



**HAL**  
open science

# Metrics in geography, their intrinsic properties and their shortest paths: correcting a few misconceptions

Benoît Kloeckner

## ► To cite this version:

Benoît Kloeckner. Metrics in geography, their intrinsic properties and their shortest paths: correcting a few misconceptions. 2023. hal-03924536

**HAL Id: hal-03924536**

**<https://hal.science/hal-03924536>**

Preprint submitted on 5 Jan 2023

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# *Metrics in geography, their intrinsic properties and their shortest paths: correcting a few misconceptions*

Benoît R. Kloeckner \*

January 5, 2023

## **1 Introduction**

This article sits at the interface between geometry, meant as a subfield of mathematics, and geography. Many geometrical notions and questions are of geographical relevance; in particular the mathematical notion of *metric space* can be used as a model for geographical spaces. While the relevance in geography of the three axioms of a metric have been discussed (concerning the triangular inequality, see in particular [FH16, KLR21] and references therein), this is not the purpose of the present article: we will assume that the geographical spaces to be dealt with are adequately modeled by metric spaces, and focus on their mathematical properties that are geographically relevant.

It appears that the geographical literature contains a number of questionable statements about metric spaces and metrics, that are not easily spotted and corrected without quite a bit of experience in mathematics. Consider the following statements:

**Claim A.** *The triangular inequality is logically equivalent to the fact that shortest paths are straight lines.*

**Claim B.** *The Euclidean metric is characterized by the fact that shortest paths are straight line segments.*

---

\*Université Paris-Est, Laboratoire d'Analyse et de Matématiques Appliquées (UMR 8050), UPEM, UPEC, CNRS, F-94010, Créteil, France

Our main goal is to show how both these claims are wrong, and explore some consequences of this (mis)understanding. The situation is quite nontrivial, since it motivated Hilbert to formulate the following problem in his famous list presented in the International Congress of Mathematicians held in Paris in 1900:

*Construct all metrics where lines are geodesics.*

Hilbert

The answers that have been given to this problem in the early XX<sup>th</sup> century will bring some light on the above claims.

Rebutting these claims only makes sense if the affirmation that these beliefs are actually held or expressed in some form is supported; to this end, we will consider quotes from the literature, not to single out their authors since it seems various variants of the above claims are prevalent in the geographical literature, but solely to show the existence of the questionable statements we aim to correct. The choice of quotes is somewhat arbitrary, mostly guided by my lectures; hence works cited are those that attracted my interest, and the misconceptions exposed here are in no way a rebuttal of these works as a whole.

To understand the above claims and how they are wrong, it will be necessary to discuss precisely what it means for a space to be Euclidean. This will involve the concept of *intrinsic properties* of a metric spaces which, in a nutshell, are the properties that do not depend on the way the metric space is *represented*. To understand the relevance of this question to geography, it suffices to mention cartographic projections: understanding how they represent (or fail to represent) a part of the surface of the Earth is a classical and important problem. Let us first discuss this situation without too much formalism.

## 1.1 Cartographic projections and non-Euclideanness of the Earth

While we will not give a formal definition of what it means for a metric space to be Euclidean just yet, let us discuss in which way a region of a sphere is *not* Euclidean. We shall use

$$S^2(r) = \{(x, y, z) \in \mathbb{R}^3 \mid x^2 + y^2 + z^2 = r\}$$

to denote the sphere of radius  $r > 0$  where the exponent 2 represents the dimension of the sphere itself (one needs exactly two coordinates, such as latitude and longitude, to define a point on  $S^2(r)$ ). Given two points on this sphere, the *distance* is meant along the sphere itself: it is the shortest length of a curve drawn on the surface  $S^2(r)$  between the points.

It is well-known that one cannot construct a perfect planar map of the sphere. The first obstruction is that it is a bounded space (every two points are at distance at most  $\pi r$ ) without boundary; so let us restrict to a region  $\Omega$ , a hemisphere say. Then it is slightly less obvious, but still true and well-known, that no perfect planar map of  $\Omega$  exists.

An interesting argument is to say that  $\Omega$  does not satisfy the theorems of Euclidean geometry: a disc of radius  $\rho$  on  $S^2(r)$ , i.e. the set  $D(p, \rho)$  of points of  $S^2(r)$  at (spherical) distance at most  $\rho$  from the given point  $p$ , has area *less* than  $\pi\rho^2$ ; and one can find an equilateral triangle  $T = \{p, q, r\}$  with three right angles in  $\Omega$ . A planar map thus cannot be perfect: it cannot send  $D(p, \rho)$  to a disc with the same radius and area, or send  $T$  to a planar equilateral triangle with three right angles, since those do not exist in the plane.

This argument could lead one to think that the necessary imperfection of a planar map is that it cannot preserve *both* distances and areas, nor *both* distances and angles, but this is misleading. Areas and angles are secondary to distances: they can be recovered from the metric by integral and differential calculus. A planar map of  $\Omega$  that would perfectly respect all pairwise distances would necessarily respect both areas and angles, and thus cannot exist.

It is precisely this that makes the sphere non-Euclidean: not having a representation on the Euclidean plane that respects distances, which is basically the same thing than not respecting the theorems of Euclidean geometry.

Let us take a simpler example to insist on the importance of representation, and consider the three-point space  $\{A, B, C\}$  where  $A$  and  $B$  are at distance 3,  $A$  and  $C$  are at distance 4 and  $B$  and  $C$  are at distance 5. This space is in any meaningful way the “same” as the three-point space  $\{p, q, r\}$  where  $p$  and  $q$  are at distance 3,  $p$  and  $r$  are at distance 4, and  $q$  and  $r$  at distance 5; and it is also the “same” as the subset  $\{(0, 0), (3, 0), (0, 4)\}$  of the Euclidean plane with the usual coordinates. Every property of these three spaces that can be expressed in terms of pairwise distances is indeed either true in all three of these spaces, or false in all three; sharing all intrinsic geometric properties, these three spaces ought to be identified as essentially the same.

More generally, an intrinsic property is a property that keeps the same truth value in any representation of the space, and we will define Euclidean-ness as an intrinsic property: in the terminology of [AH72], a space is Euclidean whenever it can be *transformed* into (part of) the Euclidean plane.

## 1.2 Straight lines and shortest paths

To discuss the precise relation between Euclidean-ness and straight lines, we will also need to define straight lines in a non-metric way: if there is a peculiarity in the fact that the usual Euclidean metric makes straight lines into shortest paths, necessarily the notion of a straight line must be defined without appealing to the Euclidean metric. Indeed if one were to *define* straight lines as shortest paths, then the statement “shortest paths are straight lines” would become the tautology “shortest paths are shortest paths”. The non-metric definition of straight lines will have us brush over what mathematicians call “affine geometry”.

In a general metric space, we may not have straight lines (even when there are shortest paths). For example on the sphere, the important statement about shortest paths is that they follow great circles; but there are no straight lines to be compared to shortest paths. This means that the relation between straight lines and Euclidean shortest paths

is specific to having the plane  $\mathbb{R}^2$  as underlying set of points; but with this restriction, the question of which metrics have straight lines as shortest paths and which ones have not will be thoroughly discussed.

### 1.3 Zoology of metric space

In order to show how wrong claims **A** and **B** are, we will provide many examples of metric spaces, which might be seen as pathological. Even pathological examples are important, as they show what cannot be a *logical* consequence of the definition; and they will hopefully show a small bit of the wealth of various spaces left possible by the axioms of a metric, both helping the conceptualization of metrics and broadening the scope of their use for geographical models. Note that all theorems and examples discussed here are well-known in the mathematical community, and there is no pretense to originality in presenting them: the contribution of this article is solely to bring these mathematical facts to the geographic community. Not all mathematical proofs will be given here; when possible, we shall try to give the main ideas, but we will routinely state some facts without proof in order to focus on our main points.

**Organization of the article.** We start in Section 2 by introducing the necessary background and definitions, giving some first examples of metric spaces on the way. Spaces without shortest paths, or where length of shortest path do not coincide with the distance between the end point are shown – but in the next section, we will concentrate on so-called *geodesic* spaces, where these phenomena do not occur.

Section 3 discusses in details Claim **A** and the role of the Triangular Inequality, while Section 4 discusses Claim **B**, first formally defining intrinsic properties and Euclidean-ness. We discuss surfaces in the space that *are* intrinsically Euclidean in a non-trivial way, contrasting with the sphere, then provide examples of intrinsically Euclidean metrics on the plane where shortest paths depart from straight lines. These spaces could seem exotic but are produced through a simple method: transforming the Euclidean metric by an arbitrary mapping. Next, we show several quite different examples of metrics on the plane that have straight lines as geodesics but are not intrinsically Euclidean. The Manhattan metric will make an appearance, and we take this opportunity to show that it has many more shortest path than one may suspect (in particular, *all* straight lines are shortest path for the Manhattan metric, even slanted ones).

## 2 Definitions: metrics, shortest paths and straight lines

### 2.1 Metric spaces

Let us recall that a *metric space* is defined as a pair  $\mathbf{X} = (X, d)$  where  $X$  is a non-empty set and  $d$  is a *metric*, i.e. a function from  $X \times X$  to  $[0, +\infty)$  satisfying the following axioms:

- i.* separation:  $\forall p, q \in X, d(p, q) = 0$  if and only if  $p = q$ ,

ii. symmetry:  $\forall p, q \in X, d(p, q) = d(q, p)$ ,

iii. triangle inequality:  $\forall p, q, r \in X, d(p, q) \leq d(p, r) + d(r, q)$ .

Given points  $p, q \in X$ , we often call the number  $d(p, q)$  the *distance* between them (sometimes,  $d$  itself is called a distance, but we shall stick to the word metric to better distinguish it, a function, from the particular values it takes).

This definition is very general, and a huge variety of metric space have been defined and studied in the mathematical literature, only a small fraction seem relevant to model geographical spaces. We shall denote by  $\mathbf{E}^2 = (\mathbb{R}^2, d_{\text{euc}})$  the Euclidean plane.

We will need some additional definitions. In a metric space  $\mathbf{X} = (X, d)$ , we denote by the *open ball* of center  $p \in X$  and radius  $r \geq 0$  by  $B(p, r)$ , or to avoid ambiguity  $B_d(p, r)$  or  $B_{\mathbf{X}}(p, r)$ , if  $X$  has been endowed with several metrics or several metric spaces are being considered. The *closed balls* are denoted by  $\bar{B}(p, r)$ :

$$B(p, r) = \{p' \in X \mid d(p, p') < r\} \quad \bar{B}(p, r) = \{p' \in X \mid d(p, p') \leq r\}.$$

Consider a subset  $Y \subset X$ ; then the *restriction* of the metric  $d$  to  $Y$  is simply the map  $Y \times Y$  sending a pair  $(p, q)$  to  $d(p, q)$ : compared to the metric of  $X$ , only the domain has changed. For example if we consider the circle  $Y = S^1 = \{(x, y) \in \mathbb{R}^2 \mid x^2 + y^2 = 1\} \subset \mathbb{R}^2$ , the restriction of the Euclidean metric is the *chordal* metric: while distances are only computed between points lying *on* the circle, they correspond to the Euclidean distance “as the crow flies”, as if the rest of the plane could be used for travel.

