex body) is a non-negatively curved space. On the other hand every smooth saddle surface in $\mathbb{R}^3$ (i.e., any surface looking locally like a hyperbolic paraboloid) is a non-positively curved space. Moreover there is the following clear connection to Riemannian geometry: Every Riemannian manifold is a non-positively (resp. non-negatively) curved space iff its sectional curvatures are non-positive (resp. non-negative), see e.g. [BBI01, chapter 6.5] for the two dimensional case.

**3.1.1 Convention.** In this chapter we will assume all length spaces to be *connected* (equivalently equipped with a finite metric) and to be *strictly intrinsic*, i.e., such that all pairs of points can be connected by a shortest path.

### 3.1.1 Comparison via distance functions

Here we introduce our first definition of a curvature bound. Let $X$ be a length space (as specified in the above convention) and let $p \in X$. Then the *distance to* $p$ is the real valued function $d_p$ on $X$ given by

$$d_p(x) = d(p, x). \tag{3.1.1}$$

Now let $[ab] = \gamma : [0, T] \to X$ be a shortest path parametrized by arc length. Then we introduce the so-called *one-dimensional distance function* by

$$g(t) = d(p, \gamma(t)) = d_p \circ \gamma(t), \tag{3.1.2}$$

i.e., the restriction of the distance function to $p$ to the segment $\gamma$.
Next we want to compare $g$ to an appropriate one-dimensional distance function in the Euclidean plane. To do so we transfer the above construction into the Euclidean plane in the following way: Choose a *comparison segment* $[\bar{a}\bar{b}]$ in the Euclidean plane of the same length as $[ab]$ and choose a *reference point* $\bar{p}$ such that its Euclidean distance to $\bar{a}$ and to $\bar{b}$ equals the distance between $p$ and $a$ respectively $b$., i.e.,

$$\|\bar{a} - \bar{p}\| = d_p(a) = d(p, a) \text{ and } \|\bar{b} - \bar{p}\| = d_p(b) = d(p, b). \tag{3.1.3}$$

Of course this *comparison configuration* is unique up to rigid motions. Now we regard the segment $[\bar{a}\bar{b}]$ as a path $\gamma_0$ parametrized by arc length, i.e., $\gamma_0(0) = \bar{a}$ and $\gamma_0(T) = \bar{b}$, and finally we define as follows.

**3.1.2 Definition (Comparison function).** *With the above notations we call* $g_0(t) := \|\bar{p} - \gamma_0(t)\|$*, i.e., the Euclidean distance from* $\bar{p}$ *restricted to the comparison segment* $[\bar{a}, \bar{b}]$ *the* comparison function *for* $g$*.*

Following the lay out specified above, the distance function of a non-positively (resp. non-negatively) curved space should be more convex (resp. more concave) than that of the Euclidean plane and we are going to define the respective spaces by the condition $g_0(t) \geq g(t)$ (resp. $g_0(t) \leq g(t)$). However, we also want our definition to be local and hence formulate it as follows.

**3.1.3 Definition (Distance condition).** *We say that a length space*[1] $(X, d)$ *is* non-positively curved *(resp.* non-negatively curved*) if every point in* $X$ *has a neighbourhood* $U$ *such that the following holds: For all points* $p \in U$ *and all segments* $\gamma \in U$ *the comparison function* $g_0$ *for the one-dimensional distance function* $g(t) = d_p(\gamma(t))$ *satisfies*

$$g_0(t) \geq g(t) \text{ (resp. } g_0(t) \leq g(t)) \text{ for all } t \in [0, T]. \tag{3.1.4}$$

To gain some working knowledge with the notions introduced above we next discuss some examples in detail.

**3.1.4 Example (Three rays glued).** We consider the space $R_{(3)}$ obtained from gluing three copies of the ray $[0, \infty)$ at the point 0, see 3.1. Then $R_{(3)}$ has non-positive curvature.
To see this denote by $O$ the common point of the three rays. Now every shortest path in $R_{(3)}$ is either a segment in one ray or a concatenation of two segments in two different rays. Now let $\gamma : [0, T] \to R_{(3)}$ be such a shortest path and let $p \in R_{(3)}$. If two of the three points $a := \gamma(0)$, $b := \gamma(T)$, and $p$ belong two the same ray then the statement is trivial. Indeed $\gamma$ and $p$ are then contained in the union of two rays which is isometric to $\mathbb{R}$.

[1] Recall Convention 3.1.1!

So we only need to consider the case when the three points $a$, $b$, and $p$ belong to different rays. For every $x \in [O,a]$ we have $|px| = |pa| - |ax|$ , see Figure 3.1. For the one-dimensional distance function this gives

$$g(t) = g(0) - t \qquad \text{for } \gamma(t) \in [O,a]. \tag{3.1.5}$$

On the other hand for the comparison function $g_0$ we have by the triangle inequality (see Figure 3.1, right) for such $t$

$$g_0(t) \geq g_0(0) - t, \tag{3.1.6}$$

and hence since $g_0(0) = g(0)$ the desired inequality $g_0(t) \geq g(t)$ follows for all $t$ with $\gamma(t) \in [O,a]$. The remaining case $\gamma(t) \in [O,b]$ is completely analogous. Hence $R_{(3)}$ is non-positively curved as claimed.

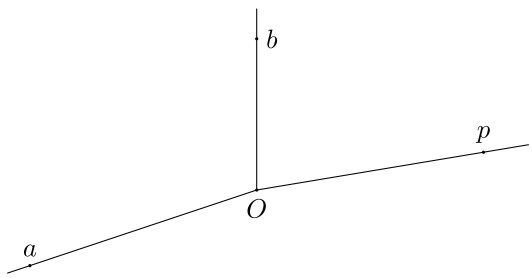



Figure 3.1: a possible picture of $R_{(3)}$

**3.1.5 Example (Cones over a circle).** Let $K$ be a cone over a circle of length $L$, see Section 2.3. Then $K$ is

- non-negatively curved iff $L \leq 2\pi$, and
- non-positively curved iff $L \geq 2\pi$.

By Lemma 2.3.13 the cone over a circle if flat outside the vertex: Every subcone over a segment of length $\alpha \leq \max\{L/2, \pi\}$ is convex and isometric to a planar sector of angular measure $\alpha$.
For a shortest path $\gamma : [0,T] \to K$ and a point $p \in K$ consider the triangle $\triangle$ composed of the three shortest paths between the points $a := \gamma(0)$, $b := \gamma(T)$, and $p$. There are two possibilities:

(1) $\triangle$ bounds a region *not* containing $O$ or one of the points $a$, $b$, $p$ coincides with $O$, and

(2) $\triangle$ bounds a region containing $O$, or some of its sides pass through $O$.

In the first case the triangle is isometric to a triangle in the plane and so the one-dimensional distance function $g$ coincides with its comparison function $g_0$.
We now consider the second case separately for $L < 2\pi$ and $L > 2\pi$. Observe that the case $L = 2\pi$ is again trivial since then the cone is isometric to $\mathbb{R}^2$.

(i) $L < 2\pi$: We cut $\mathbb{R}^2$ along the segments $[0a]$, $[0b]$, and $[0p]$, see figure 3.2. Then each of the ensuing sectors is isometric to a planar sector since $L < 2\pi$, again by Lemma 2.3.13. Now again since the sum of the angles of the sectors is bounded above by $2\pi$ we may put together these sectors in $\mathbb{R}^3$ to form a wedge with vertices $\bar{a}$, $\bar{b}$ , $\bar{p}$, and 0, see figure 3.3. The surface of this wedge with the gluing metric is isometric to $K$.

Now $\triangle\bar{a}\bar{b}\bar{p}$ which lies in the plane spanned by $\bar{a}$, $\bar{b}$ and $\bar{p}$ in $\mathbb{R}^3$ is a comparison triangle for $\triangle abp$. We then have

$$g_0(t) = d_{\text{eucl.}}(\bar{p}, \gamma(t)) \leq g(t), \tag{3.1.7}$$

where the latter is the intrinsic distance in $K$ (measured along the boundary of the wedge). So $g_0(t) \leq g(t)$ and $K$ is non-negatively curved.

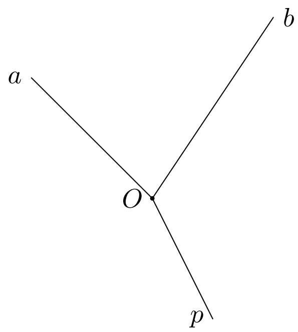



Figure 3.2

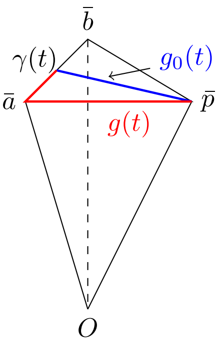



Figure 3.3

(ii) $L > 2\pi$: Again $\triangle ab0$, $\triangle ap0$, and $\triangle bp0$ are flat, i.e., isometric to planar triangles. Now consider comparison triangles $\triangle\bar{a}\bar{b}\bar{0}$ and $\triangle\bar{b}\bar{p}\bar{0}$ placed on different sides of the segment $[\bar{0}\bar{b}]$, see figure 3.4. We have for the angles $\alpha := \measuredangle a0b$, $\beta := \measuredangle p0b$ and $\gamma := \measuredangle a0p$ that $\alpha+\beta+\gamma = L$. Moreover we have (cf. Example 2.3.16) for $\bar{\alpha} := 2\pi/L\,\alpha$, $\bar{\beta} := 2\pi/L\,\beta$, and $\varphi := \measuredangle\bar{a}\bar{0}\bar{p}$ that $\bar{\alpha} + \bar{\beta} + \bar{\gamma} = 2\pi$. This implies

$$\varphi = 2\pi - (\bar{\alpha} + \bar{\beta}) = 2\pi - \frac{2\pi}{L}(\alpha + \beta) = \frac{2\pi}{L}(L - (L - \gamma)) = \frac{2\pi}{L}\gamma \leq \gamma, \tag{3.1.8}$$

and so

$$d_{\text{eucl.}}(\bar{a}, \bar{p}) = \sqrt{1 + 1 - 2\cos(\varphi)} \leq \sqrt{1 + 1 - 2\cos(\gamma)} = d_c(a, p). \tag{3.1.9}$$

Now we turn $\triangle\bar{b}\bar{0}\bar{p}$ around $\bar{b}$ until $|\bar{a}\bar{b}|$ equals $d_c(a,p)$, see figure 3.5. Next we proceed along the same lines with the other triangles to obtain the configuration shown in figure 3.6 . There we can see that we obviously have

$$g_0(t) \geq g(t), \tag{3.1.10}$$

and $K$ has non-positive curvature.

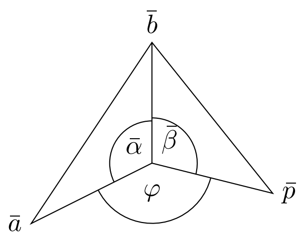



Figure 3.4

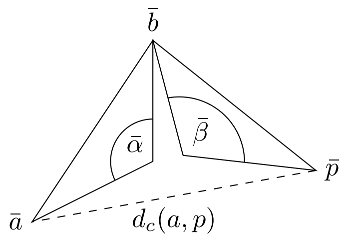



Figure 3.5

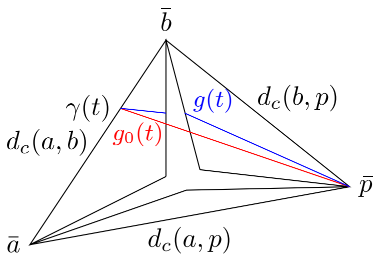



Figure 3.6

**3.1.6 Example (Cone over a segment).** Let $X$ be a cone over a segment of length $L$, then

- $X$ is non-positively curved for any $L$, and
- $X$ is non-negatively curved iff $L \leq \pi$.

Indeed by Lemma 2.3.13 $X$ is flat hence non-positively and non-negatively curved if $L \leq \pi$ and we are left with arguing that $X$ is non-positively curved for $L > \pi$.
If $L > \pi$ we can choose a segment $[ab]$ in $X$ such that $d(a,b) > \pi$, see figure 3.7 and hence $[ab]$ passes through the vertex $O$ of $X$ since by definition 2.3.15 in this case $d_c(a,b)$ is the sum of the radii $r_1, r_2$ (see figure). But then clearly the comparison function $g_0$ is larger than $g$ (the comparison triangle is 'fatter').

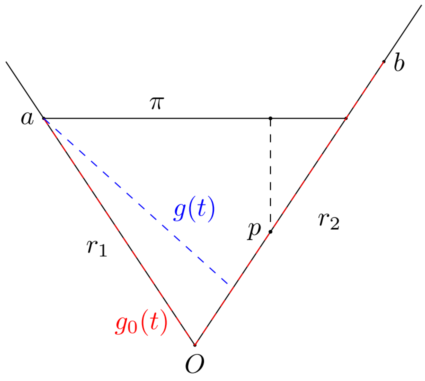



Figure 3.7

**3.1.7 Example (The 1-norm).** We next consider $\mathbb{R}^2$ with the norm $\|(x_1,x_2)\|_1 = |x_1| + |x_2|$. This space is neither non-positively nor non-negatively curved.
First observe that a straight line is always a shortest path in any normed vector space. Indeed, the length of a straight line segment is always the distance between its endpoints. However, there might be other shortest paths as well.
To begin with recall that the unit sphere in $(\mathbb{R}^2, \| \|_1)$ is the diamond-shaped square with the vertices $(0,1)$, $(1,0)$, $(0,-1)$, and $(-1,0)$. Moreover let $p = 0$.
We first consider the straight line segment $\gamma$ from $(-1,0)$ to $(0,1)$, which has length 2, see figure 3.8, left. We parametrize $\gamma$ by arc length as follows: $\gamma : [-1,1] \to \mathbb{R}^2$, $\gamma(t) = 1/2(-1+t, 1+t)$. Then obviously the one-dimensional distance function satisfies $g(t) = 1$. As comparison configuration in $(\mathbb{R}^2, \| \|_2)$ we choose $\bar{\gamma}$ the straight line segment from $(-1,0)$ to $(1,0)$ parametrized as $\bar{\gamma} : [-1,1] \to \mathbb{R}^2$, $\gamma(t) = (t,0)$, see figure 3.8, right. Then $\bar{p}$ which has to be at Euclidean distance 1 both from $(-1,0)$ and $(1,0)$ has to be $\bar{p} = 0$. But then $g_0(t) = |t|$ and so

$$g_0(t) < g(t) \text{ for all } -1 < t < 1. \tag{3.1.11}$$

Hence $g_0 \not\geq g$ and so $(\mathbb{R}^2, \| \|_1)$ is not non-positively curved.

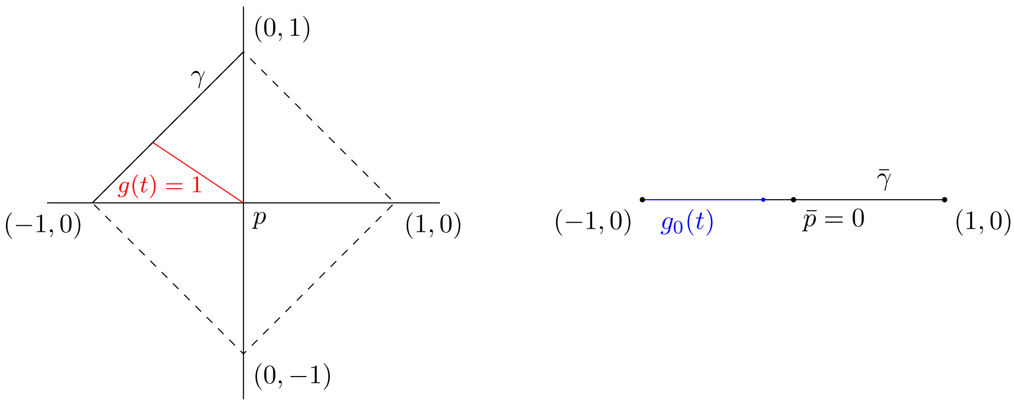



Figure 3.8

To see that the space is also not non-negatively curved we consider the straight line segment $\gamma : [0,1] \to \mathbb{R}^2$, $\gamma(t) = (1/2, 1/2 - t)$ connecting $(1/2, 1/2)$ with $(1/2, -1/2)$, see figure 3.9, left. Then $g(t) = 1/2 + |1/2 - t|$ and hence $g(0) = 1 = g(1)$, and $g(1/2) = 1/2$. Now we choose $\bar{\gamma} : [-1/2, 1/2] \to \mathbb{R}^2$, $\bar{\gamma}(t) = (t,0)$ and $\bar{p} = (0, -\sqrt{3}/2)$, see figure 3.9, right. Then $g_0(0) = 1 = g_0(1)$ but

$$g_0\Big(\frac{1}{2}\Big) = \frac{\sqrt{3}}{2} > \frac{1}{2} = g\Big(\frac{1}{2}\Big), \tag{3.1.12}$$

and so $g_0 \not\leq g$ and we are done.

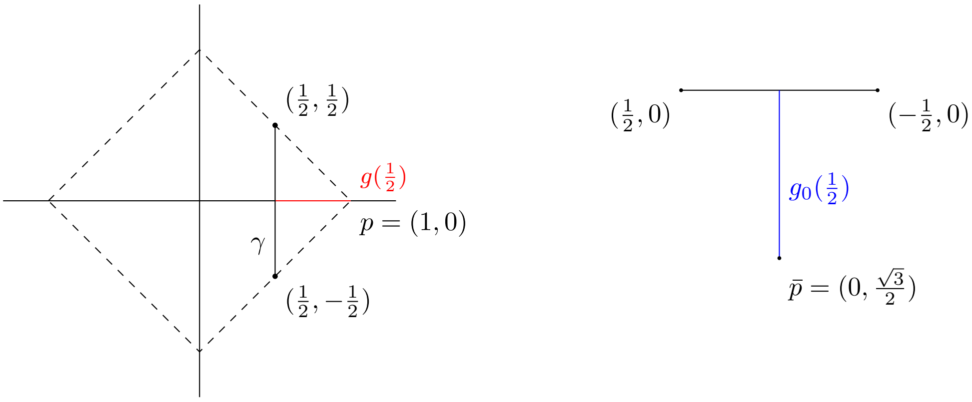



Figure 3.9

### 3.1.2 Distance comparison via triangles

In this section we reformulate the distance condition of Definition 3.1.3 to gain some additional insight into the geometry of non-positively and non-negatively curved spaces.

Let $X$ be a (connected and strictly intrinsic) length space. By a *triangle* in $X$ we mean any collection of three points $a$, $b$, $c \in X$ (called the vertices) connected by three shortest paths (called sides) denoted by $[ab]$, $[bc]$, and $[ca]$. We write $\triangle abc$ for the triangle and $|ab|$, $|bc|$, and $|ca|$ for the length of its sides. Observe that the vertices alone do *not* determine a triangle since there might be different shortest paths between the same pair of vertices.

Now for each triangle $\triangle abc$ in $X$ we construct a triangle $\triangle \bar{a}\bar{b}\bar{c}$ in the Euclidean plane with the same side lengths, i.e.,

$$|ab| = |\bar{a}\bar{b}|, \ |bc| = |\bar{b}\bar{c}|, \ \text{and} \ |ca| = |\bar{c}\bar{a}|. \tag{3.1.13}$$

We call such a triangle $\triangle \bar{a}\bar{b}\bar{c}$ a *comparison triangle* for $\triangle abc$. Clearly comparison triangles are determined up to rigid motion of the Euclidean plane.

Now we can formulate our new condition.

**3.1.8 Definition (Triangle condition).** *We say that a length space*[2] $(X, d)$ *is* non-positively curved *(resp.* non-negatively curved*) if every point in $X$ has a neighbourhood $G$ such that the following holds: For every triangle $\triangle abc \in G$ and every point $d \in [ac]$ we have*

$$|db| \leq |\bar{d}\,\bar{b}| \quad (\textit{resp. } |db| \geq |\bar{d}\,\bar{b}|), \tag{3.1.14}$$

*where $\bar{d}$ is the point on the side $[\bar{a}\bar{c}]$ of the comparison triangle $\triangle \bar{a}\bar{b}\bar{c}$ such that $|\bar{a}\bar{d}| = |ad|$.*

Observe that Definition 3.1.8 is a word by word translation of the distance condition of Definition 3.1.3: Indeed as can easily be seen from figure 3.10 the distance $|bd|$ with $d = \gamma(t)$ and $\gamma$ an arc length parametrization of the side $[ac]$ corresponds to the one-dimensional distance function via $g(t) = d(b, \gamma(t))$.