A subset  $Y \subset X$  is *open* if for every  $p \in Y$ , there is  $r > 0$  such that  $B_{\mathbf{X}}(p, r) \subset Y$ . This condition avoids “thin” subsets, e.g.  $S^1$  is not open in  $\mathbb{R}^2$ . We will sometimes consider more general regions, but openness will be convenient to state some results, and having too many definitions would obscure the main point.

## 2.2 Shortest paths

The notion of shortest path stems from the notion of *length* of a path; while length could be defined by an integral in certain context, we work here with general metric spaces and a more flexible definition is needed.

**Definition 2.1.** A *path*  $\gamma$  in a metric space  $\mathbf{X} = (X, d)$  is a continuous function from a closed interval  $[a, b] \subset \mathbb{R}$  to  $X$ ; we use the compact notation  $\gamma_t$  for the image point of the time  $t \in [a, b]$ . We say that  $\gamma$  is a path *from*  $p$  *to*  $q$  whenever  $\gamma_a = p$  and  $\gamma_b = q$ ; we denote this by  $\gamma : p \rightsquigarrow q$ . The length of a path is defined as

$$\ell(\gamma) = \sup \left\{ \sum_1^k d(\gamma_{t_{i-1}}, \gamma_{t_i}); t_0 = a \leq t_1 \dots \leq t_k = b \right\}. \quad (1)$$

In other words, we consider time discretizations of the path, adding between  $t_0 = a$  and  $t_k = b$  intermediate times  $t_1, t_2 \dots$  in arbitrary finite number and in increasing order. By the triangular inequality, adding a time to a previous list cannot decrease the sum, hence in the least upper bound will usually imply finer and finer discretizations to

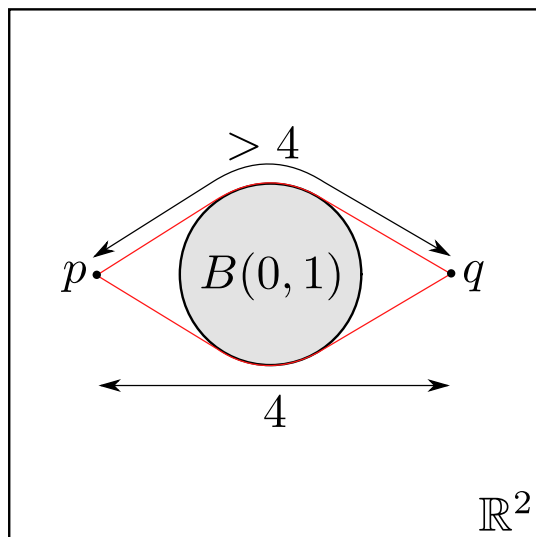


Figure 1: The plane with a disc removed and metric  $d_{\text{euc}}$  has shortest paths (in red) between every pair of points, but some shortest paths are longer than the distance between their endpoints.

be computed. Note that the length of some paths can be infinite, even in very nicely behaved spaces; it is for example the case for the famous von Koch snowflake in the plane.

It will be convenient to distinguish between two types of “efficient” paths.

**Definition 2.2.** A *shortest path* between two points  $p, q$  in a metric space  $\mathbf{X}$  is a curve  $\gamma : p \rightsquigarrow q$  achieving the least possible length, i.e. such that  $\ell(\gamma) = \inf\{\ell(\sigma) \mid \sigma : p \rightsquigarrow q\}$ .

A *geodesic* between  $p$  and  $q$  is a path  $\gamma : p \rightsquigarrow q$  such that  $\ell(\gamma) = d(p, q)$ .

Depending on the space and points, shortest paths and geodesics can exist or not, be unique or not. The triangle inequality implies that for all  $\gamma : p \rightsquigarrow q$ , we have  $\ell(\gamma) \geq d(p, q)$ . Every geodesic is therefore a shortest path, but *shortest paths* may in some cases exist without being geodesics.

**Example 2.3.** Consider the space  $\mathbf{A} = (\mathbb{R}^2 \setminus B(0, 1), d_{\text{euc}})$  of all points at Euclidean distance at least 1 from the origin, endowed with the restriction of the Euclidean distance. This means that for all pairs of point  $p, q$  in the space  $\mathbf{A}$ , thus excluding those in the open ball that has been removed, the distance between them is simply  $d_{\text{euc}}(p, q)$ .

The points  $p = (-2, 0)$  and  $q = (2, 0)$  are at distance 4, but any path  $p \rightsquigarrow q$  has to avoid the removed disc between them and its length is greater than 4 by some definite margin. There exist two shortest paths, one avoiding the removed disc from above and the other from below; both are made of two line segments tangent to the removed disc and an arc of circle along its border (Figure 1). The length of a shortest path is well-defined, but greater than the distance between the points. This provides example of shortest paths that are not geodesics.

It can also happen that no shortest path exist.

**Example 2.4.** Consider the space  $\mathbf{B} = (\mathbb{R}^2 \setminus \{(0, 0)\}, d_{\text{euc}})$  and the points  $p = (-2, 0)$ ,  $q = (2, 0)$ . For each number  $L > d_{\text{euc}}(p, q) = 4$ , we can find a path  $p \rightsquigarrow q$  of length  $L$  by joining two line segments in a way that avoids the removed origin, and for any given such path avoiding the origin one can find a shorter path grazing the origin closer still. There is no path in  $\mathbf{B}$  of length exactly 4, as it would also be the shortest path in  $\mathbf{E}^2$  and thus would be the line segment  $[p, q]$  and pass through the removed point.

In the Example 2.4, the greatest lower bound of the length of paths  $p \rightsquigarrow q$  is equal to the distance between  $p$  and  $q$ , even though it is not achieved by any path. We can combine the phenomenon of the two preceding examples into one.

**Example 2.5.** Consider the space  $\mathbf{C} = (\mathbb{R}^2 \setminus \bar{B}(0, 1), d_{\text{euc}})$  of points in the plane at distance greater than 1 from the origin, and the points  $p = (-2, 0)$ ,  $q = (2, 0)$ . Any path  $p \rightsquigarrow q$  has to avoid the removed *closed* disc between them and its length is greater than 4 by some definite margin. Additionally there is no shortest path  $p \rightsquigarrow q$  (it would have to be one of the only two shortest paths in  $\mathbf{A}$ , which are excluded since here the boundary of the unit disc has been excluded from the space as well).

### 2.3 Length spaces and geodesic spaces

While removing part of the plane and restricting  $d_{\text{euc}}$  to the remaining part could feel arbitrary, the examples above are important as they show the variety of possible behavior left open by the definitions. Other examples looking less made-up could be given, but would be more technical to analyze. In view of these somewhat pathological behavior, it makes a lot of sense to add restriction to the kind of spaces one would like to consider.

**Definition 2.6.** A metric space  $\mathbf{X} = (X, d)$  is said to be a *length space* when the distance between points is the greatest lower bound of the lengths of all paths between the points, i.e. when

$$\forall p, q \in X, \quad d(p, q) = \inf_{\gamma: p \rightsquigarrow q} \ell(\gamma).$$

We say that  $\mathbf{X}$  is a *geodesic space* when between any pair of points, there is a geodesic, i.e. when

$$\forall p, q \in X, \exists \gamma : p \rightsquigarrow q, \quad d(p, q) = \ell(\gamma).$$

One sees at once that geodesic spaces are length spaces, but the converse is not true: Example 2.4 shows a length path that is not geodesic. Examples 2.3 and 2.5 show that some metric spaces are not length spaces.

When given a metric space  $\mathbf{X} = (X, d)$ , there is a simple idea to obtain a length space from it: one simply replaces the metric  $d$  with the *associated length metric*  $d_\ell(p, q) = \inf\{\ell(\gamma) \mid \gamma : p \rightsquigarrow q\}$ .

**Proposition 2.7.** *When  $d_\ell$  is real-valued (i.e. it does not take the value  $+\infty$ , i.e. there are finite-length paths between every two points), then it is a metric on  $X$ .*



*Proof.* We only check the Triangular Inequality, as this Proposition is a good way to understand how this inequality follows from minimization, as argued in [1'H20]. One considers three points  $p, q, r$  and for all paths  $\gamma : p \rightsquigarrow r$ ,  $\eta : r \rightsquigarrow q$  observe that their concatenation is a path  $\gamma * \eta : p \rightsquigarrow q$  whose length is  $\ell(\gamma * \eta) = \ell(\gamma) + \ell(\eta)$ . Taking the greatest lower bound over all possible  $\gamma$ , then over all possible  $\eta$ , we obtain paths going from  $p$  to  $q$  (through  $r$ ) of length arbitrary close to  $d_\ell(p, r) + d_\ell(r, q)$ . The greatest lower bound of the length of *all* paths  $p \rightsquigarrow q$  cannot be larger than this, hence  $d_\ell(p, q) \leq d_\ell(p, r) + d_\ell(r, q)$ .  $\square$

One prominent family of examples is given by surfaces in the 3-dimensional Euclidean space  $\mathbf{E}^3$ ; the restriction of the euclidean metric is not usually relevant in geography (e.g. on the surface of Earth, it models well earthquakes propagation, but not traveling on the surface). Taking the length metric associated with the restriction of the Euclidean 3-dimensional metric gives the usual distance “along” the surface itself. Let us give an simple, very explicit example of this sort.

**Example 2.8.** Consider the unit circle  $S^1 = \{(x, y) \in \mathbb{R}^2 \mid x^2 + y^2 = 1\}$  first endowed with the restriction of the Euclidean metric  $d_{\text{euc}}$ . Then the points  $p = (-1, 0)$  and  $q = (1, 0)$  are at distance  $d_{\text{euc}}(p, q) = 2$ , but the shortest paths are the two half-circles between  $p$  and  $q$  and have length  $\pi > 2$ . We can consider instead the length metric  $d_\ell$  induced by  $d_{\text{euc}}$ , denoting the resulting space  $\mathbf{S}^1 = (S^1, d_\ell)$ . Then  $d_\ell(p, q) = \pi$ , and also for example  $d_\ell(p, (0, 1)) = \pi/2$ . The space  $\mathbf{S}^1$  is not only a length space, it is also geodesic. Observe that most pairs of point have a unique geodesic joining them, but antipodal points have two different geodesics between them.

A similar example is given by the round sphere  $S^2 = \{(x, y, z) \in \mathbb{R}^3 \mid x^2 + y^2 + z^2 = 1\}$ , which could be endowed with the restriction of  $d_{\text{euc}}$ , but is most commonly endowed with the associated length metric  $d_\ell$ ; one can check that  $d_\ell(p, q)$  is the angle (in radians) between the vectors  $\vec{op}$  and  $\vec{oq}$  where  $o$  is the center  $(0, 0, 0)$ . We shall denote by  $\mathbf{S}^2$  the metric space  $(S^2, d_\ell)$ .

However, for this construction to work one needs the original space to admit finite-length curves between all pair of points. Let us show an example where this fails, providing a quite spectacularly non-length space.

**Example 2.9.** The space  $(\mathbb{R}, d^{\frac{1}{2}})$  where  $d^{\frac{1}{2}}(x, y) = \sqrt{|x - y|}$  is not a length space. In fact, for all  $p \neq q$ , all paths  $p \rightsquigarrow q$  are of infinite length. (One can check it is a metric space by using the inequality  $\sqrt{x + y} \leq \sqrt{x} + \sqrt{y}$  for all  $x, y \geq 0$ ; to prove that a path  $x \rightsquigarrow y \neq x$  has infinite length, use a discretization with  $k + 1$  equally spaced times, i.e. with  $d^{\frac{1}{2}}(\gamma_{t_{i-1}}, \gamma_{t_i}) = \sqrt{|x - y|}/\sqrt{k}$  and their sum equal to  $\sqrt{|x - y|} \cdot \sqrt{k}$ , and let  $k$  go to  $\infty$ ).

While at this point the importance of distinguishing between geodesics and shortest paths has been amply stressed, note that *in a length space* the two notions coincide since the distance  $d(p, q)$  is the smallest possible length of a path  $p \rightsquigarrow q$ . When restricting to length spaces, we can thus use “shortest path” or “geodesic” indifferently. Further

restricting to geodesic space then ensures that shortest paths exist. As far as geography is concerned, it seems that, apart from the discrete case (e.g. when a finite number of locations are of interest) the metric spaces of main interest are geodesic spaces. *We shall therefore restrict from now on to geodesic metric spaces.*

Checking that a curve is a geodesic by applying the definition would necessitate to compute its length, which can be cumbersome. The following result gives a simple way to check whether a curve is a geodesic without computing its length (a simple application is given below Example 4.11).

**Proposition 2.10.** *Let  $\mathbf{X} = (X, d)$  be a metric space. A path  $\gamma$  in  $\mathbf{X}$  defined on the interval  $[a, b]$  is a geodesic if and only if for all  $t < u < v \in [a, b]$ :*

$$d(\gamma_t, \gamma_u) + d(\gamma_u, \gamma_v) = d(\gamma_t, \gamma_v). \quad (2)$$

This characterization is also relevant to our second main theme, the triangular inequality: it says that a curve is a geodesic if, and only if, for any three consecutive points on it, the triangular inequality is an equality. A geodesic is thus a curve that makes no *detour* (with respect to the metric for which it is a geodesic, see Section 2.4.1).

*Proof.* Assume that  $\gamma$  is a geodesic; by definition  $d(\gamma_a, \gamma_b) = \ell(\gamma)$ . Consider the discretization given by

$$t_0 = a, \quad t_1 = t, \quad t_2 = u, \quad t_3 = v, \quad t_4 = b$$

then by definition of length as a supremum over all discretizations,

$$\ell(\gamma) \geq d(\gamma_a, \gamma_t) + d(\gamma_t, \gamma_u) + d(\gamma_u, \gamma_v) + d(\gamma_v, \gamma_b).$$

Meanwhile, by the triangle inequality applied three times:

$$\begin{aligned} d(\gamma_a, \gamma_b) &\leq d(\gamma_a, \gamma_t) + d(\gamma_t, \gamma_b) \\ &\leq d(\gamma_a, \gamma_t) + d(\gamma_t, \gamma_v) + d(\gamma_v, \gamma_b) \\ &\leq d(\gamma_a, \gamma_t) + d(\gamma_t, \gamma_u) + d(\gamma_u, \gamma_v) + d(\gamma_v, \gamma_b) \end{aligned}$$

It follows that there is equality in all inequalities above; the last one yields (2).

Assume conversely that (2) holds for all  $t < u < v \in [a, b]$ . Then for all discretization  $a = t_0 < t_1 < \dots < t_k = b \in [a, b]$ , applying (2) repeatedly we get

$$\begin{aligned} \sum_{i=1}^k d(\gamma_{t_{i-1}}, \gamma_{t_i}) &= \sum_{i=1}^{k-2} d(\gamma_{t_{i-1}}, \gamma_{t_i}) + d(\gamma_{t_{k-2}}, \gamma_{t_k}) \\ &= \sum_{i=1}^{k-3} d(\gamma_{t_{i-1}}, \gamma_{t_i}) + d(\gamma_{t_{k-3}}, \gamma_{t_k}) \\ &= \dots \\ &= d(\gamma_{t_0}, \gamma_{t_k}) = d(\gamma_a, \gamma_b). \end{aligned}$$

Taking a supremum, this yields  $\ell(\gamma) = d(\gamma_a, \gamma_b)$ , so that  $\gamma$  is a geodesic.  $\square$

Above, when we spoke about uniqueness of geodesics, we implicitly considered as identical two paths going through the same points in the same order but that could have different parametrizations: we did not distinguish between  $(\gamma_t)_{t \in [a,b]}$  and  $(\gamma_{f(s)})_{s \in [c,d]}$  where  $f : [c, d] \rightarrow [a, b]$  is a one-to-one continuous function, called a *change of parameter*. We shall say that these geodesics *coincide*, even though they are not mathematically identical.

In some cases, it may make sense to restrict to a particular class of parametrization: a geodesic  $\gamma = (\gamma_t)_{t \in [a,b]}$  is said to *have constant speed* whenever for some number  $s \geq 0$  and all  $t_1 < t_2 \in [a, b]$  we have

$$d(\gamma_{t_1}, \gamma_{t_2}) = s|t_1 - t_2|.$$

Then  $s$  is called the *speed* of the geodesic, and must equal  $d(\gamma_b, \gamma_a)/(b - a)$ .

**Example 2.11.** Consider in  $\mathbf{E}^2$  the curves given by  $\gamma_t = (t, t)$  and  $\sigma_t = (t^2, t^2)$  where in both cases  $t$  runs over  $[0, 1]$ . They coincide, both describing the line segment from  $(0, 0)$  to  $(1, 1)$ . We can check that Proposition 2.10 validate them as geodesics: for every  $t < u < v \in [0, 1]$  we have

$$\begin{aligned} d_{\text{euc}}(\gamma_t, \gamma_u) + d_{\text{euc}}(\gamma_u, \gamma_v) &= \sqrt{(u-t)^2 + (u-t)^2} + \sqrt{(v-u)^2 + (v-u)^2} \\ &= \sqrt{2}(u-t) + \sqrt{2}(v-u) \\ &= \sqrt{2}(v-t) \\ &= d_{\text{euc}}(\gamma_t, \gamma_v) \end{aligned}$$

and also

$$\begin{aligned} d_{\text{euc}}(\sigma_t, \sigma_u) + d_{\text{euc}}(\sigma_u, \sigma_v) &= \sqrt{(u^2-t^2)^2 + (u^2-t^2)^2} + \sqrt{(v^2-u^2)^2 + (v^2-u^2)^2} \\ &= \sqrt{2}(u^2-t^2) + \sqrt{2}(v^2-u^2) \\ &= \sqrt{2}(v^2-t^2) \\ &= d_{\text{euc}}(\sigma_t, \sigma_v). \end{aligned}$$

In particular, they both have length  $\sqrt{2}$  (the distance between their endpoints). They correspond to the same travel, but with different speed histories. While  $\gamma$  has constant speed (since  $d_{\text{euc}}(\gamma_{t_1}, \gamma_{t_2}) = \sqrt{2}|t_1 - t_2|$ , and the speed is  $\sqrt{2}$ ), it is not the case of  $\sigma$ : considering times  $t$  and  $t + \varepsilon$  for some small positive  $\varepsilon$ , we have

$$\begin{aligned} d_{\text{euc}}(\sigma_t, \sigma_{t+\varepsilon}) &= \sqrt{((t+\varepsilon)^2 - t^2)^2 + ((t+\varepsilon)^2 - t^2)^2} \\ &= \sqrt{2}((t+\varepsilon)^2 - t^2) \\ &= \sqrt{2}(2t\varepsilon + \varepsilon^2) \simeq (2\sqrt{2}t)\varepsilon. \end{aligned}$$

We see that for a fixed duration  $\varepsilon$  the distance traveled depends on the considered starting time  $t$ . The curve  $\sigma$  is very slow at first, and twice as quick as  $\gamma$  at its very end. Of course, by dividing by the duration  $\varepsilon$  and taking the limit  $\varepsilon \rightarrow 0$  one recovers the instant speed as a derivative.

## 2.4 Affine structure of the plane

Let us now explain how in mathematics we define *straight lines*, as a geometric but not a metric concept.

The set  $\mathbb{R}^2$  has a natural structure of *vector space*: its elements can be added together, and can be multiplied by scalars (with nice properties such as commutativity, distributivity, etc.) As such, its element  $(0, 0)$  plays a particular role: it is the neutral element of addition of vectors. However, when  $\mathbb{R}^2$  is used as a model of “the plane”, one would like to have all its points play the same role. This is translated mathematically into the *affine structure* of  $\mathbb{R}^2$ , given by a difference operation sending a pair of points  $(p, q)$  to their difference, also known as the vector  $\vec{pq}$ . In coordinates, this is simply the operation

$$(p = (x_1, y_1), q = (x_2, y_2)) \mapsto \vec{pq} = [x_2 - x_1; y_2 - y_1]$$

where coordinates of vectors are written between brackets  $[\cdot]$ , in order to distinguish them from points.

From this, one defines three points  $p, q, r$  as *aligned* when the vectors  $\vec{pq}$  and  $\vec{pr}$  are colinear, i.e. one is a scalar multiple of the other. This gives rise to the notion of the *straight line* defined by two different points  $p \neq q$ : the subset of  $\mathbb{R}^2$  made of all points  $r$  such that  $p, q, r$  are aligned.

Note that the notion of midpoint can be defined without any reference to any metric: the midpoint of  $p$  and  $q$  is the point  $p + \frac{1}{2}\vec{pq}$  (of course, it turns out to be equidistant from  $p$  and  $q$  in the usual Euclidean metric).

The goal of this brief recapitulation of some basic definition of plane geometry is to stress that *addition of vectors is essential to defining straight lines*. This is quite a different concept than the metric notion of shortest path, and it is not a superficial statement to say that in  $\mathbf{E}^2$ , shortest paths are straight lines:<sup>1</sup> it relates properties of two different kinds of mathematical structures.

### 2.4.1 On detour and space inversion.

The above notions provide a good way to think about *detour* and *space inversion*. A detour denotes a situation where a traveler, in optimizing its journey, starts in a different direction than the direct route to its goal [l’H17] (or, if the detour happens during the journey, the traveler takes at some point a direction different from the direct route). Space inversion is the most extreme form of detour, when the initial direction of the traveler is opposed to the direct route [Bun62, Tob61], for example when to travel from Auxerre to Marseille in France, one first takes a train north to Paris, then a high-speed train south from Paris to Marseille.