To obtain a vivid picture of the triangle condition take a look at figure 3.10: In the non-positively curved situation the distance between any point $d \in [ac]$ and the vertex $b$ in the triangle $\triangle abc$ is smaller than the corresponding distance in the Euclidean comparison triangle $\triangle \bar{a}\bar{b}\bar{c}$. Hence the triangles in a non-positively curved space are 'thinner' or 'skinnier' than in the plane. Conversely the triangle in a non-negatively curved space are 'fatter' than their flat comparison triangles.

[2]Recall Convention 3.1.1!

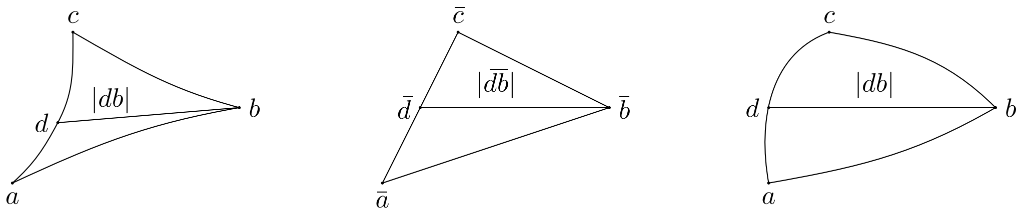



Figure 3.10: On the left a triangle in a non-positivly curved space, on the right one in a non-negatively curved space and in the center is the comparison triangle, the distance from $a$ to $d$ (resp. $\bar{a}$, $\bar{d}$) is the same in all three

Traditionally non-positively curved spaces in the sense of Definition 3.1.8 are called CAT(0)-spaces. Here CAT stand for comparison of Cartan-Alexandrov-Toponogov and (0) indicated that we compare with flat space, i.e., that we impose a zero upper curvature bound. Comparing with other spaces than the Euclidean plane (such as spheres or hyperbolas) one defines CAT(k)-spaces, $k \in \mathbb{R}$ which have curvature bounded above by $k$. In the case of lower curvature bounds one usually speaks of Alexandrov spaces with curvature bounded below by $k$. Note, however, that we can speak of non-positively (non-negatively) curved spaces or of spaces with curvature bounded above (below) *without having a notion of curvature* for length spaces at all.

A neighbourhood $G$ as in Definition 3.1.8 is called a *normal region*. Observe that one can always choose a normal region $U$ in such a way that all shortest paths starting and ending in $U$ are still contained in a possibly larger normal domain. Indeed, in any normal domain around a point $p$ choose a ball of radius $B_r(p)$ such that it has distance $2r$ from $U^c$, see figure 3.11. Then any shortest path $[xy]$ with endpoints in $B_r(p)$ has length less than $2r$ since otherwise the path from $x$ to $y$ via $p$ would be shorter than $[xy]$. Therefore $[xy]$ is entirely contained in $U$.

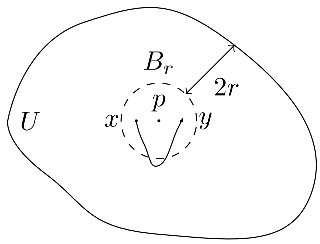



Figure 3.11

Finally we remark that contrary to the statement of [BBI01, Ex. 4.1.11] the triangle condition of Definition 3.1.8 is not implied by the simpler condition just imposing the inequalities for the midpoint $d$ of $[ac]$. A counterexample can be found in [BH99, Ex. 1.18, p. 168].

### 3.1.3 Angle comparison for triangles

By taking a look at the ‘fat’ and the ’skinny’ triangles of figure 3.10 one is lead to the following observation: ‘fat’ triangles should have larger angles than their Euclidean counterparts and ’thin’ triangles should correspondingly have smaller angles. Indeed this observation leads to yet another definition of Alexandrov spaces: A space is of non-positive curvature if the angles of all sufficiently small triangles exist and if they are not larger than the corresponding angles of a comparison triangle in the Euclidean plane. In the case of non-negative curvature in addition to the usual replacement of ‘not larger’ by ’not less’ one assumes[3] that the sum of adjacent angles equals $\pi$. The formal definition is as follows.

**3.1.9 Definition (Angle condition).** *A length space*[4] *$X$ is a space of*

*(i)* non-positive *curvature if every point of $X$ has a neighbourhood such that for every triangle $\triangle abc$ in that neighbourhood the angles $\measuredangle bac$, $\measuredangle cba$, and $\measuredangle acb$ are defined and satisfy*

$$\measuredangle bac \leq \tilde{\measuredangle} bac, \quad \measuredangle cba \leq \tilde{\measuredangle} cba, \quad \measuredangle acb \leq \tilde{\measuredangle} acb, \tag{3.1.15}$$

*where $\tilde{\measuredangle} bac$ denotes the comparison angle (cf. Definition 2.4.1), i.e., $\tilde{\measuredangle} bac = \measuredangle \bar{b}\bar{a}\bar{c}$ where $\triangle \bar{b}\bar{a}\bar{c}$ is the comparison triangle.*

[3] It is not clear whether this condition is really necessary, cf. the footnote on p. 108 of [BBI01].

[4] Recall Convention 3.1.1!

*(ii)* non-negative *curvature if every point of $X$ has a neighbourhood such that for every triangle $\triangle abc$ in that neighbourhood the angles $\measuredangle bac$, $\measuredangle cba$, and $\measuredangle acb$ are defined and satisfy*

$$\measuredangle bac \geq \tilde{\measuredangle} bac, \quad \measuredangle cba \geq \tilde{\measuredangle} cba, \quad \measuredangle acb \geq \tilde{\measuredangle} acb, \tag{3.1.16}$$

*and, in addition, the following holds: for any two shortest paths $[pq]$, $[rs]$ where $r$ is an inner point of $[pq]$ we have that (see figure 3.12)*

$$\measuredangle prs + \measuredangle srq = \pi. \tag{3.1.17}$$

While the triangle condition of Definition 3.1.8 was just a reformulation of the distance condition in Definition 3.1.3, the equivalence of these definitions with the angle condition of Definition 3.1.9 is not so obvious. In fact we will prove it only in the section 3.3 and meanwhile gain some more experience by looking at a number of examples in the next section.

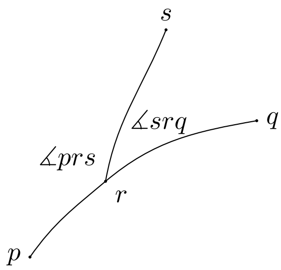



Figure 3.12

## 3.2 Examples

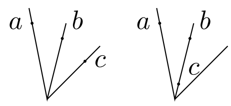
With the three definitions of non-positive and non-negative curvature introduced in Section 3.1 at hand we are going to discuss some examples. We will focus on examples which do not need any advanced techniques to be identified as Alexandrov spaces. That implies that we will not deal with those spaces that historically motivated the study of Alexandrov spaces which are convex and saddle surfaces in $\mathbb{R}^3$.

**3.2.1 Example (Euclidean space).** Euclidean space $\mathbb{R}^n$ is obviously an Alexandrov spaces with both non-positive and non-negative curvature.

**3.2.2 Example (Convex sets).** A convex set in an Alexandrov space (with the induced metric) is again an Alexandrov space with the same sign of curvature. Indeed all the shortest curves by definition remain in the convex subset, cf. Definition 2.3.3.

**3.2.3 Example (Open sets).** Open subsets of Alexandrov spaces are again Alexandrov spaces with the induced length metric. They also possess the same sign of curvature since the induced metric locally coincides with the restriction of the metric from the ambient space and all definitions of Section 3.1 are local.

**3.2.4 Example (The fan).**
The fan consisting of several segments glued together at one end is a space of non-positive curvature. One can prove this statement using the argument analogous to the one used in Example 3.1.4. However, with the angle condition 3.1.9 at our hands we can give a much shorter argument. Every triangle in the space has either all angles vanishing or is degenerate, i.e., it is entirely contained in at most two its sides, a space that is isometric to $\mathbb{R}$. In both cases the angle condition obviously holds, also see the figures on the right displaying these two possibilities.




**3.2.5 Example (A plane plus a line).** The union of the $xy$-plane and the $z$-axis in $\mathbb{R}^3$ (with the induced metric) is a space of non-positive curvature. We will discuss its generalization, i.e., a plane glued to a line at one point as a special case of a metric bouquet below. To show that this space is non-positively curved we use the angle condition 3.1.9. First note that any shortest path with its endpoints contained in the plane is entirely contained in the plane. Moreover a shortest path starting in the $z$-axis can leave it only through the origin 0. By this we see that the angle condition is trivially satisfied for all triangles which have all their vertices in the plane or all their vertices in the line. This also holds true if two vertices are contained in the line and one is in the plane. So the only remaining case is if two of vertices $a$, $b$ lie in the plane the vertex $c$ is contained in the line, see figure 3.13.

Indeed the sides emanating from $c$ have to pass through 0 so that the triangle $\triangle abc$ consist of the triangle $\triangle ab0$ and the 'tail' $[0c]$. One now sees that the angles of $\triangle abc$ are smaller that those of a corresponding comparison triangle, see figure 3.14.

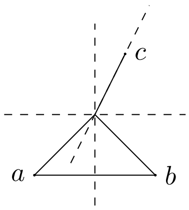


Figure 3.13

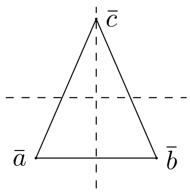


Figure 3.14

**3.2.6 Example (Planes glued at a point).** Consider several copies of $\mathbb{R}^2$ and identify their origins 0. This is again a special case of the forthcoming definition of a metric bouquet and again is a space of non-positive curvature. This can be argued essentially as in Example 3.2.5.

Now generalizing the latter three examples we introduce the following notion.

**3.2.7 Definition (Metric bouquet).** *Let $X_i$ be a family of length spaces and for all $i$ let $x_i \in X_i$. The* metric bouquet *of the spaces $X_i$ (with marked points $x_i$) is the length space obtained by gluing together all $x_i$ in the disjoint union $\dot{\bigcup} X_i$.*

**3.2.8 Remark (Metric bouquet).** It is easy to see that gluing metric spaces into a bouquet does not change their metrics. More precisely, the spaces $X_i$ are isometrically contained in the bouquet since every shortest path between points in $X_i$ never leaves it.

The latter three examples above are special cases of the following statement.

**3.2.9 Proposition (Metric bouquet and curvature).** *A metric bouquet $X$ of non-positively curved spaces is again a non-positively curved space.*

**Proof.** We use the angle condition and again the fact that every shortest path between points in some $X_i$ never leaves $X_i$.
There are three possible configurations for a triangle $\triangle abc$ in $X$. If all vertices lie in one of the $X_i$ then, since $X_i$ is non-positively curved by assumption the angle condition holds.
If all three vertices belong to different $X_i$ then all sides of $\triangle abc$ have to pass through the common point 0 and the triangle condition holds since all angles are zero. (Indeed having a nonzero angle at a point $a \in X_i$ say would contradict non-positive curvature of $X_i$: Suppose there were sides starting at $a$ and proceeding to 0 along distinct shortest paths (since the latter need not be unique) then one obtains a trivial triangle $\triangle a0b$ by choosing $b$ arbitrarily on one of the sides. The Euclidean comparison triangle will then collapse to a line and hence have a vanishing angle at $\bar{a}$.)
Finally if $a$, $b$ belong to $X_1$, say, and $c \in X_2$. Then again the angle at $c$ vanishes. Moreover the sides emanating from $c$ have to pass through the common point, making the angles at $a$ and $b$ smaller than the angles of a comparison triangle. Hence again the angle condition holds, cf. the argument in Example 3.2.5. □

**3.2.10 Example (Notebook).** The notebook from Example 1.1.20 is a non-positively curved space.
Again we consider the angle condition and similar to the above the only nontrivial configuration is if all three vertices of the triangle $\triangle abc$ lie in different half-planes, say $A$, $B$ and $C$ respectively. Indeed, if at most two half-planes are involved the condition becomes trivial since their union $A \cup B$, say, is isometric to the plane (distances are measured by 'opening the notebook').
Now denote by $L$ the common edge of the half-planes and denote by $p, q, r$ the intersection of $[ab]$, $[ac]$, $[bc]$ with $L$, see figure 3.15. Now we see that the angles $\alpha = \measuredangle bac$, $\beta = \measuredangle abc$, $\gamma = \measuredangle acb$ are

given by $\alpha = \measuredangle paq$, $\beta = \measuredangle rbp$, $\gamma = \measuredangle rcq$. Hence we construct the comparison configuration in the plane as follows, using the isometries $f_B$ of $C$ to $B$ which keeps $L$ fixed (and which amounts to turning $C$ around $L$ until it coincides with $B$) and $f_A$ of $C$ to $A$ which again keeps $L$ fixed, figure 3.16. Setting $\bar{a} = a$ and $\bar{b} = b$ we construct $\bar{c}$ as the intersection point of the circles around $\bar{a}$ with radius $[ac] = [a, f_B(c)]$ and around $\bar{b}$ with radius $[bc] = [d, f_A(c)]$. It is then obvious that we have for the angles

$$\alpha = \measuredangle paq \leq \bar{\alpha} = \measuredangle \bar{b}\bar{a}\bar{c} \quad \text{and} \quad \beta = \measuredangle rbp \leq \bar{\beta} = \measuredangle \bar{a}\bar{b}\bar{c}. \tag{3.2.1}$$

To derive the estimate on $\gamma$ we finally use the isometry $g_B$ of $A$ to $B$ keeping $L$ fixed, to construct a new comparison point $\tilde{a}$ for $a$ as the intersection of the circles centered at $f_A(c)$ and $b$ with radii $[ac]$ and $[ab]$, respectively. Then again obviously

$$\gamma = \measuredangle r f_A(c) q \leq \bar{\gamma} = \measuredangle \tilde{a} f_A(c) q. \tag{3.2.2}$$

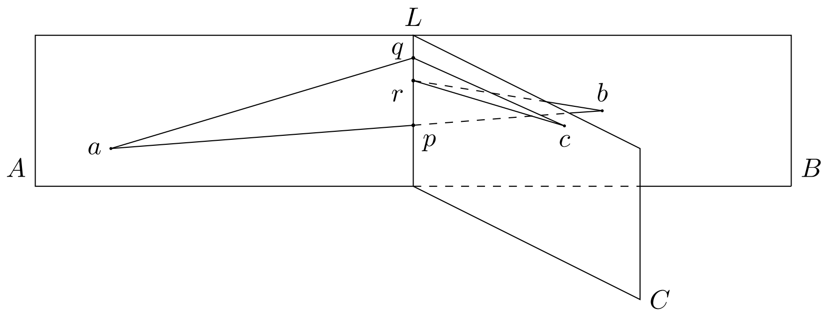



Figure 3.15

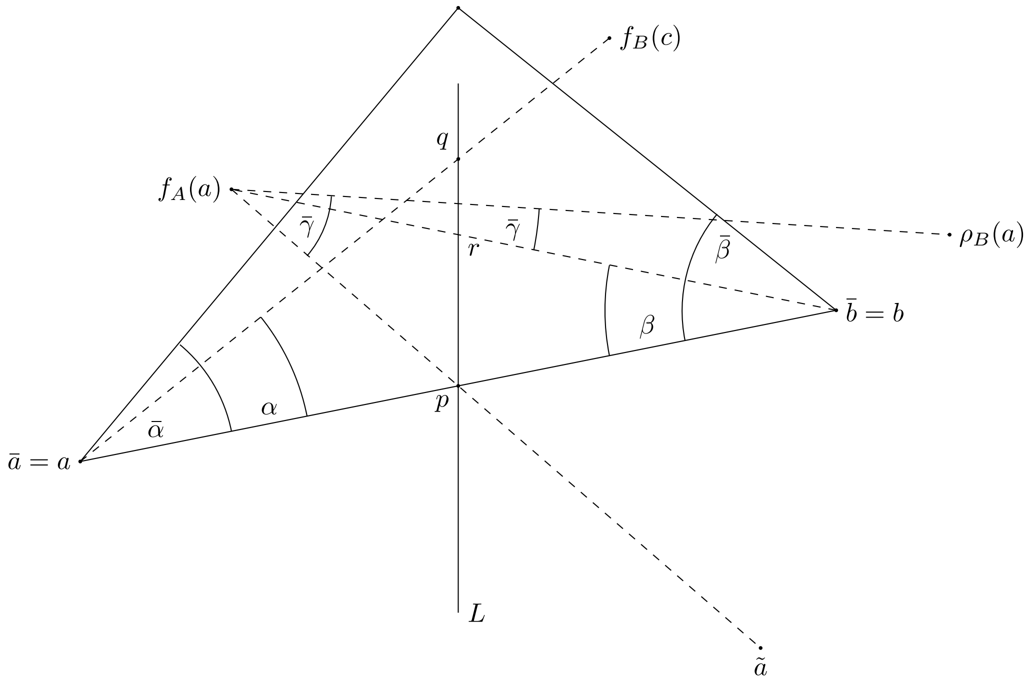



Figure 3.16

**3.2.11 Remark (Reshetnyak's theorem).** In [BBI01, Thm. 9.1.21] Reshetnyak's theorem is proved, which covers the above spaces of non-positive curvature as well as many other examples obtained by gluing.

We now turn to spaces of non-negative curvature. As a matter of fact it turns out that there are fewer distinct classes of such examples. E.g. all 2-dimensional spaces of non-negative curvature are topological manifolds, possibly with boundary. Also, all convex surfaces are non-negatively curved. Here we will prove this for polyhedral surfaces. To begin with we extract the following result from Examples 3.1.5 and 3.1.6.

**3.2.12 Lemma (Curvature of cones).** *A cone $K$ over a circle of length $L$ is non-negatively curved iff $L \leq 2\pi$, and non-positively curved iff $L \geq 2\pi$. A cone $K$ over a segment of length $L$ is non-positively curved for all $L$ and non-negatively curved iff $L \leq \pi$.*

Recall from Section 2.2 that a key-example of a 2-dimensional polyhedral space is the surface of a convex polyhedron in $\mathbb{R}^3$. We will soon see that these are in an essential way not all examples, as they are all non-negatively curved.
Observe that every point of a 2-dimensional polyhedral space has a neighbourhood which is isometric to a neighbourhood of the vertex of a cone over a graph. We call this graph the *link* of that point.

**3.2.13 Theorem (Curvature of 2-dimensional polyhedral spaces).** *A two-dimensional polyhedral space is*

*(i) non-negatively curved iff it is a topological manifold possibly with boundary such that the sum of angles around any vertex is not bigger than $2\pi$ and not bigger than $\pi$ around any vertex belonging to the boundary.*

*In other words this means that the link of every vertex is a circle of length at most $2\pi$ or a segment of length at most $\pi$.*

*(ii) non-positively curved iff the link of each vertex does not contain a subspace isometric to a circle of length less than $2\pi$.*

**Proof.** To begin with note that all curvature conditions are local in nature hence it suffices to consider small neighbourhoods of each point in the polyhedral space $X$. By the remark above the theorem we can hence restrict our attention to cones over the corresponding links. But that means that a polyhedral space $X$ is non-positively (non-negatively) curved iff for every $x \in X$ the cone over the link is non-positively (non-negatively) curved.

(i) Suppose that the link of every point is either a circle of length $L \leq 2\pi$ or a segment of length $L \leq \pi$. Then the corresponding cones are non-negatively curved by Lemma 3.2.12 and Lemma 2.3.13, respectively.

Conversely assume that $X$ is non-negatively curved. We rule out all other possible links than those mentioned above.