To formulate these definitions, we used the term “direct route” between points, say  $p$  and  $q$ ; but what is this direct route? It is of course understood as the line segment  $[pq]$ , whose direction is indicated by the vector  $\vec{pq}$ : as such, it is an affine notion.

---

<sup>1</sup>This statement is slightly imprecise: actually, shortest paths are straight line *segments*. We will often use this shortcut when no misunderstanding seems possible. When we say that straight lines are shortest paths, we mean that each of their segments are shortest paths between their extremities.

By definition when there is a detour, the shortest path  $\gamma$  is at some of its point  $\gamma_t$  not directed “toward” its endpoint  $q$ , i.e. it does not go in the direction of the vector  $\overrightarrow{\gamma_t q}$ . In other words, the shortest path does not coincide with the straight line segment between its endpoints; and conversely anytime the shortest path is not a straight line, there is a detour. The property (of e.g. the usual Euclidean metric) that straight lines be shortest paths is logically equivalent to the absence of detour. Beware that this only makes an incomplete link between Euclidean space and absence of detour, as we will see below.

In this way, detours is understood as a discrepancy between the metric governing travel and the affine geometry of the plane; but as soon as we consider the problem at a more global scale, this point of view fails: we do not have straight lines on the spherical shape of the Earth, but detour and space inversion can still occur. A second, more general interpretation is that detour occurs when there is a discrepancy between shortest paths of the metric governing travel and the shortest paths of a reference metric. In the case of the plane, the reference metric would be  $d_{\text{euc}}$ , while in the case of the sphere it would be the length metric  $d_\ell$  on  $\mathbf{S}^2$  (Example 2.8). In the former case, we recover the same notion of detour since the shortest paths of  $d_{\text{euc}}$  are straight lines; but now we can also make sense of detour on the sphere. In all cases, detour only seems paradoxical because one confuses two metrics; the shortest path does go in the right direction by definition, just not in the direction given by another metric not relevant to the travel.

### 2.4.2 Affine mappings

It is a natural question to ask which mapping of the plane onto itself preserve the affine structure, in particular send aligned triples to aligned triples and send the midpoint of any two points to the midpoint of their image; these are the *affine transformations*, of the form  $\varphi(x, y) = (a_{11}x + a_{12}y + b_1, a_{21}x + a_{22}y + b_2)$  for some  $a_{ij}, b_j \in \mathbb{R}$  (and  $\varphi$  is invertible if and only if  $a_{11}a_{22} - a_{12}a_{21} \neq 0$ ).

A very beautiful result, the “Fundamental Theorem of Affine Geometry”, states that any one-to-one mapping from  $\mathbb{R}^2$  onto itself that preserves alignment, must actually be an affine transformation, without requesting *a priori* that it preserves any other affine property such as midpoints. This falls down when one considers mappings between *parts* of  $\mathbb{R}^2$ : for example the mapping

$$\begin{aligned} \psi : \left\{ (x, y) \in \mathbb{R}^2 \mid x + y > -1 \right\} &\rightarrow \left\{ (x, y) \in \mathbb{R}^2 \mid x + y < 1 \right\} \\ (x, y) &\mapsto \left( \frac{x}{x + y + 1}, \frac{y}{x + y + 1} \right) \end{aligned}$$

sends aligned triples to aligned triples, but not midpoints to midpoints. This shows the situation is somewhat subtle and will be used below to construct some examples.

## 3 Triangular inequality and shortest paths

Let us discuss in detail Claim A, and some related quotes. They contain a certain level of confusion about the Triangular Inequality, which is in fact ubiquitous in geographical

spaces as clearly laid out by l’Hostis ([’H16] [’H20]); but our interest here is driven by the relation they seek to establish with the Euclideaness of geographical space, or lack thereof.

*Dans un espace formé par les trois lieux  $A$ ,  $B$  et  $C$ , le trajet  $AB + BC$  peut être plus rapide que le trajet direct  $AB$ ,<sup>a</sup> alors que sur la carte (euclidienne),  $AB^a$  est toujours plus court que  $AB + BC$ . Le seul moyen de représenter l’espace géographique dans toute sa complexité non euclidienne est d’user de subterfuges symboliques et de déformations, comme on le fait à l’aide de cartogrammes, par exemple, ou plus simplement dans le dessin des plans de métro.”*

Translation [by B.K.]: “In a space made of three locations  $A$ ,  $B$  and  $C$ , the travel  $AB + BC$  can be faster than the direct travel  $AB$ ,<sup>a</sup> while on the (Euclidean) map,  $AB^a$  is always shorter than  $AB + BC$ . The only way to represent the geographical space with all its non-Euclidean complexity is to rely on symbolic subterfuge and on distortions, as e.g. with cartograms, or simply in the drawing of subway maps”

<sup>a</sup> It should be “ $AC$ ” rather than “ $AB$ ”, but this is a typo that the reader can correct without thinking about it.

Poncet [Pon08]

We can pinpoint the problem with the Triangular Inequality by the following question: what would be the “direct travel” in geographical space? considering the space itself rather than any planar representation of it, the only possible answer is that the direct travel is given by the shortest path; but then if the travel  $AB + BC$  were faster than a travel from  $A$  to  $C$ , then that travel would not be the direct one. This quote also casts the Triangular Inequality as a specificity of Euclidean space, while it is an axiom of all metric spaces. It should also be observed is that it may be that (rewriting distances in our notation)  $d(A, C)$  is equal to  $d(A, B) + d(B, C)$ , even in Euclidean plane  $\mathbf{E}^2$ ; in this case, it suffices that  $B$  be on the line segment  $[A, C]$ . In a general metric space  $\mathbf{X}$ , one can define that a point  $B$  is “between” points  $A$  and  $C$  whenever  $d(A, C) = d(A, B) + d(B, C)$ ; then Proposition 2.10 can be recast as the statement that a curve is a geodesic precisely when all its points are between its endpoints. Another way to phrase this is to say that a detour occurs when the notions of betweenness given by the metric governing travel and the reference metric diverge one from another: when that point which looks in between the starting point and the objective in the reference metric, is in fact not between those points with respect to the metric governing travel.

*[...] la troisième propriété est la fameuse inégalité triangulaire qui affirme que le plus court chemin d'un point à un autre est la ligne droite.*

Translation [by B.K.]: *[...] the third property is the famous Triangle Inequality asserting that the shortest path from a point to another is the straight line.*

Lamure 1998 [Lam98]

Again, the triangular inequality is a property of all metrics, it is valid e.g. on the sphere  $\mathbf{S}^2$ , where there are no straight lines (unless one interprets “straight lines” as geodesics, but then the statement would be empty and would still have nothing to do with the Triangular Inequality).

*[...] L'une des pierres de touche de la géométrie euclidienne réside dans l'inégalité triangulaire ( $AC \leq AB + BC$ ). Or, dans la société, la vitesse (si c'est l'unité de mesure qu'on se donne) peut être plus élevée si l'on passe un point intermédiaire (par exemple en utilisant le réseau peu connexe des TGV ou un hub aérien) que si l'on prend une voie « directe ».*

Translation [by B.K.]: *[...] The Triangle Inequality is one of the touchstones of Euclidean geometry ( $AC \leq AB + BC$ ). However in society the speed (if we choose it as unit of measurement) can be higher if we pass through an intermediate point (e.g. by using the scarcely connected network of high speed trains or an aviation hub) than if we use a “direct” path.*

Lévy 2009 [Lé09]

We see again the issue with the notion of “direct” travel; the metric that gives the travel time between any two places will satisfy the Triangular Inequality even if it differs from  $d_{\text{euc}}$ , with the same argument as in the proof of Proposition 2.7.

The three above quotes seem to confuse the existence of *detours* and the failure of Triangular Inequality. Section 2.4 explains how the latter survives easily to the former: it can happen in a Euclidean representation of a metric space that a point  $B$  is not on the segment  $[A, C]$  but yet, in practice to go from  $A$  to  $C$  it is quicker to go through  $B$  than to move in a straight line. But what this means is simply that the metric governing actual travel does not have the straight line segment  $[A, C]$  as a shortest path (and in particular must differ from  $d_{\text{euc}}$ ); but that metric will still satisfy the Triangular Inequality nonetheless. Detour is only a manifestation of a discrepancy between two metrics, one governing actual travel and one underlying a representation of the geographical space (often, the representation will be on the plane and the latter metric will be  $d_{\text{euc}}$ ).

There is an important consequence of revealing this: when we try to test whether a space is Euclidean (or close to be Euclidean), we shall seek other properties than

failure of the Triangular Inequality. Such failures are indeed mirages, the Triangular Inequality being satisfied whenever the word “distance” is relevant; and the satisfaction of the Triangular Inequality is in no way a sign of Euclideanness, since the very many non-Euclidean geometries all satisfy it.

## 4 Intrinsic properties and Euclideanness

### 4.1 Intrinsic properties of metric spaces

Before considering the precise definition of an intrinsically Euclidean metric, we shall explore the word “intrinsic” in more generality. In mathematics, a property of a metric space is said to be *intrinsic* if it can be expressed solely from the metric itself; this could be somewhat cumbersome to formalize, and an elegant solution is to say that a property is intrinsic whenever it is not changed by any distance-preserving, onto mapping. Such mappings are called isometries; we thus have the following definitions.

**Definition 4.1.** Consider two metric spaces  $\mathbf{X} = (X, d)$  and  $\mathbf{Y} = (Y, \rho)$ . A mapping  $\varphi : X \rightarrow Y$  is said to be an *isometric mapping*<sup>2</sup> from  $\mathbf{X}$  to  $\mathbf{Y}$  when

$$\forall p, q \in X, \quad \rho(\varphi(p), \varphi(q)) = d(p, q).$$

If moreover  $\varphi$  is onto  $Y$ , i.e. every point of  $Y$  is the image of some point of  $X$ , then  $\varphi$  is called an *isometry*.

We say that  $\mathbf{X}$  *maps isometrically into*  $\mathbf{Y}$  when there exist an isometric mapping from  $\mathbf{X}$  to  $\mathbf{Y}$ . If there is an isometry between them, we say that  $\mathbf{X}$  and  $\mathbf{Y}$  are *isometric*.

(Note that the separation property of metrics ensures that isometric mappings are one-to-one, i.e. two different points of  $X$  must be mapped to two different points of  $Y$ .)