(a) Suppose the link of some point $x$ is not connected. Then removing $x$ makes the cone disconnected and hence $X$ locally disconnected. Now consider a triangle $\triangle abc$ with vertices $a, b$ in one component and $c$ in another one. Then the sides emanating from $c$ have to pass through $x$ and hence the angles $\alpha$, $\beta$ are smaller than the corresponding comparison angles and hence $X$ cannot be non-negatively curved, cf. the argument from 3.2.5 again.

(b) Suppose that more than two faces are adjacent to an edge. Then a small neighbourhood of a point on that edge looks like a notebook and hence $X$ cannot be non-negatively curved, cf. Example 3.2.10.

(c) By the above we know that any link is connected and every point in it has degree at most 2 (i.e., there are at most two faces joining there). So the link is either a circle or a segment. If it is a circle, then by Lemma 3.2.12 its length satisfies $L \leq 2\pi$. Moreover, if it is a segment then by Example 3.1.6 we have $L \leq \pi$.

(ii) We now pass to the case of non-positive curvature and we prove the slight reformulation of the statement:

A cone $K$ over a graph $\Gamma$ is non-positively curved iff
the length $L$ of each nontrivial loop in $\Gamma$ satisfies $L \geq 2\pi$.

First we assume that $K$ is non-positively curved and show that there is no nontrivial loop in $\Gamma$ of length $L < 2\pi$. Let $\gamma$ be a shortest nontrivial loop in $\Gamma$. Then for any pair of points $x, y \in \gamma$, at least one of the parts of $\gamma$ between $x$ and $y$ will be a shortest path in $\Gamma$. So $\gamma$ is a convex set in $\Gamma$. Hence by Theorem 2.3.17 the subcone over $\gamma$ is convex in $K$. So the subcone is non-positively curved as long as the cone over $\gamma$ is and so $L(\gamma) \geq 2\pi$, again by Lemma 3.2.12. Since $\gamma$ was assumed to be a shortest loop, all loops have length at least $2\pi$.

Conversely we now assume that all nontrivial loops in $\Gamma$ have length $L \geq 2\pi$. We show that $K$ is non-positively curved. We consider a triangle $\triangle abc \in K$ and first consider the case where none of the sides pass through the vertex 0 of $K$. Denote the projection of $\triangle abc$ to $\Gamma$ by $\triangle a'b'c'$. Then by Remark 2.3.14 the sides of $\triangle a'b'c'$ are shortest paths in $\Gamma$. We now consider two cases:

(a) Suppose that $\triangle a'b'c'$ does not contain a simple loop (i.e., a loop homeomorphic to a circle). Then the three sides of $\triangle a'b'c'$ must have a common point $d \in \Gamma$. Then the triangle $\triangle a'b'c'$ as a subset of $\Gamma$ coincides with the fan (cf. Example 3.2.4) $[a'd] \cup [b'd] \cup [c'd]$, where each one of the segments $[a'd]$, $[b'd]$, $[c'd]$ is a shortest path in $\Gamma$. So each pair of points in the fan belongs to a shortest path and hence the fan is a convex subset of $\Gamma$.

   Therefore the original triangle $\triangle abc$ is contained in the cone over the fan which is a convex subset of $K$ by 2.3.17. Also this cone consists of three sectors glued together along a ray. But now the same argument as for the notebook (Example 3.2.10) shows that the curvature is non-positive.

(b) Now suppose that $\triangle a'b'c'$ contains a simple loop which then by assumption has perimeter $L = |a'b'| + |b'c'| + |c'a'| \geq 2\pi$.

   Consider now the cone $K_L$ over a circle $S$ of length $L$ and fix a length preserving map $g : S \to \triangle a'b'c'$. This can simply be done by splitting $S$ into three arcs of length $|a'b'|$, $|b'c'|$, and $|c'a'|$, respectively and map them to the respective sides of the triangle $\triangle a'b'c'$. Now $g$ induces a map

$$\bar{g} : K_L \to K, \quad (x, t) \mapsto (g(x), t). \tag{3.2.3}$$

   Then $\bar{g}$ is an arcwise isometry, hence non-expanding. Moreover, the triangle $\triangle abc$ is the image of a triangle $\triangle a''b''c'' \in K_L$ (cf. Remark 2.3.14).

   Since $L \geq 2\pi$, $K_L$ is non-positively curved and since the triangles $\triangle abc$ in $K$ and $\triangle a''b''c'' \in K_L$ have equal side lengths they have a common comparison triangle $\triangle \bar{a}\bar{b}\bar{c}$ in $\mathbb{R}^2$. Let $d \in [bc]$ with corresponding points $d'' \in [b''c'']$ and $\bar{d} \in [\bar{b}\bar{c}]$ then we have

$$|ad| \leq |a''d''| \leq |\bar{a}\bar{d}|. \tag{3.2.4}$$

   Here the first inequality holds since $\bar{g}$ is non-expanding and the second one since $K_L$ is non-positively curved. But equation (3.2.4) shows that $K$ is non-positively curved as well.

Finally we have to deal with the case when at least one side of the triangle $\triangle abc$ passes through the vertex $O$. Suppose this is the case for $[ab]$.
If one or both of the other sides of $\triangle abc$ pass through $O$ as well, then the proof can be continued just as in Example 3.2.4. So we assume that only $[ab]$ passes through $O$ and continue the proof similar to case (b) above: The projections $[a'c']$ and $[b'c']$ of $[ac]$ and $[bc]$ to $\Gamma$ are shortest paths in $\Gamma$ and the triangle $\triangle abc$ is contained in the sub-cone $K_L$ over $[a'c'] \cup [b'c']$, which is a cone over a segment of length $L = |a'c'| + |b'c'|$ and hence non-positively curved, see Example 3.1.6. As in case (b) (cf. equation (3.2.4)) the distance condition in $\triangle abc$ reduces to one in a triangle in $K_L$ and hence $K$ again is non-positively curved. □

# 3.3 Angles in Alexandrov spaces and equivalence of definitions

The main aim of this section is to prove the equivalence of the definitions of Alexandrov spaces put forward in section 3.1. Before doing so we introduce yet another definition based on the monotonicity of angles.

## 3.3.1 Monotonicity of angles

Consider the following situation in a length space $X$ which we will generally call a *hinge*[5]. We consider two naturally parametrized shortest paths $\alpha$, $\beta$ emanating from a common point $p$ in $X$. As in section 2.4 we write $\theta(x, y) = \tilde{\angle}\alpha(x)p\beta(y)$, that is $\theta(x, y)$ is the angle at $\bar{p}$ in a comparison triangle for $\triangle\alpha(x)p\beta(y)$. The 'new' definition now is as follows.

**3.3.1 Definition (Monotonicity condition).** *A length space is called non-negatively (resp. non-positively) curved if it can be covered by neighborhoods $U$ such that for any hinge $\alpha, \beta$ in $U$ we have*

$$\theta(x, y) \quad \textit{is nonincreasing (resp. nondecreasing)} \tag{3.3.1}$$

*in each variable $x$ and $y$ with the other one fixed.*

Now since in a hinge $\theta$ is given by an arccos-function (cf. equation (2.4.2)) and hence bounded the monotonicity condition immediately implies the following fact.

**3.3.2 Proposition (Angles in Alexandrov spaces exist).** *If $X$ is an Alexandrov space in the sense of Definition 3.3.1, then the angle between any pair of shortest paths in $X$ exists.*

## 3.3.2 Equivalence of definitions

To begin with we need to discuss an elementary fact from Euclidean geometry which generally goes under the name *Alexandrov's Lemma*. It has a vivid description, which we give prior to the formal statement.

Consider a plane quadrangle in the plane. Denote two opposite diagonal points by $b$ and $d$ and assume all angles except at $d$ be less than $\pi$, see figure 3.17. Regarding the two sides emanating from $b$ as a hinge. Now we open up this hinge by straightening the hinge formed by the sides emanating from $d$ (if possible) resulting in new points $a'$ and $c'$, at the same distances from $d$ as $a$ and $c$ respectively. If the angle at $d$ was less than $\pi$ in the beginning (i.e, the quadrangle was convex, left), then this procedure causes $b$ to move nearer to $d$. Otherwise (right) $b$ moves farther away from $d$. Formally we have.

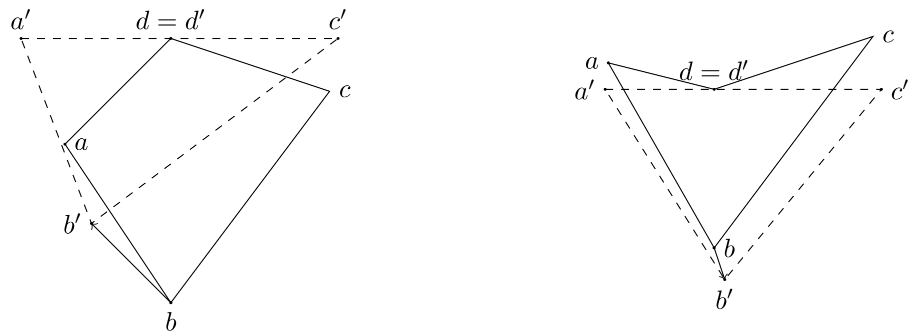



Figure 3.17

[5] Scharnier in German.

**3.3.3 Lemma (Alexandrov).** *Let $a, b, c, d$ be points in the plane $\mathbb{R}^2$ such that $a$ and $c$ lie in different half-planes with respect to the line $[bd]$. Consider the triangle $\triangle a'b'c' \in \mathbb{R}^2$ such that*

$$|ab| = |a'b'|, \quad |bc| = |b'c'|, \quad |ad| + |dc| = |a'c'| \tag{3.3.2}$$

*and let $d'$ be the point in $[a'c']$ with $|ad| = |a'd'|$.*
*Then*

$$\measuredangle adb + \measuredangle bdc < \pi \quad \text{iff} \quad |b'd'| < |bd| \tag{3.3.3}$$

*and in this case one also has $\measuredangle b'a'd' < \measuredangle bad$ and $\measuredangle b'c'd' < \measuredangle bcd$, see figure 3.17, left.*
*Also*

$$\measuredangle adb + \measuredangle bdc > \pi \quad \text{iff} \quad |b'd'| > |bd| \tag{3.3.4}$$

*and in this case one also has $\measuredangle b'a'd' > \measuredangle bad$ and $\measuredangle b'c'd' > \measuredangle bcd$, see figure 3.17, right.*

In the proof we shall use the following elementary fact of planar geometry: If the length of two sides of a planar triangle are kept fixed, then the angle between them is a monotonously increasing function of the length of the third side. More explicitly, if $|xy| = |x'y'|$ and $|yz| = |y'z'|$ for two triangles $\triangle ayz$ and $\triangle x'y'z'$ in the plane then

$$\measuredangle xyz > \measuredangle x'y'z' \quad \text{iff} \quad |xz| > |x'z'|. \tag{3.3.5}$$

**Proof.** We choose a point $c_1$ on the ray $ad$ in such a way that $d$ lies between $a$ and $c_1$ and such that $|dc| = |dc_1|$, see figure 3.18.

Suppose now that $\measuredangle adb + \measuredangle bdc > \pi$, but this implies $\measuredangle bdc_1 < \measuredangle bdc$ and hence by (3.3.5) $|bc_1| < |bc| = |b'c'|$. Now observing for the triangles $\triangle abc_1$ and $\triangle a'b'c'$ (notice that $|ab| = |a'b'|$ and $|ac_1| = |a'c'|$) we obtain $\measuredangle bac_1 < \measuredangle b'a'c'$. This implies again applying (3.3.5) now to the triangles $\triangle bad$ and $\triangle b'a'd'$ that $|bd| < |b'd'|$.
The rest of the lemma follows along the same lines.

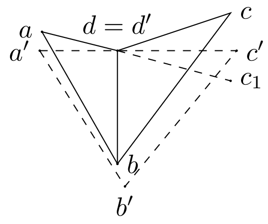



Figure 3.18

□

**3.3.4 Remark (Alexandrov's lemma in model geometries).** Alexandrov's lemma retains its validity if the triangles are placed on a sphere or a hyperbolic plane. In fact it holds true on any Riemannian manifold if the triangles are small enough. One then just works in a normal neighbourhood and the proof has to be repeated in a vector space with a (positive definite) scalar product.

Now we come to the long announced statement on the equivalence of all definitions of Alexandrov spaces.

**3.3.5 Theorem (Equivalence of definitions of Alexandrov spaces).** *All the definitions of Alexandrov spaces, that is the distance condition 3.1.3, the triangle condition 3.1.8, the angle condition 3.1.9, and the monotonicity condition 3.3.1 are equivalent.*

Observe the following technical subtlety: All definitions are local but the size of the neighbourhoods might depend on the condition at hand. E.g. the validity of one condition on a neighbourhood $U$ might imply some of the other conditions to hold only on some smaller neighbourhood $V$. We will nevertheless refer to all such neighbourhoods as *normal regions*.

**Proof.** The proofs in the cases of non-negative and non-positive curvature are very similar; so we only proof the *non-positive* case in detail and comment on the necessary modifications in the non-negative case.

(1) The distance condition 3.1.3 and the triangle condition 3.1.8 are a word by word translation of each other, hence equivalent as discussed below definition 3.1.8.

(2) We show that the triangle condition 3.1.8 implies the monotonicity condition 3.3.1: Consider a hinge of the shortest paths $\alpha = [pa]$ and $\beta = [pb]$ and a point $a_1$ in $\alpha$ beyond $a$. Now consider the comparison triangles $\triangle\bar{p}\bar{a}\bar{b}$ and $\triangle\bar{p}\bar{a}_1\bar{b}$ for the triangles $\triangle apb$ and $\triangle pa_1b$, see figure 3.19. Let $\tilde{a}$ be a point on $[\bar{p}\bar{a}_1]$ as far away from $\bar{p}$ as $a$ is from $p$, i.e., $|\bar{p}\tilde{a}| = |pa|$. Then the triangle condition implies that $|\tilde{a}\bar{b}| \geq |ab| = |\bar{a}\bar{b}|$ and so by (3.3.5) $\measuredangle\bar{a}_1\bar{p}\bar{b} \geq \measuredangle\bar{a}\bar{p}\bar{b}$, which is the monotonicity condition.

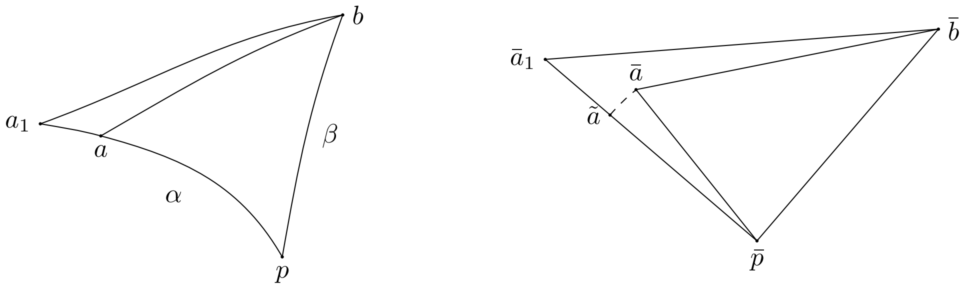



Figure 3.19

(3) The monotonicity condition 3.3.1 implies the angle condition 3.1.9: Consider a triangle $\triangle abc$ and call its sides $\alpha = [ba]$ and $\beta = [bc]$, i.e., $\alpha$, $\beta$ is a hinge at $\alpha(0) = \beta(0) = b$. By monotonicity of angles we have

$$\measuredangle abc = \measuredangle\alpha,\beta = \lim_{s,t\to 0} \theta(s,t) \leq \theta(|ba|,|bc|). \tag{3.3.6}$$

But $\theta(|ba|,|bc|) = \measuredangle\bar{a}\bar{b}\bar{c}$, hence the angle condition holds.

(4) We finally show that the angle condition 3.1.9 implies the triangle condition 3.1.8: Consider a triangle $\triangle abc$ and a point $d \in [ac]$. By the triangle inequality for angles, Theorem 2.4.7 and by Proposition 2.4.5(ii) we have (see figure 3.20, left)

$$\measuredangle adb + \measuredangle bdc \geq \measuredangle adc = \pi. \tag{3.3.7}$$

Now place comparison triangles $\triangle\bar{a}\bar{d}\bar{b}$ and $\triangle\bar{b}\bar{d}\bar{c}$ on opposite sides of the line $[\bar{b}\bar{d}]$ in the plane. Then the angle condition together with (3.3.7) implies (cf. figure 3.20, right)

$$\measuredangle\bar{a}\bar{d}\bar{b} + \measuredangle\bar{b}\bar{d}\bar{c} \geq \measuredangle adb + \measuredangle bdc \geq \pi. \tag{3.3.8}$$

Note that we have chosen the points $\bar{a}$, $\bar{b}$, $\bar{c}$ and $\bar{d}$ such that $|ac| = |cd| + |da| = |\bar{c}\bar{d}| + |\bar{d}\bar{a}|$ Setting $b' = \bar{b}$ we choose points $a', c'$ such that $|a'c'| = |ac|$ and $d'$ on the segment $|a'c'|$ such that $|a'd'| = |ad|$. Note that $\triangle a'b'c'$ is a comparison triangle for $\triangle abc$ and that by Alexandrov's lemma 3.3.3 we have

$$|bd| = |\bar{b}\bar{d}| \leq |\bar{b}\bar{d}'|, \tag{3.3.9}$$

but this is the triangle condition for $\triangle abc$ and $d \in [ac]$.

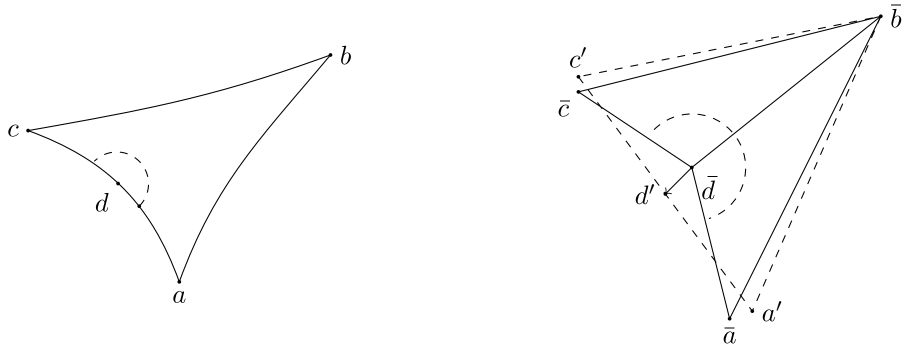



Figure 3.20

This proves the equivalence of all definitions in the case non-positive curvature.

For the case of *non-negatively* curved spaces again (1) is without problems. In the other parts of the proof one has to reverse all inequalities. This is without problem in all inequalities that directly come from the definitions and so item (2) translates accordingly. On the other hand, one sees that item (4) does not work so easily. Indeed in (3.3.7) we have used the triangle inequality for angles to show that the sum of adjacent angles is bounded below by $\pi$. Although this argument holds true in *any* length space it obviously does not work to establish the reversed bound now needed and this is why we have included it (i.e., the condition that the sum of adjacent angles is bounded above by $\pi$) in the definition (cf. 3.1.9). Hence we can indeed infer the triangle condition from the angle condition in analogy to item (4). But now the burden is shifted to (3) where one infers the angle condition from the monotonicity condition, which now includes showing the statement about the sum of adjacent angles. This step is, however, established by the following lemma. □

Besides being used in the above proof the next lemma has also numerous further applications.