Recall our goal is to define *intrinsic* properties of metric spaces, which we described as “being expressed solely from the metric”, but that last expression needs to be reflected upon. To consider a simple example, imagine that someone presents the following metric on  $\mathbb{R}^2$ :

$$d((x_1, y_1), (x_2, y_2)) = 2\sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}$$

and observes that it is not *equal* to  $d_{\text{euc}}$ , since e.g.  $d((0, 0), (1, 1)) = 2\sqrt{2} \neq \sqrt{2} = d_{\text{euc}}((0, 0), (1, 1))$ , and yet has many of its properties (e.g. is a geodesic space, satisfies Thales’ theorem, etc.). This would leave anyone unimpressed, because it is immediately realized that  $d$  is simply a zoomed-in version of  $d_{\text{euc}}$ : if one chooses new coordinates  $z = 2x$ ,  $w = 2y$ , then the points  $p_i = (x_i, y_i)$  have distance

$$d(p_1, p_2) = \sqrt{(z_1 - z_2)^2 + (w_1 - w_2)^2}$$

which is the usual expression of Euclidean distance. Here, the change of coordinates  $\varphi(x, y) = (2x, 2y)$  is an isometry from  $(\mathbb{R}^2, d)$  onto  $\mathbf{E}^2$ . We shall see below what happens

---

<sup>2</sup>In mathematics, the wording “isometric *embedding*” is more common, but the word “mapping” will be more consistent with further definitions.



when a more general change of coordinates is used, but for now this example should make the following definition seem only natural.

**Definition 4.2.** An *intrinsic property*  $\mathcal{P}$  is a mathematical statement that has an well-defined value of truth  $\mathcal{P}_{\mathbf{X}}$  for each metric space  $\mathbf{X}$ , and such that for every two *isometric* metric spaces  $\mathbf{X}, \mathbf{Y}$ , the truth values  $\mathcal{P}_{\mathbf{X}}$  and  $\mathcal{P}_{\mathbf{Y}}$  coincide.

Examples of intrinsic properties include: being a length space, being a geodesic space, being isometric to the Euclidean plane  $\mathbf{E}^2$ , being isometric to some subspace of  $\mathbf{E}^2$ , having finite diameter ( $\exists D > 0, \forall p, q \in X, d(p, q) \leq D$ ).

The property of a metric space  $\mathbf{X} = (X, d)$  asserting that its supporting set is the plane,  $X = \mathbb{R}^2$ , is an example of a well-defined property that is not intrinsic. The property to have straight lines as shortest paths is not defined in general, since  $X$  can be an arbitrary set and not have “straight lines”. If we restrict to spaces such that  $X = \mathbb{R}^2$ , the property to have straight lines as shortest paths becomes well-defined property in this class but its *not* intrinsic, as we shall see later.

Let us propose definitions which, contrary to the previous ones, are not commonly used in mathematics but fit the questions we are interested in.

**Definition 4.3.** We say that a metric space is *Euclidean* when it is isometric to  $\mathbf{E}^2$ ; it is *locally Euclidean* when every point admits a neighborhood isometric to an open domain of  $\mathbf{E}^2$  (up to choosing a smaller neighborhood, we can ask that this neighborhood be a metric ball  $B(p, \varepsilon)$ , so that the domain in  $\mathbf{E}^2$  can always be chosen to be a disc). When we want to stress that we use this definition, we may say that the space is *intrinsically* (locally) Euclidean, but most of the time we will dispense from this adjective and let is implicit.

In other words, a metric space is locally Euclidean when small enough regions can be mapped perfectly, i.e. respecting distances between all pairs of points, on the Euclidean plane. A metric space is Euclidean when the whole space can be given such a perfect planar map, taking up the whole plane. It is important to insist of the meaning of this definition: in order to have an *intrinsic* definition of Euclideaness, we *need* to declare as Euclidean any space that can be perfectly represented by the Euclidean plane, as such a space will necessarily share all the intrinsic properties of the Euclidean plane – thus including the property of being Euclidean. We shall see examples shortly.

We could also have considered the case of metric space having a perfect planar map, but that only cover a (substantial) region of the plane; we can use the locution “space admitting an isometric mapping into an open subset of  $\mathbf{E}^2$ ” to denote these spaces. The reason we do not introduce a specific word is that there are some subtleties on their geodesics depending on the form of the perfect map, into which we shall not delve here.

Note that “intrinsically *planar* (locally) Euclidean” would be a more accurate name, since Euclidean spaces of dimension different from 2 are in the above terminology, oddly, not Euclidean; but we focus here on the case of the plane, and prefer not introduce too cumbersome a wording.

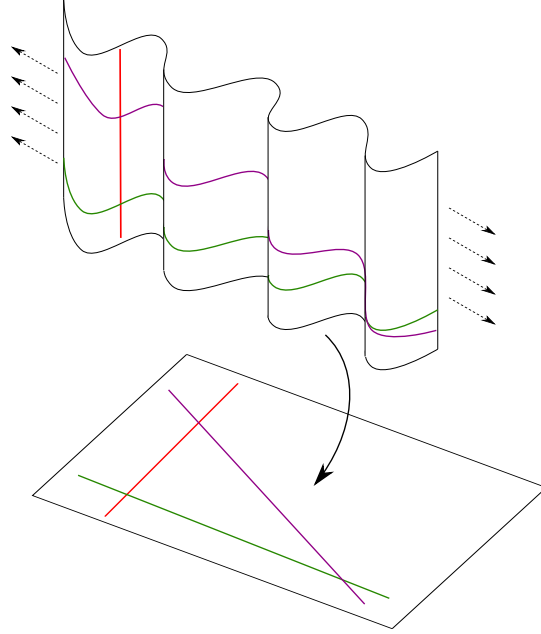


Figure 2: A corrugated surface which is intrinsically Euclidean, even though some of its geodesics wave along the surface: it can be flattened without stretching or compressing distances.

Some locally Euclidean spaces are not length spaces (e.g. Examples 2.3 and 2.5): the qualifier “local” prevents to extract information about far away points. Euclidean spaces are geodesic spaces, since being geodesic is an intrinsic property.

**Example 4.4.** A corrugated surface  $\mathbf{C} = (C, d_\ell)$  such as one found in the thickness of cardboard, e.g.

$$C = \{(x, y, z) \in \mathbb{R}^3 \mid y = \cos(x)\}$$

endowed with the length distance  $d_\ell$  induced by the Euclidean metric of  $\mathbb{R}^3$ , is Euclidean (Figure 2); this is not a completely obvious statement, but the map representing it perfectly in the Euclidean space is easy to guess: one simply iron out the surface to make it flat, and this operation does not dilate or contract any part of it, thus preserving distances. Vertical lines provide some of the geodesics of  $\mathbf{C}$ , but the other geodesic cannot be lines, they follow the waves made by the surface. A small ant living on the surface of such a surface, perceiving only the two dimensions of the surface and not the third dimension of the ambient space, would not be able to tell the difference with a flat world: if the ant were mathematically inclined it would find out that all Euclidean theorems hold, and would certainly end up giving its world an affine structure that would make, in its eyes, shortest paths be straight lines.

**Example 4.5.** An infinite cylinder in the dimension 3 Euclidean space,

$$\{(x, y, z) \in \mathbb{R}^3 \mid x^2 + y^2 = r^2\} \quad (\text{where } r > 0),$$

endowed with the length distance induced by the Euclidean metric of  $\mathbb{R}^3$ , is locally Euclidean (each small enough part of it can be “unrolled” into a plane domain without changing distances along the surface) but not Euclidean and cannot be mapped isometrically into the Euclidean plane (the horizontal circles are locally geodesics, and would thus have to be mapped to straight line segments, but they do not have endpoint).

**Example 4.6.** The unit sphere  $S^2 = \{(x, y, z) \in \mathbb{R}^3 \mid x^2 + y^2 + z^2 = 1\}$  is not locally Euclidean, neither when endowed with the restriction of the Euclidean metric, nor when endowed with the induced length metric (one sophisticated argument slightly explained below is that its Gaussian curvature, an intrinsic quantity, is not zero; a simpler argument is outlined in Example 4.12).

In example 2.8,  $(S^1, d_{\text{euc}})$  maps isometrically into the Euclidean plane, but not  $\mathbf{S}^1 = (S^1, d_\ell)$  which is locally isometric to the *line*  $\mathbf{E}^1$ .

The corrugated surface, the cylinder and the sphere hint together toward a corner stone of differential geometry: *Gauss curvature*. Gauss famously proved (in a differential context) that the value of this curvature is intrinsic, even though its definition very much involves the representation of the surface in the space. Gauss curvature is indeed defined as the product of the two “principal curvatures”, each of which is not intrinsic. In a nutshell, we can say that surfaces that at each point only bend in one direction (such as a cylinder or a corrugated surface) are locally Euclidean: one of their principal curvature is zero, hence the product is zero; and it can be proved that surfaces whose Gauss curvature vanishes are locally Euclidean. But a surface that bends in two independent directions (such as the sphere or a saddle) will have both principal curvatures non-zero, hence non-zero Gauss curvature.

Let us finally define another important class of mappings.

**Definition 4.7.** Let  $\mathbf{X} = (X, d)$  and  $\mathbf{Y} = (Y, \rho)$  be two *geodesic* spaces. A one-to-one mapping  $\varphi : X \rightarrow Y$  is said to be *geodesic* when it sends geodesics of  $\mathbf{X}$  to geodesics of  $\mathbf{Y}$ .

A significant part of our task shall be to observe the (quite large) gap between the notions of geodesic mapping and of isometric mapping. We shall in particular explore counter-examples to Claim B stated in the introduction. Both implications “A Euclidean space must have straight lines as shortest paths” and “A space with straight lines as shortest paths must be Euclidean” are incorrect (and ill-defined unless the underlying set has an affine structure, see Section 2.4), and we will consider them one by one.

## 4.2 Euclidean metrics with non-straight shortest paths

Let us start with the fallacy “A Euclidean space must have straight lines as shortest paths”, insisting on the precise meaning we shall give to this sentence. Of course *the* Euclidean plane  $\mathbf{E}^2 = (\mathbb{R}^2, d_{\text{euc}})$  does have straight lines as its shortest paths, but when

one speaks about *a* Euclidean space, what is meant? When we say that the surface of Earth is not Euclidean, we do not only mean that the underlying set is not  $\mathbb{R}^2$  and that the formula giving  $d_{\text{euc}}$  is not the one giving distance between points. We mean something stronger than that: the surface of Earth cannot be mapped isometrically onto the Euclidean plane (Example 4.6), any planar representation will distort distances. This is because the intrinsic geometric properties of a sphere are not the same as those of the plane (e.g. the Pythagoras theorem does not hold on the sphere, one can find an equilateral triangle with three square angles, etc.) In accordance with this understanding, we use the word Euclidean as per Definition 4.3, i.e. any space isometric to  $\mathbf{E}^2$ .

“Straight lines” are not defined for a general metric space as we explained in Section 2.4. We thus consider metrics on the plane  $\mathbb{R}^2$  or a region of the plane, so that the statement makes sense by using the affine structure of the plane. Disproving the above statement is then quite easy: take  $\mathbf{E}^2$ , apply to it a mapping that does not preserve alignment, and define the metric so as to make this mapping an isometry. From this idea one can construct the following example:

**Example 4.8.** The metric space  $\mathbf{W} = (\mathbb{R}^2, d_{\mathbf{W}})$  defined by

$$d_{\mathbf{W}}((x_1, y_1); (x_2, y_2)) = \sqrt{(x_1^3 - x_2^3)^2 + (y_1 - y_2)^2}$$

is Euclidean, but not all straight lines are shortest paths, nor are all shortest paths straight lines.