**3.3.6 Lemma (Adjacent angles in non-negative curvature).** *If $X$ is non-negatively curved in the sense of the monotonicity condition 3.3.1, then for each shortest path the sum of adjacent angles equals $\pi$. More precisely, if $r$ is an inner point of the shortest path $[pq]$ and $[rs]$ is also a shortest path then (see figure 3.21, left)*

$$\measuredangle prs + \measuredangle srq = \pi. \tag{3.3.10}$$

**Proof.** As already noted in the last paragraph of the proof of 3.3.5 we have by 2.4.7 and 2.4.5(ii) that $\measuredangle prs + \measuredangle srq \geq \measuredangle prq = \pi$ and it only remains to infer the reverse inequality from the monotonicity condition in case of non-negative curvature. To this end consider points $p_0$, $q_0$, and $s_0$ in the shortest paths $[pr]$, $[rq]$, and $[rs]$. Place comparison triangles $\triangle\bar{p}_0\bar{r}\bar{s}_0$ and $\triangle\bar{s}_0\bar{r}\bar{q}_0$ for the triangles $\triangle p_0rs_0$ and $\triangle s_0rq_0$ on different sides of the line $\bar{r}\bar{s}_0$ in the plane, see figure 3.21. Moreover, let $\triangle\bar{s}_1\bar{p}_1\bar{q}_1$ be a comparison triangle for $\triangle s_0p_0q_0$ and choose $\bar{p}_1 = \bar{p}_0$ and $\bar{q}_1$ on the straight line from $\bar{p}_1 = \bar{p}_0$ through $\bar{r}$, see figure 3.21, right. Since $q_0$ is farther away from $p_0$ than $r$ is, the monotonicity condition yields $\measuredangle\bar{s}_0\bar{p}_0\bar{r} \geq \measuredangle\bar{s}_1\bar{p}_1\bar{q}_1$ and so $|\bar{r}\bar{s}_1| \leq |\bar{r}\bar{s}_0|$, by (3.3.5). Consequently the first part of Alexandrov's lemma 3.3.3 implies that $\measuredangle\bar{p}_0\bar{r}\bar{s}_0 + \measuredangle\bar{s}_0\bar{r}\bar{q}_0 \leq \pi$. Finally passing to the limit as $p_0$, $q_0$, $s_0$ approach $r$ we obtain $\measuredangle prs + \measuredangle srq \leq \pi$. □

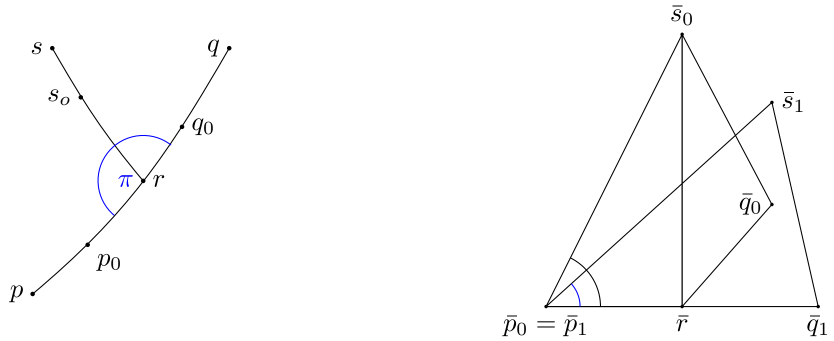



Figure 3.21

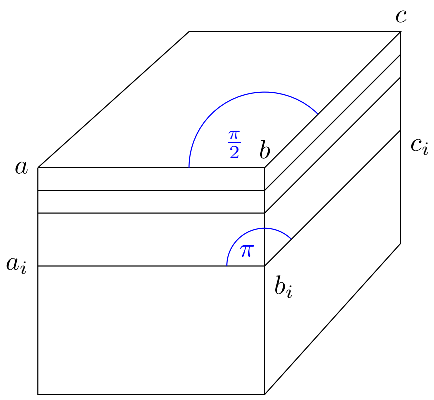
### 3.3.3 Semi-continuity of angles

We start this final subsection on angles in Alexandrov space with an observation: In the plane the angles of a triangle depend continuously on its vertices. This is, however, no longer true in length spaces. But in Alexandrov spaces semi-continuity is retained, more precisely in non-positively (non-negatively) curved spaces the angle is upper (lower) semi-continuous. We begin with some examples but first fix our notations.

Suppose the sequences of shortest paths $[a_ib_i]$ and $[a_ic_i]$ converge uniformly to shortest paths $[ab]$ and $[ac]$, respectively (and assume all shortest paths are naturally parametrized). Consequently $a_i \to a$, $b_i \to b$, and $c_i \to c$. Although uniform convergence is quite strong, nevertheless in general length spaces there is no relation between the angle $\measuredangle bac$ and the limit $\lim \measuredangle b_i a_i c_i$, even if the latter exists.

**3.3.7 Example (Collapsing angle).** Consider the surface of a cube in $\mathbb{R}^3$ with the induced length metric. Let $[ab]$ and $[bc]$ be edges of the cube and consider segments $[a_i b_i]$ and $[b_i c_i]$ parallel to $[ab]$ and $[bc]$ respectively at distances $1/i$, see figure 3.22, left. Then the hinges $([a_i b_i], [b_i c_i])$ converge to the hinge $([ab], [bc])$ but $\measuredangle abc = \pi/2$ while $\measuredangle a_i b_i c_i = \pi$ for all $i$ by 2.4.5 since $b_i$ lies on a shortest path, cf. 1.4.21.

**3.3.8 Example (Exploding angle).** Consider $\mathbb{R}^2$ with the open first quadrant $\{(x, y) : x > 0, y > 0\}$ removed and the length structure induced by the one of $\mathbb{R}^2$. Let $a = (0, 1)$, $a_i = (-1/i, 1)$, $b = (0, 0)$, $b_i = (-1/i, -1/i)$, and $c_i = (0, 1)$, $c_i = (-1/i, 1)$, see figure 3.22, right. Then the hinges $([a_i b_i], [b_i c_i])$ converge to the hinge $([a, b], [bc])$. But now $\measuredangle abc = \pi$ since $b$ is on a shortest path while $\measuredangle a_i b_i c_i = \pi/2$ for all $i$.



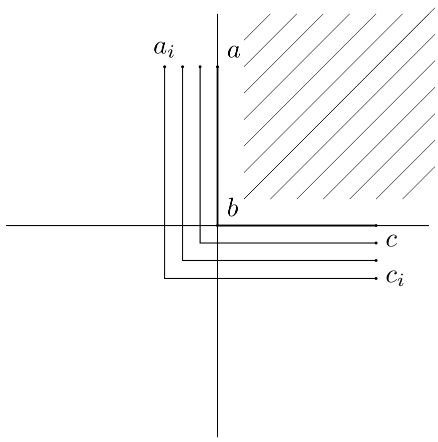




Figure 3.22

Observe that the spaces in Examples 3.3.7 and 3.3.8 are non-negatively and non-positively curved, respectively. This tells us that the following statement is optimal.

**3.3.9 Theorem (Semi-continuity of angles).** *Let $X$ be a space of non-positive (non-negative) curvature and suppose that the sequences of shortest paths $([a_i b_i])_i$ and $([b_i c_i])_i$ converge uniformly to the shortest paths $[ab]$ and $[bc]$ respectively. Then*

$$\measuredangle abc \geq \limsup_{i\to\infty} \measuredangle a_i b_i c_i \quad (\textit{resp. } \measuredangle abc \leq \liminf_{i\to\infty} \measuredangle a_i b_i c_i). \tag{3.3.11}$$

**Proof.**

Write $\alpha_i$ and $\alpha$ for the angles $\measuredangle a_i b_i c_i$ and $\measuredangle abc$, respectively. Let $s > 0$ be small and denote by $a_i'$ and $a'$ points on $[a_i b_i]$ and $[ab]$ at a distance $s$ from $b_i$ and $b$ respectively. Similarly denote by $c_i'$ and $c_i$ points on $[b_i c_i]$ and $[bc]$ at a distance $t$ from $b_i$ and $b$ respective, see figure 3.23. Now write $\theta_i(s,t)$ and $\theta(s,t)$ for the respective comparison angles $\tilde{\measuredangle} a_i' b_i c_i'$ and $\tilde{\measuredangle} a' b c'$. Observe that $|a_i' b_i| = |a' b|$, $|c_i' b_i| = |c' b|$ and that for fixed $s, t$ also $|a_i' c_i'| \to |a' c'|$.

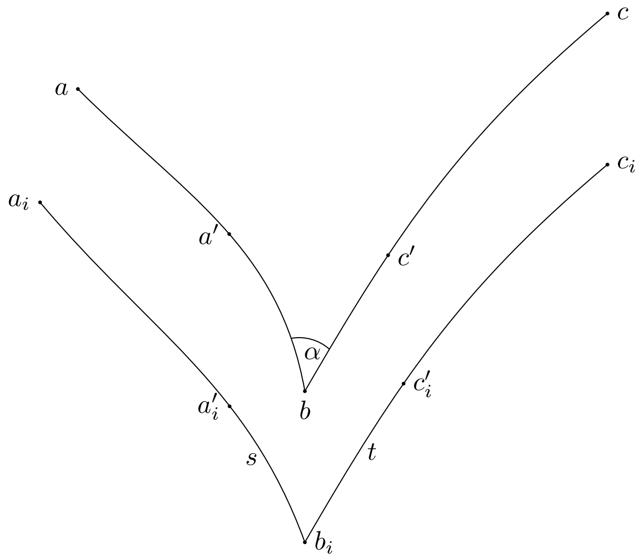



Figure 3.23

Hence we have that

$$\theta_i(s,t) \to \theta(s,t). \tag{3.3.12}$$

By definition of the angle (cf. 2.4.2) we have that $\alpha_i = \lim_{s,t\to 0} \theta_i(s,t)$and $\alpha = \lim_{s,t\to 0} \theta(s,t)$. Now if $X$ is non-positively curved then by the monotonicity condition 3.3.1 $\theta_i$ and $\theta$ are non-decreasing and so $\theta_i(s,t) \geq \alpha_i$ for all $s, t$. So by (3.3.12)

$$\theta(s,t) = \lim_i \theta_i(s,t) \geq \limsup_i \alpha_i \quad \text{and hence} \quad \alpha = \lim_{s,t\to 0} \theta(s,t) \geq \limsup_i \alpha_i. \tag{3.3.13}$$

The case of non-negatively curved spaces is completely analogous. □

## 3.4 The first variation formula

In this section we want to find formulae for the first variation of the distance between a fixed point and a point moving along a shortest path in an Alexandrov space. The term 'first variation' here refers to the derivative of this distance. To begin with, we look at a simple special case:

**3.4.1 Example.** Let $\gamma : [0, a] \to \mathbb{R}^2$ be a smooth curve of unit speed and let $p$ be a point that does not belong to $\gamma$. Denote the distance from $p$ to $\gamma(t)$ by

$$l(t) := |p\gamma(t)| = \sqrt{\langle p - \gamma(t), p - \gamma(t)\rangle}.$$

Then $l'(t) = -\frac{1}{|p-\gamma(t)|}\langle p - \gamma(t), \gamma'(t)\rangle$, i.e.,

$$\frac{dl}{dt} = -\cos \measuredangle(p - \gamma(t), \gamma'(t)). \tag{3.4.1}$$

We will show that an analogous formula holds in any space of non-positive or non-negative curvature. Let us first fix some notation. By $X$ we will denote a length space, and by $\gamma : [0, T] \to X$ a unit-speed shortest path. We set $a = \gamma(0)$, $q = \gamma(T)$, and we let $p \in X \setminus \{a\}$. For $t \in [0, T]$, set $l(t) := |p\gamma(t)|$, and we let $\sigma_t$ be a shortest path from $\gamma(t)$ to $p$ (assuming such a path exists).
Our first goal is to establish 'one half' of the desired result, namely an inequality that holds in any length space, regardless of curvature bounds.

**3.4.2 Proposition.** *Let $(X, d)$ be a length space. If the angle $\alpha = \measuredangle paq$ between the shortest paths $\gamma$ and $[ap] = \sigma_0$ exists, then*

$$\limsup_{t \to 0+} \frac{l(t) - l(0)}{t} \leq -\cos\alpha. \tag{3.4.2}$$

*Equivalently, $l(t) \leq l(0) - t\cos\alpha + o(t)$ as $t \to 0+$.*

For the proof of Proposition 3.4.2 we need some preparations.

**3.4.3 Remark.** The left hand side of (3.4.2) depends only on the length of $\sigma_0$ and not on the actual choice of a shortest path $\sigma_0$ from $a$ to $p$ itself, the claim we make in Proposition 3.4.1 can be rewritten as

$$\limsup_{t \to 0+} \frac{l(t) - l(0)}{t} \leq \inf_{\sigma_0}(-\cos\alpha),$$

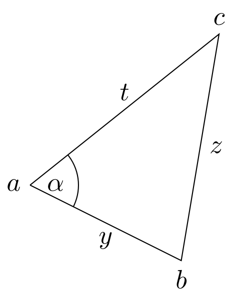
where the infimum is over all such shortest paths $\sigma_0$, or also as

$$\limsup_{t \to 0+} \frac{l(t) - l(0)}{t} \leq -\cos\alpha_{\min},$$

with $\alpha_{\min}$ the infimum over all angles between $\gamma$ and such curves $\sigma_0$.

**3.4.4 Lemma.** *Let $\Delta abc$ be a triangle in $\mathbb{R}^2$, $\alpha = \measuredangle bac$, and $t = |ac|$. Then*

$$\left|\cos\alpha - \frac{|ab| - |bc|}{t}\right| \leq \frac{t}{|ab|}.$$




**Proof.** Set $y := |ab|$ and $z := |bc|$. The law of cosines can be written as

$$\cos\alpha = \frac{t^2 + y^2 - z^2}{2ty} = \frac{y - z}{t}\frac{y + z}{2y} + \frac{t}{2y}.$$

By the triangle inequality, $\left|\frac{y-z}{t}\right| \leq 1$ and $\left|\frac{y+z}{2y} - 1\right| \leq \frac{t}{2y}$, so

$$\left|\cos\alpha - \frac{y - z}{t}\right| = \left|\frac{y - z}{t}\frac{y + z}{2y} + \frac{t}{2y} - \frac{y - z}{t}\right| \leq \left|\frac{y - z}{t}\right| \cdot \left|\frac{y + z}{2y} - 1\right| + \frac{t}{2y} \leq \frac{t}{2y} + \frac{t}{2y} = \frac{t}{y}.$$

□

**Proof of Proposition 3.4.2.** We consider two variable points: a point $b$ on the shortest path $\sigma_0 = [ap]$ and a point $c = \gamma(t)$. (see figure 3.24). Then by the triangle inequality

$$|ab| - |bc| \leq |ap| - (|bp| + |bc|) \leq l(0) - l(t).$$

Now we apply Lemma 3.4.4 to the comparison triangle for $\Delta abc$ to obtain

$$\cos\tilde{\measuredangle} bac \leq \frac{|ab| - |bc|}{t} + \frac{t}{|ab|} \leq -\frac{l(t) - l(0)}{t} + \frac{t}{|ab|}.$$

Here, $t = |ac|$ because $\gamma$ has unit speed. Since the angle $\measuredangle paq$ exists by assumption, we can in particular let $b$ and $c$ converge to $a$ in this inequality in such a way that $t/|ab| \to 0$ to have the left hand side converge to $\cos\alpha$. This gives the claim.

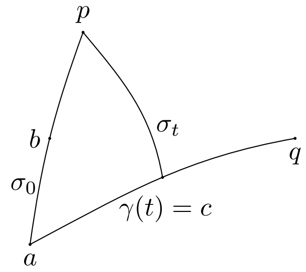



Figure 3.24

□

**3.4.5 Theorem (First variation theorem).** *Let $X$ be a length space of non-positive or non-negative curvature and let $\gamma, \sigma_t$ and $l(t)$ as above. Suppose that some sequence $\sigma_{t_i}$ converges uniformly to $\sigma_0$, where $t_i \to 0$ as $i \to \infty$. Then[6] we have*

$$\lim_{i\to\infty} \frac{l(t_i) - l(0)}{t_i} = -\cos\alpha. \tag{3.4.3}$$

**Proof.** Due to Proposition 3.4.2 it only remains to prove that

$$\liminf_{i\to\infty} \frac{l(t_i) - l(0)}{t_i} \geq -\cos\alpha.$$

Pick $r > 0$ so small that $|ap| > 5r$ and such that $B_{5r}(a)$ is a normal region for non-positive resp. non-negative curvature — this ball will contain all the triangles used in the constructions below. Also, we may assume that $\gamma(t_i) \in B_r(a)$ for all $i \in \mathbb{N}$. Let $c_i = \gamma(t_i)$ and let $b_i$ be the point on the shortest path $\sigma_{t_i} = [c_i p]$ such that $|b_i c_i| = r$ (see figure 3.25, left). We will show that

$$\limsup_{i\to\infty} \tilde{\measuredangle} a c_i b_i \leq \pi - \alpha. \tag{3.4.4}$$

In fact, suppose for the moment that (3.4.4) has already been shown. Then applying Lemma 3.4.4 to a comparison triangle for $\Delta b_i c_i a$ (setting $b = b_i$, $a = c_i$ and $c = a$ in that Lemma) gives

$$l(0) = |pa| \leq |pb_i| + |b_i a| \leq |pb_i| + |b_i c_i| - t_i \cos\tilde{\measuredangle} b_i c_i a + \frac{t_i^2}{|b_i c_i|}.$$

Now $|pb_i| + |b_i c_i| = l(t_i)$, so

$$\frac{l(t_i) - l(0)}{t_i} \geq \cos\tilde{\measuredangle} b_i c_i a - \frac{t_i}{|b_i c_i|} = \cos\tilde{\measuredangle} b_i c_i a - \frac{t_i}{r}.$$

Combining this with (3.4.4) we arrive at

$$\liminf_{i\to\infty} \frac{l(t_i) - l(0)}{t_i} \geq \liminf_{i\to\infty} (\cos\tilde{\measuredangle} b_i c_i a) \geq \cos(\pi - \alpha) = -\cos\alpha,$$

from which the Theorem follows.
It remains to show (3.4.4), which we do separately for each curvature assumption.
Suppose first that $X$ has non-negative curvature. Then by the angle condition from Definition 3.1.9 and Lemma 3.3.6 we get

$$\tilde{\measuredangle} b_i c_i a \leq \measuredangle b_i c_i a = \pi - \measuredangle b_i c_i q.$$

From this, (3.4.4) follows because $\liminf_{i\to\infty} \measuredangle b_i c_i q \geq \alpha$ by semicontinuity of angles, see Theorem 3.3.9 (it is here that we need the uniform convergence of $(\sigma_{t_i})$).

[6] since $X$ is a non-positively resp. non-negatively curved space, the angle $\alpha$ between $\sigma_0$ and $\gamma$ exists

Finally, let $X$ be a space of non-positive curvature. Let $b$ be the point on $\sigma_0 = [ap]$ such that $|ab| = r$ (see figure 3.25, right). Then by Definition 3.1.9, $\measuredangle bab_i \le \tilde{\measuredangle} bab_i$, and $\tilde{\measuredangle} bab_i \to 0$ as $i \to \infty$ because $|b_i b| \to 0$ and both $|ab|$ and $|ab_i|$ stay bounded away from 0. Consequently, $\measuredangle bab_i \to 0$ as $i \to \infty$. Combining this with the triangle inequality for angles (Theorem 2.4.7) we obtain

$$\measuredangle c_i ab_i \le \alpha \le \measuredangle c_i ab_i + \measuredangle bab_i \Rightarrow 0 \le \alpha - \measuredangle c_i ab_i \le \measuredangle bab_i \to 0 \quad (i \to \infty).$$

From this, the angle condition entails

$$\liminf_{i\to\infty} \tilde{\measuredangle} c_i ab_i \ge \liminf_{i\to\infty} \measuredangle c_i ab_i = \alpha. \tag{3.4.5}$$

Furthermore, since $\tilde{\measuredangle} c_i ab_i + \tilde{\measuredangle} ac_i b_i + \tilde{\measuredangle} ab_i c_i = \pi$ and $\tilde{\measuredangle} ab_i c_i \to 0$, it follows that $\tilde{\measuredangle} c_i ab_i + \tilde{\measuredangle} ac_i b_i \to \pi$ as $i \to \infty$. Therefore, using (3.4.5) we finally arrive at

$$\limsup_{i\to\infty} \tilde{\measuredangle} ac_i b_i = \pi - \liminf_{i\to\infty} \tilde{\measuredangle} c_i ab_i \le \pi - \alpha.$$

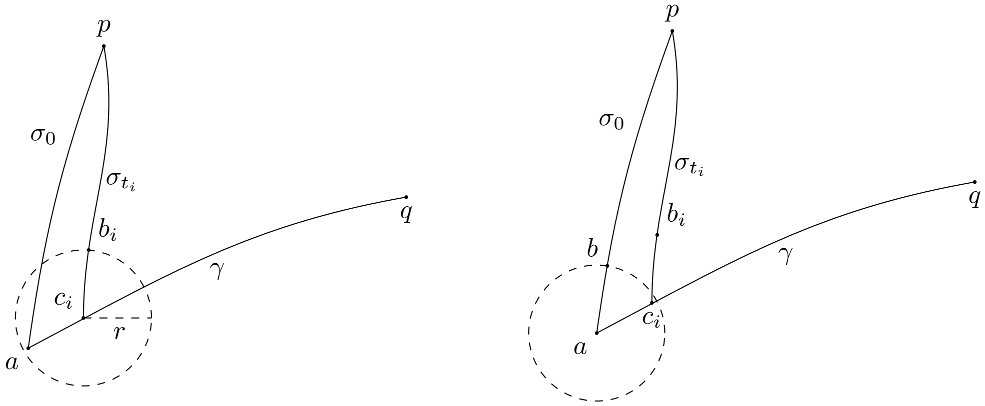



Figure 3.25

□

It immediately follows that if $\{\sigma_t : t \in [0,T]\}$ is a continuous family of shortest paths from $p$ to $\gamma(t)$, then the right derivative of $l$ at $t = 0$ exists and equals $-\cos\alpha$. In fact, one does not even need to assume uniqueness of shortest paths. More precisely, we have:

**3.4.6 Corollary.** *Let $X$ be a non-positively or non-negatively curved complete, locally compact length space. Let $\gamma : [0,T] \to X$ be a geodesic parametrized by arclength, $p \in X$, $p \ne \gamma(0)$. Then the function $t \mapsto l(t) = |p\gamma(t)|$ possesses the right derivative*

$$\lim_{t\to 0+} \frac{l(t) - l(0)}{t} = -\cos\alpha_{\min},$$

*where $\alpha_{\min}$ is the infimum (and in fact minimum) of angles between $\gamma$ and shortest paths connecting $\gamma(0)$ to $p$.*

**Proof.** From Proposition 3.4.2 and Remark 3.4.3 we know

$$\limsup_{t\to 0+} \frac{l(t) - l(0)}{t} \le -\cos\alpha_{\min}. \tag{3.4.6}$$

Now pick $t_i \to 0+$ with

$$\frac{l(t_i) - l(0)}{t_i} \to \liminf_{t\to 0+} \frac{l(t) - l(0)}{t} \quad (i \to \infty)$$

and pick shortest paths $\sigma_{t_i}$ from $p$ to $\gamma(t_i)$. By Theorem 1.4.10(Arzela-Ascoli) (together with Proposition 1.4.16 and the proof of Corollary 1.4.11), $(\sigma_{t_i})$ contains a subsequence that uniformly converges to a path $\sigma_0$. Moreover, $\sigma_0$ is a shortest path by Proposition 1.4.13. Noting that the angle between shortest paths always exists in Alexandrov spaces (cf. 3.3.2), we can apply Theorem 3.4.5 to conclude

$$\lim_{i\to\infty} \frac{l(t_i)-l(0)}{t_i} = -\cos\alpha,$$

where $\alpha$ is the angle between $\gamma$ and $\sigma_0$. Thus

$$\liminf_{t\to 0+} \frac{l(t)-l(0)}{t} = -\cos\alpha \geq -\cos\alpha_{\min}.$$

Finally, combining this with (3.4.6) we get $-\cos\alpha_{\min} \leq -\cos\alpha \leq -\cos\alpha_{\min}$, implying that $\alpha_{\min} = \alpha$, so the infimum is indeed a minimum. □

## 3.5 Non-zero curvature bounds and globalization

So far we have have only looked at spaces of non-negative or non-positive curvature, via comparison with $\mathbb{R}^2$. In the present section we also introduce non-zero curvature bounds. The idea is again to use comparison with spaces of constant, but this time non-zero, curvature. In fact, it will suffice to consider spaces of constant curvature $+1$ or $-1$, since any other positive or negative curvature can be reduced to these cases via scaling.
The comparison spaces of constant curvature are the so-called *Riemannian model spaces* that will be studied in detail in Section 4.2. For the moment, we will take some basic properties of these spaces for granted and only refer to Chapter 4 for proofs of these properties. In order to have a common notation for all occurring cases, we introduce the following (cf. also Definition 4.2.16):

**3.5.1 Definition.** *Let $k \in \mathbb{R}$. The $k$-plane is one of the following spaces:*

- $\mathbb{R}^2$ *if* $k = 0$.
- *The Euclidean sphere of radius* $1/\sqrt{k}$ *if* $k > 0$.
- *The hyperbolic space of radius* $1/\sqrt{-k}$ *if* $k < 0$.

The $k$-plane is bounded (i.e., has finite diameter) for $k > 0$ and unbounded for $k \leq 0$. We denote the diameter of the $k$-plane by $D_k$, so

$$D_k = \begin{cases} \frac{\pi}{\sqrt{k}}, & k > 0 \\ \infty & k \leq 0. \end{cases} \tag{3.5.1}$$

As we shall prove in Proposition 4.2.20, if $a, b, c > 0$ satisfy $a + b + c < 2D_k$ then there exists a triangle in the $k$-plane with sides of lengths $a, b, c$, which moreover is unique up to isometry. Consequently, for every sufficiently small (if $k > 0$) triangle in a length space there exists a unique (up to rigid motion) comparison triangle in the $k$-plane.
Now that we have fixed suitable comparison spaces, we can introduce general curvature bounds in complete analogy to the definition of non-positive and non-negative curvature, i.e., either via distance functions, triangles, angles, or monotonicity. We explicitly write out the triangle definition:

**3.5.2 Definition.** *Let $k \in \mathbb{R}$. A strictly intrinsic connected length space $X$ is a space of curvature $\geq k$ (resp. $\leq k$) if every point $x \in X$ has a neighborhood $U$ such that for any triangle $\Delta abc$ contained in $U$ and any point $d \in [ac]$ the inequality $|bd| \geq |\bar{b}\bar{d}|$ (respectively $|bd| \leq |\bar{b}\bar{d}|$) holds, where $\Delta\bar{a}\bar{b}\bar{c}$ is a comparison triangle in the $k$-plane and $\bar{d} \in [\bar{a}\bar{c}]$ is the point with $|\bar{a}\bar{d}| = |ad|$.*

The other definitions can be transferred analogously.

**3.5.3 Remark (Equivalence of definitions for curvature bounds).** It is important to note that theorem 3.3.5, stating the equivalence of all three definitions of curvature bounds, also holds for $k \neq 0$. Indeed, the proof of that result relied only on the Alexandrov Lemma 3.3.3. As we shall see in Remark 4.2.21, this Lemma remains valid for general $k$, so the claim follows.

There are also Alexandrov spaces with variable curvature bounds (i.e., where $k$ may depend on the point):

**3.5.4 Definition.** *A strictly intrinsic length space $X$ is a space of curvature bounded above (resp. below) if every point $x \in X$ has a neighborhood that is a space of curvature $\leq k$ (resp. $\geq k$) for some $k \in \mathbb{R}$.*

Another far-reaching concept is that of Alexandrov spaces in which triangle comparison works 'in the large', i.e., for all triangles regardless of size:

**3.5.5 Definition (arg1).** *A strictly intrinsic connected length space $X$ is said to satisfy a curvature bound ($\geq k$ or $\leq k$)* globally*, if for some $k \in \mathbb{R}$ the triangle condition from definition 3.5.2 is satisfied for all triangles $\Delta abc$ in $X$ for which there exists a comparison triangle in the $k$-plane and is unique up to rigid motion.*

In the case $k > 0$ this definition requires a bit of care. In fact, in this case the $k$-plane is a sphere of radius $\pi/\sqrt{k}$, so there do not exist any comparison triangles whose perimeter exceeds $2D_k = 2\pi/\sqrt{k}$. If the perimeter equals $2\pi/\sqrt{k}$, then there are two cases:

(i) All sides are shorter than $\pi/\sqrt{k}$. In this case, the comparison triangle is unique, namely it equals a great circle with three points marked as vertices.

(ii) One of the sides, say $|ab|$, equals $\pi/\sqrt{k}$. In this case the comparison triangle is no longer unique: e.g., one can take two antipodal points on the sphere for $\bar{a}$ and $\bar{b}$ and connect them by any two great half-circles. Then one places $\bar{c}$ on one of these half-circles and considers the other one as the side $[\bar{a}\bar{b}]$.

It follows from these considerations in order to have unique comparison triangles one needs to impose the following conditions on $\Delta abc$: $\max(|ab|, |ac|, |bc|) < \pi/\sqrt{k}$ and $|ab|+|ac|+|bc| \leq 2\pi/\sqrt{k}$. Finally, we mention without proof the following globalization theorems ([BBI01, Th. 9.2.9, Th. 10.3.1]):

**3.5.6 Remark (Globalization Theorems).**

(i) Globalization theorem for non-positive curvature: Every complete simply connected space of curvature $\leq k$, where $k \leq 0$, is a space of curvature $\leq k$ in the large.

(ii) Toponogov's globalization theorem: For any $k \in \mathbb{R}$, every complete space of curvature $\geq k$ is a space of curvature $\geq k$ in the large.

# Chapter 4

# Riemannian Manifolds and Length Structures

In this chapter we build a bridge between metric and Riemannian geometry. First, in Section 4.1 we demonstrate that any Riemannian manifold can be viewed as a length space, where the length of curves is measured via the Riemannian metric. In fact, we also show that differentiability of the Riemannian metric is not required. Finally, in Section 4.2 we study the Riemannian model spaces of constant curvature that we already used in the previous chapter for defining non-zero curvature bounds.

Throughout this chapter we assume a certain familiarity with Riemannian geometry. Our standard reference will be the lecture notes [KS20]

## 4.1 Riemannian manifolds as length spaces

In this section we mainly follow the recent article [Bur15] in analyzing the natural length structure of Riemannian manifolds. Let $(M, g)$ be a connected smooth Riemannian manifold. If $\gamma : [a, b] \to M$ is a piecewise smooth curve then its length is defined by (cf. [KS20, Def. 2.3.1])

$$L(\gamma) := \int_a^b \|\gamma'(t)\|_g \, dt. \tag{4.1.1}$$

Here, for any $v \in T_pM$, $\|v\|_g = g_p(v, v)^{1/2} \equiv \langle v, v\rangle^{1/2}$ denotes the norm on the tangent space induced by $g$. Let us denote by $A_\infty$ the space of all piecewise smooth curves into $M$. Then $A_\infty$ satisfies all the requirements of a length structure, imposed by Definition 1.1.2: the only non-obvious point is (L4), which follows from the fact that distance-balls form a neighborhood basis of any point (cf. [KS20, Prop. 2.3.6]).

The Riemannian distance function is given by

$$d(p, q) \equiv d_L(p, q) := \inf\{L(\gamma) \,|\, \gamma \in A_\infty, \gamma(a) = p, \gamma(b) = q\}, \qquad p, q \in M. \tag{4.1.2}$$

Comparison with Definition 1.1.6 reveals that $d$ is precisely the length metric with respect to the length structure $(M, A_\infty, L)$. In this way, $(M, d)$ becomes a length space. Furthermore, [KS20, Th. 2.3.9] shows that $d$ induces the natural manifold topology on $M$.

According to Definition 1.2.7, the metric $d$, in turn, induces a variational length $L_d$ on the class $C$ of all continuous paths (with monotonous surjective reparametrizations), giving rise to a length structure $(M, C, L_d)$. The corresponding intrinsic metric $\hat{d} = d_{L_d}$ induced by $L_d$ then coincides with $d$ by Proposition 1.3.1.

Before we proceed we need the following auxiliary result.

**4.1.1 Lemma.** *Let $(M, g)$, $(N, h)$ be smooth Riemannian manifolds and let $f : M \to N$ be $C^1$. Then $f$ is locally Lipschitz as a map $(M, d_g) \to (N, d_h)$.*

**Proof.** Let $U$ be a relatively compact, convex set in $M$ (cf. [KS20, Th. 2.2.7]) and let $p, q \in U$. There exists a geodesic $\gamma : [a, b] \to U$ from $p$ to $q$ realizing the distance between $p$ and $q$: $L_g(\gamma) = d_g(p, q)$. Also, since $\overline{U}$ is compact, the operator norm of $Tf$ on $\overline{U}$ is bounded:

$$C := \sup\{\|T_x f(v)\|_h : x \in \overline{U},\ v \in T_x M,\ \|v\|_g = 1\} < \infty.$$

Since $f \circ \gamma$ connects $f(p)$ to $f(q)$,

$$d(f(p), f(q)) \leq L_h(f \circ \gamma) = \int_a^b \|T_{\gamma(t)} f(\gamma'(t))\|_h \, dt \leq C \int_a^b \|\gamma'(t)\|_g \, dt = C d(p, q),$$

so $C$ is a Lipschitz constant for $f$ on $U$. □

Returning to our main topic, we note that the Riemannian length (4.1.1) is in fact well-defined for a much larger class of curves than just the piecewise smooth ones. Indeed, let $\gamma : [a, b] \to M$ be absolutely continuous (cf. Definition 1.6.6), then the image $\gamma([a, b])$ of $\gamma$ can be covered by finitely, many relatively compact coordinate domains. For any such chart $(U, \varphi)$ and any open set $V \subseteq \overline{V} \Subset U$, $\varphi$ is a Lipschitz map from $(V, d)$ to $\mathbb{R}^n$ by Lemma 4.1.1. We may therefore apply 1.6.7 (iii) to infer that $\varphi \circ \gamma = (\gamma^1, \ldots, \gamma^n)$ is absolutely continuous. Its derivative therefore exists almost everywhere and is integrable (by Proposition 1.6.7 (iv)). Consequently, so is

$$\|\gamma'(t)\|_g = \sqrt{g(\gamma', \gamma')} = \Big| \sum_{i,j} g_{ij}(\gamma(t)) \frac{d\gamma^i}{dt} \frac{d\gamma^j}{dt} \Big|^{1/2} \leq C \sum_i \Big| \frac{d\gamma^i}{dt} \Big|.$$

Thus

$$L(\gamma) = \int_a^b \|\gamma'(t)\| \, dt \tag{4.1.3}$$

exists and is finite. Let us denote by $A_{ac}$ the class of all absolutely continuous curves into $M$. By Proposition 1.6.7 (v), any $\gamma \in A_{ac}$ is rectifiable.
For any curve $\gamma \in A_{ac}$ it appears that we now have two notions of length at hand, namely the Riemannian length $L(\gamma)$ from (4.1.1) and the variational length $L_d(\gamma)$. The following result shows, however, that these notions actually coincide:

**4.1.2 Theorem.** *Let $M$ be a connected manifold with smooth Riemannian metric $g$ and let $\gamma : [a, b] \to M$ be absolutely continuous. Then*

*(i) For almost every $t \in [a, b]$, the metric derivative $v_\gamma(t)$ of $\gamma$ exists and*

$$v_\gamma(t) \equiv \lim_{\varepsilon \to 0} \frac{d(\gamma(t + \varepsilon), \gamma(t))}{|\varepsilon|} = \|\gamma'(t)\|_g.$$

*Moreover, $v_\gamma \in L^1([a, b])$.*

*(ii) $\gamma$ is rectifiable and $L(\gamma) = L_d(\gamma)$.*

**Proof.** (i): By what was said above, combined with Theorem 1.6.8, the set of all $t \in [a, b]$ such that both $v_\gamma(t)$ and $\gamma'(t)$ exist is of full measure. Let $t \in (a, b)$ be any such point.
The exponential map $\exp_{\gamma(t)}$ defines a diffeomorphism on a neighborhood $U$ of $\gamma(t)$. Let $\varepsilon > 0$ such that $\gamma([t - \varepsilon, t + \varepsilon]) \subseteq U$. Then

$$\frac{1}{\varepsilon} d(\gamma(t), \gamma(t + \varepsilon)) = \frac{1}{\varepsilon} \left\| \exp^{-1}_{\gamma(t)}(\gamma(t + \varepsilon)) \right\|_{g_{\gamma(t)}} = \left\| \frac{1}{\varepsilon} \exp^{-1}_{\gamma(t)}(\gamma(t + \varepsilon)) \right\|_{g_{\gamma(t)}},$$

Consequently,

$$v_\gamma(t) = \lim_{\varepsilon \to 0+} \frac{d(\gamma(t), \gamma(t + \varepsilon))}{\varepsilon} = \left\| \frac{d}{d\varepsilon}\Big|_0 \exp^{-1}_{\gamma(t)}(\gamma(t + \varepsilon)) \right\|_{g_{\gamma(t)}} = \Big\| (\underbrace{T_0 \exp_{\gamma(t)}}_{=\mathrm{id}})^{-1}(\gamma'(t)) \Big\|_{g_{\gamma(t)}}$$
$$= \|\gamma'(t)\|_{g_{\gamma(t)}}.$$

Integrability of $v_\gamma$ follows from Theorem 1.6.8 (or, alternatively, from that of $\|\gamma'\|_g$ noted above).
(ii): Using (i) and the rectifiability of $\gamma$ noted above, this is immediate from Theorem 1.6.8. □

Analogously to (4.1.2), we may introduce

$$d_{ac}(p,q) := \inf\{L_d(\gamma) \,|\, \gamma \in A_{ac}, \gamma(a) = p, \gamma(b) = q\}, \qquad p, q \in M. \tag{4.1.4}$$

As it turns out, this notion of distance is precisely the one we started with:

**4.1.3 Proposition.** *For any $p, q \in M$,*

$$d_{ac}(p,q) = \inf\{L(\gamma) \,|\, \gamma \in A_{ac}, \gamma(a) = p, \gamma(b) = q\} = d(p,q).$$

**Proof.** The first equality is immediate from Theorem 4.1.2 (ii). Concerning the second one, we have to show that, for any $p, q \in M$,

$$\inf\{L(\gamma) \,|\, \gamma \in A_{ac}, \gamma(a) = p, \gamma(b) = q\} = \inf\{L(\gamma) \,|\, \gamma \in A_\infty, \gamma(a) = p, \gamma(b) = q\}.$$

Here, $\leq$ is immediate from $A_\infty \subseteq A_{ac}$. Now assume that there exist $p, q \in M$ such that $d_{ac}(p,q) < d(p,q)$ and let $\varepsilon > 0$ be such that $d_{ac}(p,q) + \varepsilon < d(p,q)$. Then there would exist an absolutely continuous curve $\gamma : [a,b] \to M$ from $p$ to $q$ with $L(\gamma) < d(p,q) - \varepsilon$, hence strictly shorter than any piecewise smooth curve connecting $p$ and $q$. Now cover $\gamma([a,b])$ by finitely many convex sets (cf. [KS20, Th. 2.2.7]) $(U_i)_{i=1}^m$ and pick points $\gamma(t_i)$ ($t_1 = a < \cdots < t_m = b$) such that $\gamma([t_i, t_{i+1}]) \subseteq U_i$ for $i = 1, \ldots, m-1$. Let $\sigma$ be the concatenation of the radial geodesics connecting $\gamma(t_i)$ to $\gamma(t_{i+1})$. As $L(\gamma) < L(\sigma)$, there must exist some $i$ such that $\gamma|_{[t_i,t_{i+1}]}$ is strictly shorter than the radial geodesic $\sigma|_{[t_i,t_{i+1}]}$. But radial geodesics have minimal length in any normal neighborhood even among all curves that are merely absolutely continuous: the arguments in the proof of this result for piecewise smooth curves (see [KS20, Prop. 2.3.4]) continue to hold almost everywhere and still give the desired conclusion. So we arrive at a contradiction, implying that indeed $d_{ac}(p,q) = d(p,q)$. □

**4.1.4 Remark.** The class $A_{ac}$ is admissible (with the length of curves measured by $L$, the Riemannian length from (4.1.3)): again the only non-obvious property is (L4), which however follows from Proposition 4.1.3, together with the fact that $d$ induces the natural manifold topology. Now if we start out from the length space $(M, A_{ac}, L)$, then by Proposition 4.1.3 the induced length metric $d_{ac}$ is precisely the usual Riemannian distance function $d$ from (4.1.2). Thus $(M, A_\infty, L)$ and $(M, A_{ac}, L)$ give rise to the same metric space $(M, d)$.