This example shows that a seemingly complicated formula can actually define a metric with exactly the same intrinsic properties than the Euclidean one. Stated more vividly, if we picture ourselves as ants living on a plane where the length of our travels (and other physical phenomena, such as course of light along shortest paths) would be defined by this metric, it would be impossible for us to make any difference with the metric  $d_{\text{euc}}$ .

To understand this example and check the stated properties, consider the one-to-one, onto mapping  $\varphi : \mathbb{R}^2 \rightarrow \mathbb{R}^2$  defined by  $\varphi(x, y) = (x^3, y)$ . For all  $p, q \in \mathbb{R}^2$ , we have  $d_{\text{euc}}(\varphi(p), \varphi(q)) = d_{\mathbf{W}}(p, q)$ . This proves at once that  $d_{\mathbf{W}}$  is a metric (all three axioms follow by applying this formula and using that  $d_{\text{euc}}$  is a metric), and that  $\varphi$  is an isometry from  $\mathbf{W}$  to  $\mathbf{E}^2$  (and its inverse mapping  $\varphi^{-1}$  is an isometry from the latter to the former). From this it follows that  $\mathbf{W}$  is isometric to  $\mathbf{E}^2$  and that shortest paths of  $d_{\mathbf{W}}$  are the images of shortest paths of  $\mathbf{E}^2$  by  $\varphi^{-1}$ .

Let us work out the shortest path in  $\mathbf{W}$  from  $p = (0, 0)$  to  $q = (1, 1)$ . It is obtained by first taking the images of these points by  $\varphi$  in  $\mathbf{E}^2$ , which turn out to be themselves:  $(0, 0)$  and  $(1, 1)$ ; then considering the shortest path in  $\mathbf{E}^2$  between these images, given by  $\gamma_t = (t, t)$ ,  $t \in [0, 1]$ ; and taking the image by  $\varphi^{-1}$  of this path:  $\varphi^{-1}(\gamma_t) = (t^{\frac{1}{3}}, t)$ . This is not a straight line segment, as it goes through  $(1/2, 1/8)$ . Since the shortest path between any two points of  $\mathbf{E}^2$  is unique and this is an intrinsic property,  $\mathbf{W}$  shares it. The straight line segment between  $p$  and  $q$  is thus not a shortest path of  $\mathbf{W}$ . The shortest paths issued from  $(0, 0)$  and from  $(1, 0.5)$  are shown in Figure 3.

In this example, some straight lines are sent to straight lines by  $\varphi$ , so that some shortest paths of  $\mathbf{W}$  are straight lines. It is not difficult to devise a different map

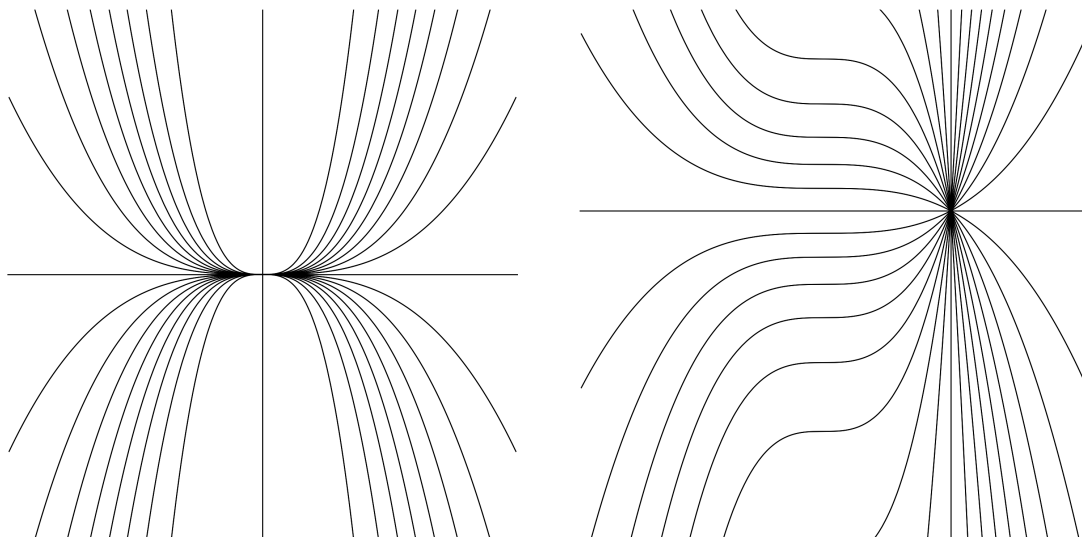


Figure 3: Shortest paths of the intrinsically Euclidean metric  $d_{\mathbf{W}}$  issued from the points  $(0, 0)$  (left) and  $(1, 0.5)$  (right).

$\varphi$  such that the metric on  $\mathbb{R}^2$  constructed to make  $\varphi$  an isometry has *no* shortest path that is also a straight line, but the formulas are more intricate and not very enlightening.

In view of this example, we would like to discuss the following quote:

*The second set of error, which is the most widespread, involves a confusion between the Euclidean straight line and the minimum path. The errors consist in considering the presence of a detour as a violation of the TI, while this situation simply corresponds to a non-Euclidean distance.*

l'Hostis 2016 [l'H16]

There is no error since l'Hostis most certainly used “non-Euclidean” to mean “not equal to  $d_{\text{euc}}$ ”. However, the above example shows that *intrinsically* Euclidean metrics can also cause detour: it can be that we observe that the shortest paths are not straight lines, causing apparent detour, but that the metric could still be represented perfectly in the plane, through a well-chosen change of coordinates. Arguably, this could be a very rare occurrence in geography; but it shows that to understand the core matter, we have to separate the intrinsic properties and the properties that depend on a particular representation of space, be it the way it lies before our eyes.

We can use the mapping  $\psi$  of Subsection 2.4 to make an amusing “fake counter-example”.

**Proposition 4.9.** *The metric space  $\mathbf{V} = (V, d_{\mathbf{V}})$  where  $V \subset \mathbb{R}^2$  is defined by the*

inequality  $(x + y > -1)$  and

$$d_{\mathbf{V}}((x_1, y_1); (x_2, y_2)) = \sqrt{\left(\frac{x_1}{x_1 + y_1 + 1} - \frac{x_2}{x_2 + y_2 + 1}\right)^2 + \left(\frac{y_1}{x_1 + y_1 + 1} - \frac{y_2}{x_2 + y_2 + 1}\right)^2}$$

is isometric to a convex open subset of the Euclidean plane and its shortest paths are straight lines.

The proofs follows exactly the same lines than the previous one, but instead of  $\varphi$  we use the map  $\psi$  defined at the end of Section 2.4, which sends all straight lines to straight lines. Then shortest paths are straight lines despite the complicated formula. Note that *constant-speed* geodesics do not have the usual parametrization  $(at, bt)$ .

The metric space  $\mathbf{V}$  is only isometric to *a part* of  $\mathbf{E}^2$ , though. To obtain a space isometric to the whole of  $\mathbf{E}^2$  and with straight lines as shortest paths, by the Fundamental Theorem of Affine Geometry one has no choice but use an affine transformation for  $\varphi$ , leading to a metric formula of the form

$$d((x_1, y_1); (x_2, y_2)) = \sqrt{(\alpha(x_1 - x_2))^2 + \beta(x_1 - x_2)(y_1 - y_2) + \gamma(y_1 - y_2)^2} \quad (3)$$

easily identified as Euclidean, only with a linear change of coordinates. For example, with  $\beta = 0$  and  $\alpha = \gamma$ , we obtain a rescaled version of the Euclidean distance (zoomed out when  $\alpha = \gamma > 1$ , zoomed in when  $\alpha = \gamma < 1$ ).

### 4.3 Non-Euclidean metrics with straight lines as shortest paths

We shall now turn to second implication in Claim B, namely “A space with straight lines as shortest paths must be Euclidean”; for “straight lines” to make sense, we again consider metrics defined on  $\mathbb{R}^2$  or on one of its open subsets. Let us consider two relevant quotes.

*Our most familiar concept of distance is of course Euclidean (or straight-line) distance between points, which represents the shortest-path distance in Euclidean space.*

Huriot, Smith & Thisse 1989 [HST89]

*The standard or default assumption in the Euclidean model is that the length-metric is the straight-line segment between any two locations; this corresponds to the Euclidean distance between the pair.*

Miller & Wentz 2003 [MW03]

We find in both quotes a conflation between the metric, or equivalently the collection of distances between every pair of points, and the shape of shortest paths. It is true that these notions are related since (assuming the metric to be geodesic) the distance is the length of a shortest path; but this relation is not as strong in the other direction: the length of a given path is only defined once the metric is specified, hence two different metrics can in principle have the same shortest paths, but assign to them different lengths. We will now see a number of examples showing how this principle can be actually instantiated.

### 4.3.1 Normed spaces

Our first examples come in a large family. For each  $p \in [1, +\infty)$ , the  $\ell_p$  norm on  $\mathbb{R}^2$  is defined by

$$\|[x; y]\|_p = \left(|x|^p + |y|^p\right)^{\frac{1}{p}}$$

and the corresponding distance by

$$d_p(A, B) = \|\overrightarrow{AB}\|_p \quad \text{i.e.} \quad d_p((x_1, y_1), (x_2, y_2)) = \left(|x_1 - x_2|^p + |y_1 - y_2|^p\right)^{\frac{1}{p}}$$

(here the customary notation  $p$  for the exponent makes us use other letters for points). Of the three metric axioms, only the triangle inequality is any trouble to check; it follows from the following classical statement.

**Proposition 4.10** (Minkowski's inequality). *For all  $p \geq 1$  and all  $\vec{u}, \vec{v} \in \mathbb{R}^2$ , we have*

$$\|\vec{u} + \vec{v}\|_p \leq \|\vec{u}\|_p + \|\vec{v}\|_p.$$

*Moreover, equality happens in the following cases, and the following cases only:*

- *when  $p = 1$  and in the coordinates  $[x; y]$  of  $\vec{u}$  and  $[x'; y']$  of  $\vec{v}$ ,  $x$  and  $x'$  have the same sign and  $y$  and  $y'$  have the same sign.*
- *when  $p > 1$ , if  $\vec{u}$  and  $\vec{v}$  are positively colinear, i.e. one of them is zero or there exist  $\lambda > 0$  such that  $\vec{u} = \lambda\vec{v}$ .*

Here, “sign” is to be understood in the broadest sense, i.e. 0 has the same sign as any real number.

**Example 4.11.** For any  $p \in [1, +\infty)$ , consider the metric space  $\mathbf{L}_p = (\mathbb{R}^2, d_p)$ . For all  $p \geq 1$ , all straight lines are shortest paths, and if  $p > 1$  then conversely all shortest paths are straight lines. For all  $p \neq 2$ ,  $\mathbf{L}_p$  is not intrinsically Euclidean, not even locally, and not isometric to any subset of the Euclidean plane.