Our next goal is to generalize the previous results to manifolds with continuous Riemannian metrics. By this we mean a continuous $(0,2)$-tensor field on $M$ whose restriction to any $T_pM \times T_pM$ is a positive definite scalar product. Such a Riemannian metric also induces a distance function via (4.1.2). However, since we do not have the standard tools of Riemannian geometry available for metrics that are merely continuous (like the exponential map or convex neighborhoods, as employed in the proof of [KS20, Th. 2.3.9]), we first need an independent proof of the fact that the corresponding distance function is indeed a metric and that it induces the natural manifold topology. This will be done in Lemma 4.1.7. To prove it (and also for some considerations later on) we need:

**4.1.5 Lemma.** *Let $g, h$ be continuous Riemannian metrics on $M$, and let $K \Subset M$. Then there exist constants $C_1, C_2 > 0$ such that for all $v \in TM|_K$ we have*

$$C_1 \|v\|_h \leq \|v\|_g \leq C_2 \|v\|_h.$$

**Proof.** Since $K$ can be covered by finitely many coordinate domains, we may without loss of generality suppose that it is in fact contained in a single chart $(\varphi = (x^1, \ldots, x^n), U)$. For $x \in \varphi(K)$ and $\xi \in \mathbb{R}^n \setminus \{0\}$ set

$$f(x,\xi) := \frac{\|T_x\varphi^{-1}(\xi)\|_g}{\|T_x\varphi^{-1}(\xi)\|_h}$$

Then with $\| \, \|_e$ the Euclidean norm on $\mathbb{R}^n$ we have $f(x,\xi) = f(x, \xi/\|\xi\|_e)$. Since $f$ is continuous it therefore suffices to set

$$C_1 := \min\{f(x,\xi) : x \in \varphi(K),\ \xi \in S^{n-1}\} \quad C_2 := \max\{f(x,\xi) : x \in \varphi(K),\ \xi \in S^{n-1}\}.$$

□

**4.1.6 Corollary.** *Under the assumptions of Lemma 4.1.5, if $\gamma \in A_\infty$ has its image contained in $K$, then*

$$C_1 L_h(\gamma) \leq L_g(\gamma) \leq C_2 L_h(\gamma).$$

**Proof.** Immediate from (4.1.1) and Lemma 4.1.5. □

**4.1.7 Lemma.** *Let $g$ be a continuous Riemannian metric on a connected smooth manifold $M$. The corresponding distance function $d$ from (4.1.2) is a metric that induces the natural manifold topology on $M$.*

**Proof.** Clearly, $d$ is non-negative, finite and symmetric. Also the triangle inequality follows as in the proof of Lemma 1.1.7. Note, however, that we cannot simply apply that Lemma here to conclude that $d$ is a metric because we do not (yet) have (L4). In fact, we will infer (L4) from the fact that $d$ is a metric that induces the manifold topology.
Let $(\varphi = (x^1, \dots, x^n), U)$ be a chart around $p \in M$ with $\varphi(p) = 0$. Choose $r > 0$ so small that $B := \overline{B_r(0)} = \{x \in \mathbb{R}^n : \|x\|_e \leq r\} \subseteq \varphi(U)$ and let $\tilde{g} := \varphi_* g$ be the push-forward Riemannian metric on $\varphi(U)$. Then for any curve $\alpha$ in $\varphi(U)$, $L_{\tilde{g}}(\alpha) = L_g(\varphi^{-1} \circ \alpha)$. Since $B$ is compact, Lemma 4.1.5 implies the existence of $C_1, C_2 > 0$ such that

$$C_1 \|v\|_e \leq \tilde{g}_x(v,v)^{1/2} \leq C_2 \|v\|_e \quad \forall x \in B \ \forall v \in \mathbb{R}^n.$$

Given $x \in B$, let $\alpha : [0,1] \to B$, $\alpha(t) := tx$. Then

$$L_{\tilde{g}}(\alpha) = \int_0^1 \tilde{g}_{\alpha(t)}(\alpha'(t), \alpha'(t))^{1/2}\, dt \leq C_2 \int_0^1 \|\alpha'(t)\|_e \, dt = C_2 \|x\|_e.$$

Consequently, $d(p, \varphi^{-1}(x)) \leq C_2 \|x\|_e$ for all $x \in B$, and we conclude that

$$\varphi^{-1}(\overline{B_\rho(0)}) \subseteq \{q : d(p,q) \leq C_2 \rho\} \quad \forall \rho \in (0, r]. \tag{4.1.5}$$

Conversely, let $\beta : [a,b] \to B$ be any piecewise smooth curve in $B$ from 0 to $x$. Then

$$C_1 \|x\|_e = C_1 \Big\| \int_a^b \beta'(t)\, dt \Big\|_e \leq C_1 \int_a^b \|\beta'(t)\|_e \, dt \leq L_{\tilde{g}}(\beta).$$

Consequently, $C_1 \|x\|_e \leq L_g(\gamma)$ for any piecewise smooth curve $\gamma$ from $p$ to $\varphi^{-1}(x)$ that runs entirely in $\varphi^{-1}(B)$. On the other hand, if such a connecting curve leaves $\varphi^{-1}(B)$, then we can repeat the argument with $\beta$ the initial part of $\varphi \circ \gamma$ until the parameter value where $\gamma$ leaves $\varphi^{-1}(B)$ for the first time. It then follows that $C_1 \|x\|_e \leq C_1 r \leq L_{\tilde{g}}(\beta) \leq L_g(\gamma)$ also in this case. Thus we conclude that $C_1 \|x\|_e \leq d(p, \varphi^{-1}(x))$ for all $x \in B$, hence

$$\{q : d(p,q) \leq C_1 \rho\} \subseteq \varphi^{-1}(\overline{B_\rho(0)}) \quad \forall \rho \in (0, r]. \tag{4.1.6}$$

Now if $q \in \varphi^{-1}(B)$ and $q \neq p$, then $0 \neq x := \varphi(q)$ and the above shows that $d(p,q) > 0$. On the other hand, if $q \notin \varphi^{-1}(B)$ then by (4.1.6) we get $d(p,q) > C_1 r > 0$. Summing up, $d(p,q) > 0$ if and only if $p \neq q$.
Finally, that $d$ induces the manifold topology also follows from (4.1.5) and (4.1.6). □

Corollary 4.1.6 implies an analogous result for the Riemannian distance functions induced by continuous metrics:

**4.1.8 Proposition.** *Let $g, h$ be continuous Riemannian metrics on $M$ with corresponding Riemannian distances $d_g$, $d_h$, and let $K \Subset M$. Then there exist constants $C_1, C_2 > 0$ such that for all $p, q \in K$ we have*

$$C_1 d_h(p,q) \leq d_g(p,q) \leq C_2 d_h(p,q).$$

**Proof.** By symmetry it suffices to prove the second inequality. We do this indirectly: suppose that, for any $m \in \mathbb{N}$ there exist $p_m, q_m \in K$ such that

$$d_g(p_m, q_m) > m d_h(p_m, q_m). \tag{4.1.7}$$

Extracting subsequences if necessary, by compactness of $K$ we may assume that $p_m \to p$ and $q_m \to q$ as $m \to \infty$. Since $d_g$ and $d_h$ are continuous with respect to the manifold topology by Lemma 4.1.7, $d_g(p_m, q_m) \to d_g(p,q)$ and $d_h(p_m, q_m) \to d_h(p,q)$. But then (4.1.7) implies that in fact $p = q$.
Since $M$ is locally compact we may choose a compact neighborhood $V$ of $p$. Again by Lemma 4.1.7, we may choose some $r > 0$ such that the $4r$-ball $B^{d_h}_{4r}(p)$ with respect to $d_h$ is contained in $V$. For any $x, y \in B^{d_h}_r(p)$ and for any $\varepsilon \in (0, r)$, by definition of $d_h$ there exists a piecewise smooth path $\gamma : [a,b] \to M$ from $x$ to $y$ with $L_h(\gamma) < d_h(x,y) + \varepsilon$. Any such $\gamma$ cannot leave $B^{d_h}_{4r}(p)$ because for any $t \in [a,b]$ we have

$$\begin{aligned} d_h(p, \gamma(t)) &\leq d_h(p,x) + d_h(x, \gamma(t)) \leq d_h(p,x) + L_h(\gamma|_{[a,t]}) \\ &\leq d_h(p,x) + L_h(\gamma) \leq d_h(p,x) + d_h(x,y) + \varepsilon < 4r. \end{aligned}$$

By Corollary 4.1.6 there exists a constant $C > 0$ such that for any piecewise smooth curve $\sigma$ in $V$ we have $L_g(\sigma) \leq C L_h(\sigma)$. Therefore, for all $x, y \in B^{d_h}_r(p)$ we obtain

$$d_g(x,y) \leq L_g(\gamma) \leq C L_h(\gamma) \leq C d_h(x,y) + C\varepsilon.$$

Letting $\varepsilon \to 0$ it follows that $d_g(x,y) \leq C d_h(x,y)$ for all $x, y \in B^{d_h}_r(p)$. But for $m$ so large that $p_m$ and $q_m$ are contained in $B^{d_h}_r(p)$ and $m > C$, this contradicts (4.1.7). □

**4.1.9 Corollary.** *The conclusion of Lemma 4.1.1 remains valid for continuous Riemannian metrics $g$, $h$.*

We now want to also extend the validity of Theorem 4.1.2 and Corollary 4.1.3 to continuous Riemannian metrics. The difficulty we face is that the proof of Theorem 4.1.2 made essential use of the exponential map, which we no longer have at our disposal for continuous Riemannian metrics. A remedy for this problem lies in approximating continuous by smooth Riemannian metrics. To not overburden the proof of the next result let us briefly recall how to uniformly approximate continuous functions on $\mathbb{R}^n$ by convolution with a mollifier: Let $f : \mathbb{R}^n \to \mathbb{R}$ be continuous and compactly supported and let $\rho$ be a standard mollifier, i.e., $\rho$ is smooth, $\mathrm{supp}(\rho) \subseteq B_1(0)$, and $\int \rho(x)\, dx = 1$. For $m \in \mathbb{N}$, set $\rho_m(x) := m^n \rho(mx)$. Then

$$f * \rho_m(x) = \int f(x-y)\rho_m(y)\, dy$$

is smooth and

$$|f(x) - f * \rho_m(x)| = \Big| \int (f(x) - f(x-y)) m^n \rho(my)\, dy \Big| \leq \int |f(x) - f(x - z/m)||\rho(z)|\, dz \to 0,$$

$(m \to \infty)$, uniformly on $\mathbb{R}^n$.

**4.1.10 Theorem.** *Let $M$ be a connected smooth manifold with a continuous Riemannian metric $g$. Then there exists a sequence of smooth Riemannian metrics $(g_m)_{m \in \mathbb{N}}$ such that*

*(i) The $g_m$ converge locally uniformly to $g$ on $M$.*

*(ii)*

$$\frac{m-1}{m}\|v\|_g \le \|v\|_{g_m} \le \frac{m+1}{m}\|v\|_g, \qquad v \in T_pM, p \in M. \tag{4.1.8}$$

*(iii) For any $\gamma \in A_{ac}$,*

$$\frac{m-1}{m}L_g(\gamma) \le L_{g_m}(\gamma) \le \frac{m+1}{m}L_g(\gamma).$$

*(iv) The induced distance functions $d_m \equiv d_{g_m}$ converge uniformly to $d \equiv d_g$ on $M$. In fact, for all $m \in \mathbb{N}$ and all $p, q \in M$:*

$$\frac{m}{m+1}d_m(p,q) \le d(p,q) \le \frac{m}{m-1}d_m(p,q).$$

**Proof.** (i) and (ii): Let $p \in M$ and let $K_p$ be a compact neighborhood of $p$ contained in a single chart $(\varphi = (x^1, \dots, x^n), U)$. By component-wise convolution with mollifiers as above, we can approximate $\varphi_* g$ by a sequence of smooth Riemannian metrics $\tilde{h}^p_m$ on $\varphi(U)$ (positive definiteness is an open condition because eigenvalues of a matrix depend continuously on the coefficients). Setting $h^p_m := \varphi^* \tilde{h}^p_m$ we obtain a sequence of smooth local Riemannian metrics that converge to $g$ uniformly on $K_p$. As in the proof of Lemma 4.1.5, for $x \in \varphi(K_p)$ and $\xi \in \mathbb{R}^n \setminus \{0\}$ set

$$f_m(x,\xi) := \frac{\|T_x\varphi^{-1}(\xi)\|_{h^p_m}}{\|T_x\varphi^{-1}(\xi)\|_g}.$$

Then $f_m \to 1$ uniformly on $\varphi(K_p) \times S^{n-1}$, and so by extracting a subsequence if necessary we may assume that

$$\frac{m-1}{m}\|v\|_g \le \|v\|_{h^p_m} \le \frac{m+1}{m}\|v\|_g, \qquad v \in T_qM, q \in K_p.$$

Now pick a partition of unity $\{\chi_p\}_{p\in M}$ subordinate to the cover $\{\mathrm{int}(K_p)\}_{p\in M}$ to patch these local approximations $h^p_m$ of $g$ together and obtain a sequence of approximating smooth Riemannian metrics $g_m := \sum_{p\in M}\chi_p h^p_m$ on $M$ (cf. [KS20, Th. 2.4.4]) satisfying the above estimate globally, i.e. that satisfy (4.1.8).
(iii) This follows exactly as in Corollary 4.1.6.
(iv)Let $p, q \in M$. By definition of $d$, for every $\varepsilon > 0$ there exists a curve $\gamma^\varepsilon \in A_\infty$ connecting $p$ and $q$ such that $L_g(\gamma^\varepsilon) < d(p,q) + \varepsilon$, and thus by (iii),

$$d(p,q) + \varepsilon > L_g(\gamma^\varepsilon) \ge \frac{m}{m+1}L_{g_m}(\gamma^\varepsilon) \ge \frac{m}{m+1}d_m(p,q),$$

Therefore,

$$\frac{m+1}{m}d(p,q) \ge d_m(p,q), \qquad p, q \in M.$$

Similarly, for every $d_m$ there exists a curve $\gamma^\varepsilon_m$ from $p$ to $q$ in $A_\infty$ satisfying $L_{g_m}(\gamma^\varepsilon_m) < d_m(p,q)+\varepsilon$, so by (iii) we get

$$d(p,q) \le L_g(\gamma^\varepsilon_m) \le \frac{m}{m-1}L_{g_m}(\gamma^\varepsilon_m) < \frac{m}{m-1}d_m(p,q) + \frac{m}{m-1}\varepsilon.$$

Letting $\varepsilon \to 0$ then implies

$$\frac{m-1}{m}d(p,q) \le d_m(p,q), \qquad p, q \in M.$$

□

Based on this we can now generalize both Theorem 4.1.2 and Proposition 4.1.3 to continuous Riemannian metrics:

**4.1.11 Theorem.** *Let $M$ be a connected manifold with a continuous Riemannian metric $g$ with induced distance function $d = d_g$ given by (4.1.2), and let $\gamma : [a, b] \to M$ be absolutely continuous. Then*

*(i) For almost every $t \in [a, b]$, the metric derivative $v_\gamma(t)$ of $\gamma$ exists and*

$$v_\gamma(t) \equiv \lim_{\varepsilon\to 0} \frac{d(\gamma(t+\varepsilon), \gamma(t))}{|\varepsilon|} = \|\gamma'(t)\|_g.$$

*Moreover, $v_\gamma \in L^1([a, b])$.*

*(ii) $\gamma$ is rectifiable and $L(\gamma) = L_d(\gamma)$.*

*(iii) For any $p, q \in M$,*

$$d_{ac}(p, q) = \inf\{L(\gamma) \mid \gamma \in A_{ac}, \gamma(a) = p, \gamma(b) = q\} = d(p, q).$$

**Proof.** (i) Let $g_m$ be a sequence of smooth Riemannian metrics as in Theorem 4.1.10. Let $t \in [a, b]$ be such that $\gamma'(t)$ exists, as well as $v_\gamma^{(m)}(t)$ (the metric derivative of $\gamma$ with respect to $d_m$) for each $m \in \mathbb{N}$. The set of all such points is of full measure in $[a, b]$. For each $m \in \mathbb{N}$ and each $\varepsilon \neq 0$ we have

$$\frac{m}{m+1}\frac{d_m(\gamma(t+\varepsilon), \gamma(t))}{|\varepsilon|} \le \frac{d_g(\gamma(t+\varepsilon), \gamma(t))}{|\varepsilon|} \le \frac{m}{m-1}\frac{d_m(\gamma(t+\varepsilon), \gamma(t))}{|\varepsilon|}.$$

If, for fixed $m$, we let $\varepsilon \to 0$, then by Theorem 4.1.2 we obtain

$$\frac{m}{m+1}\|\gamma'(t)\|_{g_m} \le \liminf_{\varepsilon\to 0} \frac{d_g(\gamma(t+\varepsilon), \gamma(t))}{|\varepsilon|} \le \limsup_{\varepsilon\to 0} \frac{d_g(\gamma(t+\varepsilon), \gamma(t))}{|\varepsilon|} \le \frac{m}{m-1}\|\gamma'(t)\|_{g_m}.$$

Letting $m \to \infty$ (and recalling (4.1.8)) gives the claim. Integrability of $v_\gamma$ follows from Theorem 1.6.8 (or, alternatively from that of $\|\gamma'\|_g$ — the arguments from (4.1.3) remain valid for continuous Riemannian metrics).
(ii) This is immediate by combining Theorem 1.6.8 with (i).
(iii) The first equality is clear by (ii). To prove the second one, again $\le$ is immediate from $A_\infty \subseteq A_{ac}$. Suppose now that this inequality were strict for some $p, q \in M$. Then there would exist some $\varepsilon > 0$ such that $d_{ac}(p, q) + \varepsilon < d(p, q)$. Hence there would be an absolutely continuous curve $\gamma : [a, b] \to M$ from $p$ to $q$ with $L(\gamma) < d(p, q) - \varepsilon$. By Theorem 4.1.10 (iii) and (iv), then, for $m$ sufficiently large we would obtain

$$L_{g_m}(\gamma) < d_m(p, q) - \varepsilon,$$

contradicting Proposition 4.1.3. □

**4.1.12 Corollary.** *For any continuous Riemannian metric $g$ on a connected smooth manifold $M$, the triple $(M, A_{ac}, L)$ (with $L$ the Riemannian length induced by $g$) is a length structure. Moreover, $(M, A_\infty, L)$ and $(M, A_{ac}, L)$ give rise to the same metric length space $(M, d)$.*

**Proof.** Using Lemma 4.1.7 and Theorem 4.1.11 (iii), this follows exactly as in Remark 4.1.4. □

**4.1.13 Remark.** Denoting by $A_1$, $A_{\mathrm{Lip}}$, $A_{H^1}$ the spaces of piecewise $C^1$, Lipschitz, and $H^1$-Sobolev curves (absolutely continuous curves whose derivative is in $L^2$), respectively, we have

$$A_\infty \subseteq A_1 \subseteq A_{\mathrm{Lip}} \subseteq A_{H^1} \subseteq A_{ac}.$$

Therefore, Corollary 4.1.12 remains valid upon replacing $A_{ac}$ by any of these classes.

## 4.2 The Riemannian model spaces

As we have seen in section 3.5, curvature bounds in length spaces rely on triangle comparison with respect to certain Riemannian manifolds of constant sectional curvature: Euclidean space $\mathbb{R}^n$, the $n$-sphere $\mathbb{S}^n$, or the hyperbolic $n$-space $\mathbb{H}^n$. To unify notation, we shall set

$$M_k^n := \begin{cases} \mathbb{S}^n & k = 1 \\ \mathbb{R}^n & k = 0 \\ \mathbb{H}^n & k = -1 \end{cases}$$

$M_k^n$ is a Riemannian manifold of constant curvature $k$. The important information in $k$ is its sign only since spaces of arbitrary curvature of the same sign can be constructed simply by scaling. In this section we derive some fundamental facts about the model spaces $M_k^n$, following [BH99, Ch. I.2].

### 4.2.1 Euclidean space $\mathbb{R}^n$

This is the simplest of the model spaces: $\mathbb{R}^n$ is a Riemannian manifold with metric $g = \sum_{i=1}^n (dx^i)^2$, i.e., $\langle x, y \rangle = \sum_{i=1}^n x^i y^i$. The Riemann tensor of $M_0^n = \mathbb{R}^n$ vanishes identically, hence so does the sectional curvature $k$. Geodesics are straight lines.

If $H$ is a hyperplane in $\mathbb{R}^n$ (i.e., a subspace of codimension 1) and $x \in H$, then using a unit normal vector $u$ to $H$ we may write

$$H = \{y \in \mathbb{R}^n : \langle y - x, u \rangle = 0\}.$$

Every hyperplane can be written as the set of points equidistant from appropriate points $A, B \in \mathbb{R}^n$, the so-called *hyperplane-bisector* of $A$ and $B$. Also, any hyperplane $H$ comes with an associated isometry $r_H$, the *reflection through* $H$: If $P \in H$ and $u$ is a unit normal vector to $H$, then for $A \in \mathbb{R}^n$ we have

$$r_H(A) = A - 2\langle A - P, u \rangle u.$$

Then $H$ is the set of fixed points of $r_H$. If $A \notin H$, then $H$ is the hyperplane bisector of $A$ and $r_H(A)$. Conversely, if $H$ is the hyperplane bisector of $A$ and $B$, then $r_H(A) = B$.

### 4.2.2 The sphere $\mathbb{S}^n$

In $\mathbb{R}^{n+1}$ with the standard scalar product, we consider the $n$-dimensional sphere

$$\mathbb{S}^n := \{x \in \mathbb{R}^{n+1} : \langle x, x \rangle = 1\}.$$

With the metric induced by the standard metric on $\mathbb{R}^n$, $\mathbb{S}^n$ is a Riemannian manifold. It can be shown (cf. [O'N83, Cor. 4.11]) that the geodesics in $\mathbb{S}^n$ are precisely the (constant speed reparametrizations of parts of) great circles. They are segments within the intersection of $\mathbb{S}^n$ with planes through the origin in $\mathbb{R}^{n+1}$. As we do not assume familiarity with the theory of Riemannian submanifolds, we will equip $\mathbb{S}^n$ with a metric directly:

Given $A, B \in \mathbb{S}^n$, let $d(A, B)$ be the unique element of $[0, \pi]$ such that $\cos d(A, B) = \langle A, B \rangle$.

**4.2.1 Lemma.** *The map $d : \mathbb{S}^n \times \mathbb{S}^n \to \mathbb{R}$ is a metric.*

**Proof.** Non-negativity and symmetry are obvious, as is the implication $d(A, B) = 0 \Leftarrow A = B$. Next, note that $d(A, B)$ is precisely the angle between the ($\mathbb{R}^{n+1}$-)geodesic segments $[O, A]$ and $[O, B]$. The triangle inequality for $d$ therefore follows from Theorem 2.4.7. □

In order to be consistent with our approach to the metric geometry of the hyperbolic space $\mathbb{H}^n$ in the next section, we now give an alternative proof of the triangle inequality in $\mathbb{S}^n$, this time based on the spherical law of cosines. To this end we first observe that there is a natural parametrization for any great circle on $\mathbb{S}^n$. In fact, let $A \in \mathbb{S}^n$ and let $u \in \mathbb{R}^{n+1}$ be a unit vector with $\langle u, A \rangle = 0$. Then for any $a \in [0, \pi]$, consider the path $c : [0, a] \to \mathbb{S}^n$,

$$c(t) = (\cos t)A + (\sin t)u. \qquad (4.2.1)$$

Then

$$\cos(d(c(t), c(t'))) = \langle c(t), c(t')\rangle = \cos t \cos t' + \sin t \sin t' = \cos(t - t'),$$

so $d(c(t), c(t')) = |t - t'|$ for all $t, t' \in [0, a]$. We will call $c$ the great arc with initial position $A$ and initial velocity $u$. For $a = \pi$ we have $c(\pi) = -A$ irrespective of the choice of $u$. If, on the other hand, $d(A, B) < \pi$ then there is a unique minimal great arc from $A$ to $B$. If $A \neq B$ then the initial velocity vector $u$ is the unit vector in the direction of $B - \langle A, B\rangle A$.
We define the *spherical angle* between two minimal arcs emanating from the same point of $\mathbb{S}^n$ with initial velocities $u$ and $v$ to be the unique $\alpha \in [0, \pi]$ such that $\cos\alpha = \langle u, v\rangle$. A spherical triangle $\Delta = (ABC)$ is a choice of three distinct points $A, B, C \in \mathbb{S}^n$ and three minimal great arcs, its sides, joining the vertices. The vertex angle at, say, $C$, is defined to be the spherical angle between the sides $CA$ and $CB$.

**4.2.2 Proposition (The spherical law of cosines).** *Let $\Delta = (ABC)$ be a spherical triangle with sidelengths $a = d(B, C)$, $b = d(C, A)$, and $c = d(A, B)$. Then for the vertex angle $\gamma$ at $C$ we have:*

$$\cos c = \cos a \cos b + \sin a \sin b \cos\gamma.$$

**Proof.** Denoting by $u$ and $v$ the initial vectors of the sides $CA$ and $CB$, respectively, by definition we have $\cos\gamma = \langle u, v\rangle$. Moreover,

$$\begin{aligned}\cos c = \langle A, B\rangle &= \langle(\cos b)C + (\sin b)u, (\cos a)C + (\sin a)v\rangle\\ &= \cos a \cos b\langle C, C\rangle + \sin a \sin b\langle u, v\rangle\\ &= \cos a \cos b + \sin a \sin b \cos\gamma.\end{aligned}$$

□

Using this result, we can now re-prove the triangle inequality for $d$, plus several other fundamental property of $\mathbb{S}^n$ as a metric space:

**4.2.3 Theorem.**

*(i) For all $A, B, C \in \mathbb{S}^n$, $d(A, B) \leq d(A, C) + d(C, B)$. Equality holds if and only if $C$ lies on a minimal great arc connecting $A$ and $B$.*

*(ii) $(\mathbb{S}^n, d)$ is a geodesic (i.e., strictly intrinsic) length space.*

*(iii) The geodesic segments in $\mathbb{S}^n$ are the minimal great arcs.*

*(iv) If $d(A, B) < \pi$ then there is a unique geodesic joining $A$ and $B$.*

*(v) Any open (resp. closed) ball of radius $r \leq \pi/2$ (resp. $r < \pi/2$) in $\mathbb{S}^n$ is convex.*

**Proof.** (i) Let $a = d(C, B)$, $b = d(A, C)$, $c = d(A, B)$, with $A$, $B$, $C$ distinct and let $\Delta = (ABC)$ with vertex angle $\gamma$ at $C$. Since cos is strictly decreasing on $[0, \pi]$, for fixed $a$ and $b$, as $\gamma$ increases from 0 to $\pi$, the function $\gamma \mapsto \cos a \cos b + \sin a \sin b \cos\gamma$ decreases from $\cos(b - a)$ to $\cos(b + a)$. Hence the spherical law of cosines implies that $\cos c \geq \cos(b + a)$, and thereby $c \leq a + b$, with equality if and only if $\gamma = \pi$ and $b + a \leq \pi$. In particular, equality holds if and only if $C$ lies on a minimal great arc from $A$ to $B$.
(ii), (iii) and (iv) are clear from (i).
(v) If $d(A, B) < \pi$ then there is a unique minimal great arc from $A$ to $B$, which is the intersection of $\mathbb{S}^n$ with the positive cone in $\mathbb{R}^{n+1}$ spanned by $A$ and $B$. Any element of this intersection can be written in the form $\lambda A + \mu B$, where $\lambda + \mu \geq 1$. Now by definition of $d$, any point $C \in \mathbb{S}^n$ lies in the closed $r$-ball around $P$ if and only if $\langle C, P\rangle \geq \cos r$. Hence if $A, B \in \bar{B}_r(P)$ for some $r < \pi/2$, then for any element of the minimal great arc connecting them we obtain

$$\langle\lambda A + \mu B, P\rangle = \lambda\langle A, P\rangle + \mu\langle B, P\rangle \geq (\lambda + \mu)\cos r \geq r.$$

This means that the entire geodesic arc from $A$ to $B$ is contained in $\bar{B}_r(P)$. The other claim follows in the same way. □

**4.2.4 Remark.** Theorem 4.2.3 allows us to conclude that the distance function $d$ introduced above coincides with the distance function induced by the Riemannian metric on $\mathbb{S}^n$. Indeed, we already know that for both these metrics the geodesics are precisely the great arcs. Denote by $d_g$ the distance function generated by the Riemannian metric $g$ induced on $\mathbb{S}^n$ by the Euclidean metric on $\mathbb{R}^{n+1}$. Then using the representation of a great arc given in (4.2.1), we have $c'(t) = -(\sin t)A + (\cos t)u$, and so $\|c'(t)\|_g = 1$, giving

$$d_g(c(t), c(t')) = \int_t^{t'} 1\,dt = t' - t = d(c(t), c(t')) \qquad (t \le t').$$

**4.2.5 Remark (Hyperplanes in $\mathbb{S}^n$).** We define a hyperplane $H$ in $\mathbb{S}^n$ to be the intersection of $\mathbb{S}^n$ with an $n$-dimensional subspace of $\mathbb{R}^{n+1}$. Then $H$, equipped with the induced metric from $\mathbb{S}^n$ is isometric to $\mathbb{S}^{n-1}$. The *reflection* $r_H$ through $H$ is defined to be the isometry of $\mathbb{S}^n$ obtained by restricting to $\mathbb{S}^n$ the Euclidean reflection through the $\mathbb{R}^{n+1}$-hyperplane generating $H$. For any two distinct points $A, B \in \mathbb{S}^n$, the set of points equidistant to $A$ and $B$ is a hyperplane, the so-called *hyperplane-bisector* for $A$ and $B$. It is the intersection of $\mathbb{S}^n$ with the subspace of $\mathbb{R}^{n+1}$ orthogonal to the connecting vector $A - B$. If $A \in \mathbb{S}^n$ does not belong to $H$, then $H$ is precisely the hyperplane bisector of $A$ and $r_H(A)$. Conversely, for $H$ the hyperplane bisector of $A$ and $B$ we have $r_H(A) = B$ and $r_H(B) = A$.

### 4.2.3 Hyperbolic space $\mathbb{H}^n$

There are several equivalent ways of representing hyperbolic space. We will focus on realizing it in a manner similar to the one we followed for describing $\mathbb{S}^n$ as a subspace of $\mathbb{R}^{n+1}$. However, this time we will need to equip $\mathbb{R}^{n+1}$ with a metric of Lorentzian signature: By $\mathbb{R}^{n+1}_1$ we denote the $(n+1)$-dimensional Minkowski space. We write $x = (x_0, \dots, x_n) \equiv (x_0, x') \in \mathbb{R}^{n+1}_1$ for the coordinates and use the Lorentzian metric

$$\langle x, y\rangle := -x_0 y_0 + \sum_{i=1}^n x_i y_i.$$

Vectors $v \in \mathbb{R}^{n+1}_1$ are called timelike, spacelike, or null, if $\langle v, v\rangle < 0$, $\langle v, v\rangle > 0$, or $\langle v, v\rangle = 0$, respectively. The orthogonal complement of a timelike vector is a spacelike hypersurface, whereas the orthogonal complement of a null vector is tangent to the light cone $\{x : \langle x, x\rangle = 0\}$ (cf. [O'N83, Ch. 5]).

**4.2.6 Definition (Hyperbolic space).** *Hyperbolic $n$-space $\mathbb{H}^n$ is defined as*

$$\mathbb{H}^n := \{x \in \mathbb{R}^{n+1}_1 : \langle x, x\rangle = -1,\ x_0 > 0\}.$$

Thus $\mathbb{H}^n$ is the upper sheet of the hyperboloid $\{x \in \mathbb{R}^{n+1}_1 : \langle x, x\rangle = -1\}$. If $x \in \mathbb{H}^n$, then $x_0 \ge 1$, and equality holds if and only if $x_i = 0$ for all $i = 1, \dots, n$.

**4.2.7 Remark.** $\mathbb{H}^n$, with the induced metric from $\mathbb{R}^{n+1}_1$ is a Riemannian manifold: in fact, let $p \in \mathbb{H}^n$ and let $v \in T_p\mathbb{H}^n$. Let $c$ be a smooth curve in $\mathbb{H}^n$ with $c(0) = p$ and $c'(0) = v$. Then differentiating the constant function $t \mapsto \langle c(t), c(t)\rangle \equiv -1$ at $t = 0$ it follows that $p \perp v$. Thus the timelike vector $p$ is orthogonal to the tangent plane at $p$, implying that the latter is spacelike. In other words, the restriction of the Minkowski metric to $\mathbb{H}^n$ is Riemannian.

**4.2.8 Lemma.** *For any $x, y \in \mathbb{H}^n$ we have $\langle x, y\rangle \le -1$, with equality if and only if $x = y$.*

**Proof.**

$$\begin{aligned}\langle x, y\rangle &= -x_0 y_0 + \sum_{i=1}^n x_i y_i \le -x_0 y_0 + \Big(\sum_{i=1}^n x_i^2\Big)^{\frac12}\Big(\sum_{i=1}^n y_i^2\Big)^{\frac12} \\ &= -x_0 y_0 + (x_0^2 - 1)^{\frac12}(y_0^2 - 1)^{\frac12} =: f(x_0, y_0).\end{aligned} \tag{4.2.2}$$

By what was said above, $f : [1, \infty) \times [1, \infty) \to \mathbb{R}$ and one easily checks that on this domain $f \leq -1$, with $f(x_0, y_0) = -1$ if and only if $x_0 = y_0$. Inserting this in (4.2.2), we conclude that $\langle x, y \rangle \leq -1$, with equality if and only if $x_0 = y_0$ and thereby, since $x, y \in \mathbb{H}^n$, $\sum_{i=1}^n x_i^2 = \sum_{i=1}^n y_i^2$. It follows that $\langle x, y \rangle = -1$ if and only if there is equality everywhere in (4.2.2). Applying the Cauchy-Schwarz-inequality on $\mathbb{R}^n$ implies that $x'$ must be proportional to $y'$. Combining all these constraints, the only remaining possibility is $x = y$, as claimed. □

It follows that, for all $x, y \in \mathbb{H}^n$, we have $\langle x - y, x - y \rangle = -2(1 + \langle x, y \rangle) \geq 0$, with equality if and only if $x = y$. Also, we may now introduce a metric on $\mathbb{H}^n$:

**4.2.9 Proposition (Metric on $\mathbb{H}^n$).** *Let $d : \mathbb{H}^n \times \mathbb{H}^n \to \mathbb{R}$ be the map that assigns to each pair $(A, B) \in \mathbb{H}^n \times \mathbb{H}^n$ the unique non-negative number $d(A, B) \geq 0$ such that*

$$\cosh d(A, B) = -\langle A, B \rangle.$$

*Then $d$ is a metric on $\mathbb{H}^n$.*

**Proof.** By Lemma 4.2.8, $-\langle A, B \rangle \geq 1$ for every $A$, $B$, so there is indeed a unique non-negative number satisfying the above requirement, i.e., $d$ is well-defined. The lemma also implies that $d$ is positive definite, and symmetry is obvious. It remains to show that the triangle inequality holds. As in the case of $\mathbb{S}^n$ we will derive this from a suitable version of the law of cosines, see theorem 4.2.12 (i) below. □

It will turn out that, as for $\mathbb{S}^n$, also for $\mathbb{H}^n$ geodesics are precisely the sub-arcs of intersections of $\mathbb{H}^n$ with two-dimensional subspaces of Minkowski space $\mathbb{R}^{n+1}_1$. We therefore define a *hyperbolic segment* to be such a curve: Given $A \in \mathbb{H}^n$ and a unit vector $u \in A^\perp \subseteq \mathbb{R}^{n+1}_1$ (i.e., $u$ spacelike, $\langle u, u \rangle = 1$ and $\langle A, u \rangle = 0$), define $c : \mathbb{R} \to \mathbb{H}^n$ by

$$c(t) := (\cosh t)A + (\sinh t)u. \tag{4.2.3}$$

Then $\langle c(t), c(t) \rangle = -\cosh^2(t) + \sinh^2(t) = -1$, so $c$ is indeed a curve in $\mathbb{H}^n$. Moreover,

$$\cosh(d(c(t), c(t'))) = -\langle c(t), c(t') \rangle = \cosh(t)\cosh(t') - \sinh(t)\sinh(t') = \cosh(t - t'),$$

so

$$d(c(t), c(t')) = |t - t'| \tag{4.2.4}$$

for all $t, t' \in \mathbb{R}$. For any $a > 0$, we call $c([0, a])$ the hyperbolic segment from $A$ to $c(a)$, also denoted by $[A, c(a)]$.

Now let $A \neq B$ be two points in $\mathbb{H}^n$ and let $u$ be the unit vector in the direction of $B + \langle A, B \rangle A$. Then $\langle u, A \rangle = 0$, so $u \in A^\perp$. We claim that $u$ is the unique unit vector perpendicular to $A$ such that

$$B = (\cosh a)A + (\sinh a)u, \tag{4.2.5}$$

where $a := d(A, B)$. In fact, $\cosh a = -\langle A, B \rangle$, and

$$\langle B + \langle A, B \rangle A, B + \langle A, B \rangle A \rangle = -1 + \langle A, B \rangle^2 = -1 + \cosh^2(a) = \sinh^2(a).$$

We call $u$ the initial vector of the hyperbolic segment $[AB]$.

**4.2.10 Definition.** *The* hyperbolic angle *between two segments $[AB]$ and $[AC]$, with initial vectors $u$ and $v$, respectively, is the unique $\alpha \in [0, \pi]$ such that $\cos \alpha = \langle u, v \rangle$.*

Note that $u, v$, being elements of $A^\perp$, are spacelike, and $\langle \, , \, \rangle$ is positive definite on $A^\perp$, so there indeed is a unique $\alpha \in [0, \pi]$ satisfying these requirements.

A *hyperbolic triangle* $\Delta = (ABC)$ is a choice of three distinct vertices $A, B, C \in \mathbb{H}^n$, together with three hyperbolic segments joining the vertices, its sides. The *vertex angle* at $C$ is the hyperbolic angle between the segments $[CA]$ and $[CB]$.