That straight lines are shortest paths follows from the characterization 2.10 and the case of equality in Minkowski's inequality (case  $p = 1$  treated in more detail below).

When  $p > 1$ , that there exist no other shortest path than straight lines follows from the more restricted case of equality in Minkowski's inequality (if vectors  $\overrightarrow{AB}$  and  $\overrightarrow{BC}$  are positively colinear, then the points  $A, B, C$  lie on a line in this order).

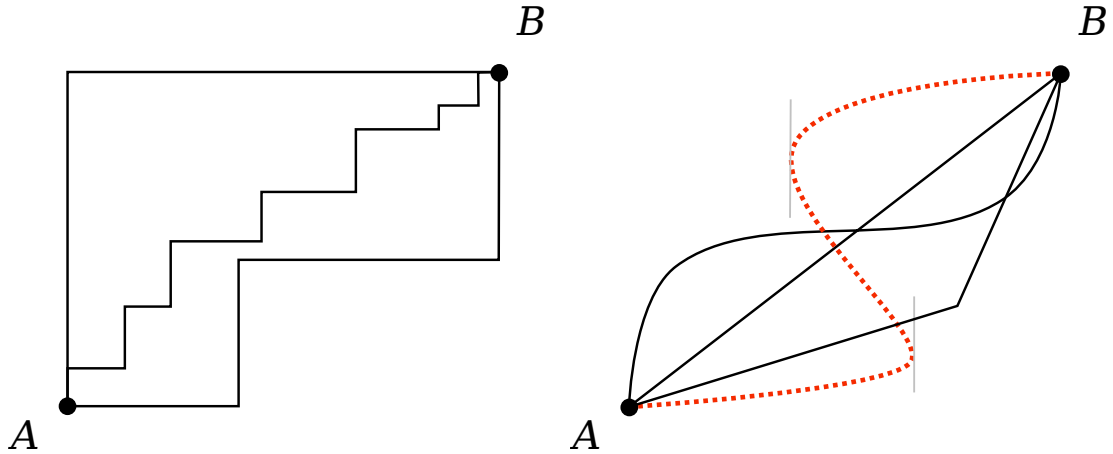


Figure 4: Some geodesics of the Manhattan metric  $d_1$ . Left: the piecewise horizontal/vertical geodesics giving the metric its name. Right: other geodesics (solid black) including the straight line segment, and a non-geodesic curve (dotted red). It cannot be stressed enough: all black curves have the same length according to  $d_1$ .

That when  $p \neq 2$ ,  $\mathbf{L}_p$  is not isometric to a subset of  $\mathbf{E}^2$  (in particular not Euclidean) can be checked in a variety of ways; one possibility is to observe that the quadrangle with vertices  $(0, 0)$ ,  $(1, 0)$ ,  $(1, 1)$ ,  $(0, 1)$  has all sides of length 1 and both diagonals of length  $2^{\frac{1}{p}}$ . In the Euclidean plane, a quadrangle with unit sides and with equal diagonals must be a square, thus have diagonal length  $\sqrt{2}$ .

Actually, these examples could be vastly generalized to the family of “norms” on  $\mathbb{R}^2$  (the additional property that only straight lines are geodesics is enjoyed when the norm is “strictly convex”); but we prefer to delve deeper into the prominent case of  $d_1$ .

### 4.3.2 Geodesics of the Manhattan metric

In the case  $p = 1$ , the name “Manhattan” metric evokes the picture of polylines with vertical and horizontal segments (Figure 4, left), but there are many others geodesics,<sup>3</sup> including straight lines. Let us check this precisely when  $p = 1$  in a specific case, say for the points  $A = (0, 0)$  and  $B = (1, 2)$ . The straight line segment  $[A, B]$  can be parametrized by  $\gamma_t = (t, 2t)$  for  $t \in [0, 1]$ , and for all  $t < u < v \in [0, 1]$  we get

$$\begin{aligned}
 d_1(\gamma_t, \gamma_u) + d_1(\gamma_u, \gamma_v) &= |u - t| + |2u - 2t| + |v - u| + |2v - 2u| \\
 &= 3u - 3t + 3v - 3u \\
 &= 3v - 3t \\
 &= d_1(\gamma_t, \gamma_v)
 \end{aligned}$$

proving that  $\gamma$  is a geodesic. Actually, any curve each of whose coordinate is either nondecreasing or nonincreasing a function of the parameter is a geodesic, as can be

<sup>3</sup>Unless  $A$  and  $B$  are on the same vertical or horizontal line, in which case the straight line segment is the only geodesic.



shown in the same way as above (monotonicity allowing to remove the absolute parts), see Figure 4, right.

In the light of these facts, the following quotes are interesting.

*Other distance metrics are possible, since interaction along straight-line paths is the exception rather than the rule. For example, distance metrics implied by actual travel distance at urban and regional scales are typically  $1 \leq p \leq 2$ .*

Miller & Wentz 2003 [MW03]

Since the metrics  $d_p$  do have straight lines as shortest paths even for  $p \neq 2$ , their difference with the Euclidean metric cannot be explained by a lack of “interaction along straight-line paths” in any obvious way.

*Another important special case is  $p = 1$ , or the “Manhattan metric,” where shortest paths are polylines with segments parallel to one of the axes*

Miller & Wentz 2003 [MW03]

Here we see that the many other shortest paths of this metric seem to have been neglected; one can guess why by analyzing the two quotes together.

The “actual travel distances” above certainly refers to the Euclidean distance in formula (1) to define the length of paths, but with travel constrained by the shape of the transportation network to follow certain paths, such as roads aligned to form a grid for  $p = 1$ . And indeed, if one computes length of paths using the Euclidean metric but constrains travel to follow polylines with segments parallel to the axes, the resulting length metric is  $d_1$ . However, if one considers  $d_1$  itself, without reference to such constrained travel, then straight lines *also* are shortest paths in this metric, as we have mentioned above: the constraint by which the metric was defined is not actually carried in the metric itself and the notion of shortest path.

For  $1 < p < 2$  a similar interpretation can be made: a network of roads that are not strictly aligned on a grid, but cover in each point a limited, non-uniformly distributed range of directions, could result in a metric close to  $d_p$  at large scale. But remember that the only shortest paths for  $d_p$  are the straight line: looking at the metric alone becomes misleading with regard to paths actually followed during travels, since the constrained paths giving birth to the metric are not shortest paths for it!

These examples are thus quite important in practice: a metric that fits well the data of time travel could be a poor predictor of the *shape* of travels. This would certainly deserve a deeper investigation.

### 4.3.3 Beltrami’s examples

Among metrics, a family particularly cherished by differential geometers are the *Riemannian* metrics; we shall not give the formal definition, but let us try to convey its core idea: Riemannian metrics are the metrics that, when zooming in at any point of

the space, converge to the usual Euclidean metric up to a linear change of coordinates. Important cases are that of length metrics induced on differentiable surfaces of  $\mathbf{E}^3$  and, in geography, the case of a Euclidean domain with a *velocity field* [AH72].<sup>4</sup>

To better understand this concept, it is useful to look at examples of *non*-Riemannian metrics; we choose the  $\mathbf{L}_p$  spaces of the previous section for  $p \neq 2$ . Indeed, if we consider a first point  $A = (x_1, y_1)$  and a second point very close to the first one, thought of as a perturbation of  $A$ , say  $A_\varepsilon = (x_1 + \varepsilon x', y_1 + \varepsilon y')$  where  $[x'; y']$  are the coordinate of a fixed vector giving the direction of  $\overrightarrow{AA_\varepsilon}$  and  $1 \gg \varepsilon > 0$  giving its magnitude, we get  $d_p(A, A_\varepsilon) = (|\varepsilon x'|^p + |\varepsilon y'|^p)^{\frac{1}{p}} = \varepsilon \|[x'; y']\|_p$ . Rescaling by the magnitude of the perturbation thus yields  $\frac{1}{\varepsilon} d_p(A, A_\varepsilon) = \|[x'; y']\|_p$ : zooming in at any point does not change the non-Euclidean character of the metric. The informal definition above can be slightly precised in the following way: a metric on a Domain of  $\mathbb{R}^2$  is *Riemannian* when for all point  $(x, y)$  and all vector  $[x'; y']$  we have

$$\lim_{\varepsilon \rightarrow 0} \frac{1}{\varepsilon} d((x, y), (x + \varepsilon x', y + \varepsilon y')) = \sqrt{\alpha_{x,y} x'^2 + \beta_{x,y} x' y' + \gamma_{x,y} y'^2}$$

(one recognizes the distance to the origin, after a linear change of coordinates, in  $d_{\text{euc}}$ , see formula (3)). Then the three functions  $\alpha_{x,y}, \beta_{x,y}, \gamma_{x,y}$  can be used to recover the distance; the usual formal definition actually gives them the first role.

Note that being Riemannian in this sense is *not* an intrinsic property; being isometric to a Riemannian space of course is intrinsic; and any *differentiable* isometry between metrics defined on domains of  $\mathbb{R}^2$  (or on surfaces embedded in  $\mathbb{R}^3$ ) preserves the property of being Riemannian.

One could expect that a *Riemannian* space with straight lines as shortest paths must be Euclidean. It is known at least since the XIX<sup>th</sup> century that this is not the case; the following counter-example is in particular well-known in cartography.

**Example 4.12** (gnomonic projection). Consider the space  $\mathbf{G} = (\mathbb{R}^2, d_{\mathbf{G}})$  where

$$d_{\mathbf{G}}((x_1, y_1), (x_2, y_2)) = \arccos \left( \frac{x_1 x_2 + y_1 y_2 + 1}{\sqrt{x_1^2 + y_1^2 + 1} \sqrt{x_2^2 + y_2^2 + 1}} \right).$$

Then  $d_{\mathbf{G}}$  is a geodesic, Riemannian metric, its shortest paths are exactly the straight lines, and it is not intrinsically Euclidean, not even locally.

The formula of  $d_{\mathbf{G}}$  is not very enlightening, and in fact the proof of the stated facts consists in explaining how it is constructed:  $d_{\mathbf{G}}$  simply translates usual length metric on a unit hemisphere through a gnomonic projection.

Let us give some details. Let  $S^+ = \{(x, y, z) \in \mathbb{R}^3 \mid x^2 + y^2 + z^2 = 1 \text{ and } z > 0\}$  be the upper unit hemisphere and let  $d_\ell$  be the restriction to  $S^+$  of the usual spherical metric,

---

<sup>4</sup>The celebrated *Uniformization Theorem* implies that any surface can be represented in the plane with a suitable velocity field, at least locally; however the converse problem is still open for infinitely differentiable velocity fields, and has counter-examples of limited regularity [Gho17].

i.e.  $d_\ell(p, q)$  is the angle, in radian, between the vectors  $\vec{op}$  and  $\vec{oq}$  where  $o = (0, 0, 0)$  is the origin. Consider the mapping  $g : S^+ \rightarrow \mathbb{R}^2$  defined by  $g(x, y, z) = \begin{pmatrix} x/z \\ y/z \end{pmatrix}$ . Then a simple computation, using that for all  $(x, y)$  there must be some  $\lambda > 0$  such that  $g^{-1}(x, y) = (\lambda x, \lambda y, \lambda)$ , shows that  $d_{\mathbf{G}}(p, q) = d_\ell(g^{-1}(p), g^{-1}(q))$ .