**4.2.11 Proposition (The hyperbolic law of cosines).** *Let $\Delta = (ABC)$ be a hyperbolic triangle and set $a = d(B,C)$, $b = d(C,A)$, and $c = d(A,B)$. Then for the vertex angle $\gamma$ at $C$ we have:*

$$\cosh c = \cosh a \cosh b - \sinh a \sinh b \cos\gamma.$$

**Proof.** With $u$ and $v$ the initial vectors of the hyperbolic segments from $C$ to $A$ and $C$ to $B$, respectively, we have $\cos\gamma = \langle u, v\rangle$. Since $\langle u, C\rangle = \langle v, C\rangle = 0$ and $\langle C, C\rangle = -1$, using (4.2.5) we calculate:

$$\begin{aligned}\cosh c &= -\langle A, B\rangle = -\langle(\cosh b)C + (\sinh b)u, (\cosh a)C + (\sinh a)v\rangle\\ &= -\cosh a \cosh b\langle C, C\rangle - \sinh a \sinh b\langle u, v\rangle\\ &= \cosh a \cosh b - \sinh a \sinh b \cos\gamma.\end{aligned}$$

□

Using this we obtain the following analogue of Theorem 4.2.3:

**4.2.12 Theorem.**

*(i) For all $A, B, C \in \mathbb{H}^n$, $d(A,B) \le d(A,C) + d(C,B)$. Equality holds if and only if $C$ lies on the hyperbolic segment connecting $A$ and $B$.*

*(ii) $(\mathbb{H}^n, d)$ is a geodesic length space.*

*(iii) The geodesic segments in $\mathbb{H}^n$ are precisely the hyperbolic segments $[AB]$.*

*(iv) If the intersection of a two-dimensional subspace of $\mathbb{R}^{n+1}_1$ with $\mathbb{H}^n$ is non-empty, then it is a geodesic, and all geodesics in $\mathbb{H}^n$ arise in this way.*

*(v) Any open (resp. closed) ball in $\mathbb{H}^n$ is convex.*

**Proof.** Again let $a = d(C,B)$, $b = d(A,C)$, $c = d(A,B)$, with $A$, $B$, $C$ distinct and let $\Delta = (ABC)$ be the corresponding hyperbolic triangle with vertex angle $\gamma$ at $C$. Then for fixed $a$ and $b$, as $\gamma$ increases from 0 to $\pi$, the function $\gamma \mapsto \cosh a \cosh b - \sinh a \sinh b \cos\gamma$ strictly increases from $\cosh(b-a)$ to $\cosh(b+a)$. Thus Proposition 4.2.11 implies that $\cosh c \le \cosh(b+a)$ and thereby $c \le a + b$, with equality if and only if $\gamma = \pi$. On the other hand, $\gamma = \pi$ if and only if $C \in [AB]$. This gives (i).
Moreover, since along any geodesic the triangle inequality becomes an equality, the above implies that it must lie in a hyperbolic segment, i.e., it must be a parametrization of such a segment. Then (4.2.4) shows that it must in fact *be* such a segment (when parametrized by arclength). Conversely, the intersection of a two-dimensional subspace of $\mathbb{R}^{n+1}_1$ with $\mathbb{H}^n$, if non-empty, is a hyperbolic segment, giving (ii) and (iii).
The proof of (v) is analogous to that of (v) in Theorem 4.2.3, only this time, elements of $[AB]$ are of the form $\lambda A + \mu B$, with $\lambda + \mu \le 1$ and $\lambda, \mu \ge 0$. Now $d(P,C) \le r$ if and only if $-\langle P, C\rangle \le \cosh r$. Thus for $A, B \in \bar{B}_r(P)$,

$$-\langle P, \lambda A + \mu B\rangle = -\lambda\langle P, A\rangle - \mu\langle P, B\rangle \le (\lambda + \mu)\cosh r \le \cosh r,$$

showing that $\lambda A + \mu B \in \bar{B}_r(P)$ as well. □

Again as in the case of $\mathbb{S}^n$ we have:

**4.2.13 Remark.** Theorem 4.2.12 implies that the metric from Proposition 4.2.9 coincides with the distance function $d_g$ induced by the Lorentzian metric $g$ on $\mathbb{R}^{n+1}_1$. In fact, by the above and [O'N83, Prop. 4.28], the geodesics in both cases are intersections of 2-planes with $\mathbb{H}^n$, i.e., hyperbolic segments. For such a segment we have by (4.2.3) $c'(t) = (\sinh t)A + (\cosh t)u$, and so $\|c'(t)\|_g = 1$, giving

$$d_g(c(t), c(t')) = \int_t^{t'} 1\, dt = t' - t = d(c(t), c(t')) \qquad (t \le t').$$

**4.2.14 Remark (Hyperplanes in $\mathbb{H}^n$).** We define a hyperplane $H$ in $\mathbb{H}^n$ to be the intersection of $\mathbb{H}^n$ with an $n$-dimensional subspace of $\mathbb{R}^{n+1}_1$. Then $H$ equipped with the induced metric from $\mathbb{H}^n$ is isometric to $\mathbb{H}^{n-1}$. The *reflection* $r_H$ through $H$ is given by $X \mapsto X - 2\langle X, u\rangle u$, and is easily checked to be an isometry. For any two distinct points $A, B \in \mathbb{H}^n$, the set of points equidistant to $A$ and $B$ is a hyperplane, the *hyperplane-bisector* for $A$ and $B$. It is the intersection of $\mathbb{H}^n$ with the subspace of $\mathbb{R}^{n+1}_1$ orthogonal to the connecting vector $A - B$. If $A \in \mathbb{H}^n$ does not belong to $H$, then $H$ is precisely the hyperplane bisector of $A$ and $r_H(A)$. Conversely, for $H$ the hyperplane bisector of $A$ and $B$ we have $r_H(A) = B$ and $r_H(B) = A$. Finally, $H$ is the set of fixed points of $r_H$.

Both $\mathbb{S}^n$ and $\mathbb{H}^n$ are length spaces, so at the moment we seemingly have two concepts of angle between geodesics: spherical and hyperbolic angles on the one hand, and Alexandrov angles (i.e., those given in Definition 2.4.2) on the other. In fact, these notions coincide:

**4.2.15 Proposition.** *The spherical (resp. hyperbolic) angle between two geodesic segments $[CA]$ and $[CB]$ in $\mathbb{S}^n$ (resp. $\mathbb{H}^n$) is equal to the Alexandrov angle between them (in particular, the latter exists).*

**Proof.** We only give the proof in the hyperbolic case, the one for $\mathbb{S}^n$ proceeds analogously. Let $\gamma$ be the hyperbolic angle between $[CA]$ and $[CB]$. Then letting $\alpha$ and $\beta$ be arclength-parametrizations of $[CA]$ and $[CB]$, by (2.4.2) we have to show that $\lim_{s,t\to 0} \theta(s,t) = \gamma$.
Set $c_{s,t} := d(\alpha(s), \beta(t))$, then

$$c_{s,t}^2 = s^2 + t^2 - 2st\cos\theta(s,t).$$

By Proposition 4.2.11,

$$\cosh c_{s,t} = \cosh s \cosh t - \sinh s \sinh t \cos\gamma. \tag{4.2.6}$$

The problem is that the inverse of $\cosh|_{[0,\infty]}$ is not differentiable at 0, so simply applying it to (4.2.6) will not suffice. Instead, we introduce the auxiliary function $h : \mathbb{R} \to \mathbb{R}$,

$$h(x) := \sum_{i=1}^{\infty} \frac{1}{(2i)!} x^i.$$

This is an analytic function on $\mathbb{R}$ and $\cosh x - 1 = h(x^2)$. Also, $h(0) = 0$ and $h'(0) = 1/2$, so $h$ is invertible in a neighborhood of 0, with inverse given by a power series of the form

$$h^{-1}(x) = 2x + \sum_{i=2}^{\infty} a_i x^i. \tag{4.2.7}$$

Using (4.2.6) we have

$$\begin{aligned} h(c_{s,t}^2) &= \cosh c_{s,t} - 1 = \cosh s \cosh t - 1 - \sinh s \sinh t \cos\gamma \\ &= (\cosh s - 1)\cosh t + (\cosh t - 1) - \sinh s \sinh t \cos\gamma \\ &= h(s^2)\cosh t + h(t^2) - \sinh s \sinh t \cos\gamma =: r(s,t). \end{aligned}$$

Here, $r : \mathbb{R}^2 \to \mathbb{R}$ satisfies

$$r(0,0) = 0, \quad r(s,0) = h(s^2), \quad r(0,t) = h(t^2). \tag{4.2.8}$$

In addition, the coefficient of $st$ in the power series expansion of $r$ is $-\cos\gamma$. Thus the function $f := h^{-1} \circ r$, which equals $c_{s,t}^2$ for small positive $s$ and $t$, is defined on some disc around the origin and has there a power series expansion that (due to $f(0,0) = h^{-1}(r(0,0)) = 0$) is of the form

$$f(s,t) = \sum_{j=1}^{\infty} f_{j,0} s^j + \sum_{k=1}^{\infty} f_{0,k} t^k - st\Big(\sum_{j=1}^{\infty}\sum_{k=1}^{\infty} f_{j,k} s^{j-1} t^{k-1}\Big). \tag{4.2.9}$$

By (4.2.8), $h(f(s,0)) = r(s,0) = h(s^2)$, and $h(f(0,t)) = r(0,t) = h(t^2)$, so the first and second sum in (4.2.9) equal $s^2$ and $t^2$, respectively. Consequently, for $s$, $t$ small and positive, the term in brackets is $(s^2 + t^2 - c_{s,t}^2)/(st)$. Also, (4.2.7), together with the remark about the $st$-coefficient of $r$ show that $f_{1,1} = 2\cos\gamma$. Thus, finally,

$$\cos\theta(s,t) = \frac{s^2 + t^2 - c_{s,t}^2}{2st} = \cos\gamma + \frac{1}{2}\sum_{k>1 \text{ or } j>1} f_{j,k}s^{j-1}t^{k-1} \to \cos\gamma \quad (s,t \to 0+).$$

□

So far we have only considered spheres and hyperbolic spaces of radius 1. More generally we define:

**4.2.16 Definition (The model spaces $\mathbf{M_k^n}$).** *For any $k \in \mathbb{R}$, we denote by $M_k^n$ the following metric spaces:*

*(i) $M_0^n$ is the Euclidean space $\mathbb{R}^n$.*

*(ii) If $k > 0$ then $M_k^n$ is obtained from $\mathbb{S}^n$ by multiplying the distance function by $1/\sqrt{k}$.*

*(iii) If $k < 0$ then $M_k^n$ is obtained from $\mathbb{H}^n$ by multiplying the distance function by $1/\sqrt{-k}$.*

*For $n = 2$, the space $M_k^2$ is called $k$-plane.*

**4.2.17 Remark.** As Riemannian manifolds, the spaces $M_k^n$ can be realized as spheres resp. hyperbolic spaces of appropriately scaled radius, i.e.,

$$M_k^n = \begin{cases} \{x \in \mathbb{R}^{n+1} : \langle x,x\rangle = 1/k\} & k > 0 \\ \mathbb{R}^n & k = 0 \\ \{x \in \mathbb{R}_1^{n+1} : \langle x,x\rangle = 1/k,\ x_0 > 0\} & k < 0 \end{cases}$$

We extend the diameter-convention (3.5.1) to general $n$:

**4.2.18 Definition (Diameter-convention).** *By $D_k$ we denote the diameter of $M_k^n$. Thus, $D_k = \pi/\sqrt{k}$ for $k > 0$, and $D_k := \infty$ for $k \le 0$.*

**4.2.19 Proposition (Law of cosines in $M_k^n$).** *Let $\Delta = (ABC)$ be a geodesic triangle in $M_k^n$ and set $a = d(B,C)$, $b = d(C,A)$, and $c = d(A,B)$. Then for the vertex angle $\gamma$ at $C$ we have:*

$$\begin{aligned} c^2 &= a^2 + b^2 - 2ab\cos(\gamma) && \text{for } k = 0 \\ \cosh(\sqrt{-k}c) &= \cosh(\sqrt{-k}a)\cosh(\sqrt{-k}b) - \sinh(\sqrt{-k}a)\sinh(\sqrt{-k}b)\cos\gamma && \text{for } k < 0 \\ \cos(\sqrt{k}c) &= \cos(\sqrt{k}a)\cos(\sqrt{k}b) + \sin(\sqrt{k}a)\sin(\sqrt{k}b)\cos\gamma && \text{for } k > 0. \end{aligned}$$

*It follows that, for $a$, $b$ and $k$ fixed, $c$ is a strictly increasing function of $\gamma$, varying from $|b-a|$ to $a+b$ as $\gamma$ varies from $0$ to $\pi$.*

**Proof.** The formulae follow from those of $\mathbb{S}^n$ resp. $\mathbb{H}^n$ by rescaling. □

We also note that a unified notation can be obtained from $\sqrt{-k} = i\sqrt{k}$ for $k > 0$, as well as $\cos(it) = \cosh(t)$ and $\sin(it) = i\sinh(t)$.

Our next aim is to supply the foundations for triangle comparison in the model spaces $M_k^2$. As observed in Section 3.5, this makes it possible to define non-zero curvature bounds. Contrary to the flat case, if $k \ne 0$ then it is a priori not clear that suitable comparison triangles exist. The following result shows that this is indeed the case.

**4.2.20 Proposition (Existence of comparison triangles in $M_k^2$).** *Let $k \in \mathbb{R}$ and let $p,q,r$ be points in a metric space $X$. If $k > 0$ then assume that $d(p,q) + d(q,r) + d(r,p) < 2D_k$. Then there exist points $\bar{p},\bar{q},\bar{r} \in M_k^2$ such that (denoting the metrics in $X$ and $M_k^2$ with the same letter) $d(p,q) = d(\bar{p},\bar{q})$, $d(q,r) = d(\bar{q},\bar{r})$, $d(r,p) = d(\bar{r},\bar{p})$.*

*A triangle $\Delta(\bar{p},\bar{q},\bar{r}) \subseteq M_k^2$ with vertices $\bar{p},\bar{q},\bar{r}$ is called a comparison triangle for the triple $(p,q,r)$. It is unique up to an isometry of $M_k^2$. If $\Delta \subseteq X$ is a geodesic triangle in $X$ with vertices $p,q,r$, then $\Delta(\bar{p},\bar{q},\bar{r})$ is also called a comparison triangle for $\Delta$.*

**Proof.** Let $a = d(p, q)$, $b = d(p, r)$, and $c = d(q, r)$. Without loss of generality, suppose that $a \leq b \leq c$. Then by the triangle inequality in $X$, $c \leq a + b$, so that $c < \pi/\sqrt{k}$ in case $k > 0$. Thus using the corresponding law of cosines (proposition 4.2.19) we obtain a unique $\gamma \in [0, \pi]$. Now fix any point $\bar{p} \in M_k^2$ and construct two geodesic segments $[\bar{p}, \bar{q}]$ and $[\bar{p}, \bar{r}]$ of lengths $a$ and $b$, respectively, forming an angle of $\gamma$. Then the law of cosines implies that $d(\bar{q}, \bar{r}) = c$.
Uniqueness up to isometry will follow from Theorem 4.2.22 below. □

A main tool for comparison geometry is the Alexandrov Lemma 3.3.3. As noted in Remark 3.3.4, it remains true also for model spaces other than the Euclidean plane. We now have the tools at hand to verify this:

**4.2.21 Remark (Alexandrov Lemma for arbitrary model spaces).** The conclusions of Lemma 3.3.3 remain valid for points $A, A', B, B', C, C', D, D' \in M_k^2$ for arbitrary $k \in \mathbb{R}$. The only additional assumption required in the case $k > 0$ is that $d(A, B) + d(B, C) + d(C, D) + d(D, A) < 2D_k$. In fact, as already noted in the proof of Lemma 3.3.3, the only fact used in the proof is that the angle between two sides of a triangle is a monotonically increasing function of the third side. That this remains valid in any model space was established in Proposition 4.2.19.

It remains to prove that comparison triangles in model spaces are unique up to isometries. We will in fact show more, in particular that any isometry is a composition of reflections and that the isometry groups acts transitively on the model spaces.

**4.2.22 Theorem (Transitivity of the isometry group of $M_k^n$).** *Let $k \in \mathbb{N}$ and let $A_1, \ldots, A_m$ and $B_1, \ldots, B_m$ be points in $M_k^n$ such that $d(A_i, A_j) = d(B_i, B_j)$ for all $1 \leq i, j \leq m$. Then there exists an isometry $\Phi$ of $M_k^n$ mapping $A_i$ to $B_i$ for each $i = 1, \ldots, m$. In fact, $\Phi$ can be written as a composition of at most $m$ reflections through hyperplanes.*

**Proof.** By rescaling it suffices to consider the cases $\mathbb{R}^n$, $\mathbb{S}^n$ and $\mathbb{H}^n$. For $k = 1$ and $A_1 \neq B_1$ we may set $\Phi$ equal to the reflection $r_H$, where $H$ is the hyperplane bisector of $A_1$ and $B_1$. Now suppose that the result has already been established for $m-1$ points and let $\Phi$ be an isometry such that $\Phi(A_i) = B_i$ for $i = 1, \ldots, m-1$. If $\Phi(A_m) = \Phi(B_m)$ there is nothing to prove. Otherwise, let $H$ be the hyperplane bisector of $\Phi(A_m)$ and $B_m$. Then $B_i \in H$ for each $i = 1, \ldots, m-1$ because $d(B_i, \Phi(A_m)) = d(\Phi(A_i), \Phi(A_m)) = d(A_i, A_m) = d(B_i, B_m)$. Therefore, $r_H \circ \Phi$ (a product of at most $m$ hyperplane reflections) is the desired isometry. □

**4.2.23 Proposition (Isometries of model spaces).** *Let $\Phi$ be an isometry of $M_k^n$.*

*(i) If $\Phi$ is not the identity, then the set of points it fixes lies in some hyperplane.*

*(ii) If $\Phi$ acts as the identity on some hyperplane $H$, then either $\Phi = \mathrm{id}$ or $\Phi = r_H$.*

*(iii) $\Phi$ can be written as a composition of at most $n + 1$ reflections through hyperplanes.*

**Proof.** (i) Suppose there exists some $A$ such that $\Phi(A) \neq A$. Then all points fixed by $\Phi$ lie in the hyperplane bisector $H$ of $A$ and $\Phi(A)$: In fact, if $B = \Phi(B)$ then $d(A, B) = d(\Phi(A), \Phi(B)) = d(\Phi(A), B)$, so $B \in H$.
(ii) Let $\Phi|_H = \mathrm{id}_H$. By what was shown in (i), if $A \neq \Phi(A)$ then $H$ must be contained in, and hence be equal to, the hyperplane bisector of $A$ and $\Phi(A)$. Consequently, $r_H(A) = \Phi(A)$. Since $A$ was arbitrary, $\Phi = r_H$.
(iii) Fix a set of $n+1$ point $A_0, \ldots, A_n$ that is not contained in any hyperplane, and set $B_i := \Phi(A_i)$ for $i = 0, \ldots, n$. By Theorem 4.2.22, there exists an isometry $\Psi$ sending $A_i$ to $B_i$ for each $i$, and which can be written as a composition of at most $n + 1$ reflections through hyperplanes, then the isometry $\Psi^{-1} \circ \Phi$ fixes each $A_i$. It is therefore the identity by (i). □

# Bibliography


[Bär10] Christian Bär. *Elementary differential geometry.* Cambridge University Press, Cambridge, 2010.

[BBI01] Dmitri Burago, Yuri Burago, and Sergei Ivanov. *A course in metric geometry*, volume 33 of *Graduate Studies in Mathematics.* American Mathematical Society, Providence, RI, 2001.

[BH99] Martin R. Bridson and André Haefliger. *Metric spaces of non-positive curvature*, volume 319 of *Grundlehren der Mathematischen Wissenschaften.* Springer-Verlag, Berlin, 1999.

[Bur15] Annegret Y. Burtscher. Length structures on manifolds with continuous Riemannian metrics. *New York J. Math.*, 21:273–296, 2015.

[Coh13] Donald L. Cohn. *Measure theory.* Birkhäuser Advanced Texts: Basler Lehrbücher. Birkhäuser/Springer, New York, second edition, 2013.

[Die69] Jean Dieudonné. *Foundations of modern analysis.* Academic Press, New York-London, 1969. Enlarged and corrected printing, Pure and Applied Mathematics, Vol. 10-I.

[Els05] Jürgen Elstrodt. *Maß- und Integrationstheorie.* Springer-Lehrbuch. Springer-Verlag, Berlin, fourth edition, 2005. Grundwissen Mathematik.

[KS20] Michael Kunzinger and Roland Steinbauer. Riemannian geometry. Lecture notes, 2020.

[Kun16] Michael Kunzinger. General topology. Lecture course, 2016.

[O'N83] Barrett O'Neill. *Semi-Riemannian geometry*, volume 103 of *Pure and Applied Mathematics.* Academic Press, Inc. New York, 1983.

[Pap14] Athanase Papadopoulos. *Metric spaces, convexity and non-positive curvature*, volume 6 of *IRMA Lectures in Mathematics and Theoretical Physics.* European Mathematical Society (EMS), Zürich, second edition, 2014.

[Sch18] Benedict Schinnerl. Length structures and geodesics on riemannian manifolds of low regularity. Master's Thesis, 2018.

[Wil04] Stephen Willard. *General topology.* Dover Publications, Inc., Mineola, NY, 2004. Reprint of the 1970 original.

# Index