That  $d_{\mathbf{G}}$  is Riemannian then follows from the fact that  $d_\ell$  is Riemannian and  $g$  is differentiable. The fact that it is not intrinsically locally Euclidean follows from the fact that it is isometric to  $d_\ell$ . A low-tech proof showing that  $\mathbf{G}$  it is not isometric to any subset of the Euclidean plane consists in exhibiting configurations of four points whose 6 pairwise distances are impossible to realize in the Euclidean plane. This is for example the case with spherical “squares”, i.e. quadrilaterals in  $S^+$  with equal sides and equal diagonals. If  $\ell$  is the side length, the diagonals have length  $2 \arccos \sqrt{\cos \ell} > \sqrt{2}\ell$ .

Last, the fact that shortest paths are straight lines is easily understood once one observes that the gnomonic projection  $g$  is the projection from the center of the sphere to the plane of equation  $(z = 1)$ , up to the obvious identification between this plane and  $\mathbb{R}^2$ . We know that shortest paths of  $d_\ell$  are the arcs of great circles, i.e. the intersections of  $S^+$  with planes containing the origin. The projection from the origin preserves such planes, so that the image by  $g$  of a great circle is the intersection of two planes, i.e. a straight line.

**Example 4.13.** Consider the space  $\mathbf{H} = (\mathbb{D}, d_{\mathbf{H}})$  where  $\mathbb{D} = \{(x, y) \in \mathbb{R}^2 \mid x^2 + y^2 < 1\}$  is the unit disc and, given points  $p, q \in \mathbb{D}$  and denoting by  $a, b$  the points of intersection of the line  $(pq)$  and the boundary of  $\mathbb{D}$ , such that  $a, p, q, b$  are aligned in this order,

$$d_{\mathbf{H}}(p, q) = \frac{1}{2} \ln \frac{d_{\text{euc}}(a, q)d_{\text{euc}}(p, b)}{d_{\text{euc}}(a, p)d_{\text{euc}}(q, b)}. \quad (4)$$

Then  $d_{\mathbf{H}}$  is a geodesic, Riemannian metric, its shortest paths are exactly the straight lines segments, and it is not intrinsically Euclidean, not even locally.

It is a non-obvious but well-known fact that this somewhat mysterious formula defines a metric. Lifting its mystery would need to enter the realm of *projective geometry* and its *cross-ratio*; while it is far beyond the scope of this article to do so, we shall see a generalization below. The space  $\mathbf{H}$  is in fact isometric to the *hyperbolic plane*, which has constant negative curvature; this particular model is known as the *Beltrami-Klein disc*.

We have seen that at least three kind of Riemannian metrics, spherical, hyperbolic and of course Euclidean can be represented in the plane in a way that makes straight lines geodesics. These three spaces are also known as the only 2-dimensional spaces of constant curvature (up to “covering”). It is a theorem of Beltrami [Bel66] that in fact, this list is essentially exhaustive: a Riemannian metric on a domain of  $\mathbb{R}^2$  for which straight lines are geodesics must have constant curvature. The following quote is thus misleading:

*Of course we know from Beltrami that only surfaces of zero curvature can be mapped in a geodesic fashion onto a plane.*

Tobler 1993 [Tob93]

Either Tobler made a confusion between “constant curvature” and “zero curvature”, an odd mistake for a man well-acquainted with the gnomonic projection, or he had in mind a more restrictive meaning for “geodesic mapping” (either distance-preserving, but then the theorem should not be attributed to Beltrami, or mapping constant-speed geodesics to constant-speed geodesic, which would be an uncommon definition). Let us compare with the following less ambiguous quote:

*The shortest paths on any map which has been produced by a projection of a spherical surface into the plane cannot be straight lines. In fact, since the Earth is spherical, it is impossible to construct a flat map of the world which correctly represents distances.*

Angel & Hyman 1976 [AH76]

Here the mistake of forgetting the gnomonic projection is clear in the first sentence, and we can observe in the second sentence the origin of the mistake: the non-existence of distance-preserving mapping from a spherical surface into the plane (which is correct) has been confused with the non-existence of geodesic mappings (which is not). The shape of shortest paths carries in fact less information than the values of the distances.

#### 4.3.4 Hilbert geometries

We would like to end this section by mentioning a family of metric space of much elegance, which is a kind of bridge between the normed spaces and the Beltrami-Klein model of the hyperbolic plane.

**Example 4.14.** Let  $C \subset \mathbb{R}^2$  be any bounded open convex set. Given points  $p, q$ , we denote by  $a, b$  the points of intersection of the line  $(p, q)$  with the boundary of  $\mathbb{D}$  such that  $a, p, q, b$  are aligned in this order, and we define the *Hilbert metric* by

$$d_C(p, q) = \frac{1}{2} \ln \frac{d_{\text{euc}}(a, q)d_{\text{euc}}(p, b)}{d_{\text{euc}}(a, p)d_{\text{euc}}(q, b)}. \quad (5)$$

Then  $(C, d_C)$  is a geodesic metric and its shortest paths are exactly the straight lines segments, and it is not intrinsically Euclidean, not even locally.

When  $C$  is an ellipse, it is relatively easy to see that  $d_C$  is Riemannian and  $(C, d_C)$  is isometric to the hyperbolic plane (when  $C$  is a disc, we get  $\mathbf{H}$ , and it can be shown that an affine transformation between convex sets  $C$  and  $C'$  is an isometry of their hilbert metrics). If  $C$  is not an ellipse, then  $d_C$  is not Riemannian but “Finslerian”, i.e. zooming in to a point it converges to a norm. The shape of  $C$  has a huge influence on the geometric properties of  $d_C$ , and there are many mathematical works exploring these geometries.

## 5 Conclusion

We have seen through counter-examples to widely claimed statements that the different aspects of the geometry of a metric space need to be meticulously articulated to avoid any confusion.

First, the Triangular Inequality is consubstantial to the notion of a metric, without regards to its Euclideaness, to the shape of its shortest paths, or even to the existence of shortest paths.

Second, while the shape of shortest path (when they exist) is determined by the metric, the converse is not true: different metric can have the same shortest paths. The shape of shortest path thus only gives limited information on the metric; in particular for a metric on a region of the plane, having straight lines as shortest paths does not make it necessarily Euclidean.

This last statement is well illustrated geographically by the gnomonic projection; while geographers cannot be suspected of not knowing that projection, it might not be well understood that each projection defines a metric: the metric which, to any two points of the range of the projection, associates the distance along the sphere between the points they represent. This construction of *pushing forward* the spherical metric by a projection to the plane can be generalized, and can be used to define *intrinsic* geometric properties: properties that are not changed under such a push-forward.

While it makes a great deal of sense to define Euclideaness as an intrinsic property, having straight lines as shortest paths is not a intrinsic property, but is a particular relation between the given metric and the *affine* structure of the plane. One can thus build Euclidean metric with shortest paths different from straight lines. While these construction may feel artificial, they show that non-Euclideaness cannot be seen from the shape of shortest paths.

The mathematical subfield of geometry provides powerful tools to think geographical spaces; but it is easy to underestimate its subtleties. It thus seems that a stronger co-operation between geographers and mathematicians is called for; I hope that this article will be a small but significant contribution to bridge the gap between our communities.

**Acknowledgment** I warmly thank Alain l’Hostis for introducing me to geography and for his encouragement and many relevant comments; Thomas Richard for introducing me to Alain l’Hostis; both of them for our previous collaboration that gave me much to think about; and Françoise Bahoken for inviting me to the Tribute To Tobler workshop, a great opportunity to learn more about geography.

## References

- [AH72] S. ANGEL & G. M. HYMAN – “Transformations and geographic theory”, *Geographical Analysis* 4 (1972), no. 4, p. 350–367. [1.1](#), [4.3.3](#)
- [AH76] — , *Urban fields*, Pion limited, 1976. [4.3.3](#)

- [Bel66] E. BELTRAMI – “Résolution du problème de reporter les points d’une surface sur un plan, de manière que les lignes géodésiques soient représentée par des lignes droites”, *Annali di Matematica* (1866). 4.3.3
- [Bun62] W. BUNGE – *Theoretical geography.*, Lund studies in geography, Gleerup, 1962. 2.4.1
- [Gho17] M. GHOMI – “Open problems in geometry of curves and surfaces”, Available at: [people.math.gatech.edu/~ghomi/Papers/op.pdf](http://people.math.gatech.edu/~ghomi/Papers/op.pdf), 2017. 4
- [HST89] J.-M. HURIOT, T. E. SMITH & J.-F. THISSE – “Minimum-cost distances in spatial analysis”, *Geographical Analysis* **21** (1989), no. 4, p. 294–315. 4.3
- [KLR21] B. R. KLOECKNER, A. L’HOSTIS & T. RICHARD – “Contextual metrics: A mathematical definition for a comprehensive approach of geographical distances”, *Geographical Analysis* **53** (2021), no. 4, p. 736–766. 1
- [Lam98] M. LAMURE – “Proximité(s), voisinage et distance”, in *Approches multiformes de la proximité* (M. Bellet, T. Kirat & C. Largeron, eds.), Hermès, 1998, p. 9–12. 3
- [l’H16] A. L’HOSTIS – “Misunderstanding geographical distances: two errors and an issue in the interpretation of violations of triangle inequality”, *Cybergeog: European Journal of Geography* (2016). 1, 3
- [l’H17] —, “Detour and break optimising distance, a new perspective on transport and urbanism”, *Environment and Planning B: Urban Analytics and City Science* **44** (2017), no. 3, p. 441–463. 2.4.1
- [l’H20] —, “All geographical distances are optimal”, *Cybergeog: European Journal of Geography* (2020). 2.3, 3
- [Lé09] J. LÉVY – “Entre contact et écart: la distance au coeur de la réflexion”, *Atala* **12** (2009), p. 175–185. 3
- [MW03] H. J. MILLER & E. A. WENTZ – “Representation and spatial analysis in geographic information systems”, *Annals of the Association of American Geographers* **93** (2003), no. 3, p. 574–594. 4.3, 4.3.2
- [Pon08] P. PONCET – “Visions du monde”, in *L’invention du Monde* (J. Lévy, éd.), Presses de Sciences Po, 2008, p. 81–93. 3
- [Tob61] W. R. TOBLER – “Map transformations of geographic space”, Thèse, 1961. 2.4.1
- [Tob93] W. TOBLER – “Three presentations on geographical analysis and modeling”, National Center For Geographic Information And Analysis, Technical Report 93-1, 1993. 4.3.3