Introduction: Some historical markers

The aim of this introduction is threefold. First, we give a brief account of a paper by Jacques Hadamard which is at the genesis of many works on global properties of surfaces of nonpositive Gaussian curvature. Then, we make a brief account of several different notions of nonpositive curvature in metric spaces, each of which generalizes in its own manner the notion of nonpositive Gaussian curvature for surfaces or of nonpositive sectional curvature for Riemannian manifolds. In particular, we mention important works by Menger, Busemann and Alexandrov in this domain. Finally, we review some of the connections between convexity theory and the theory of nonpositive curvature, which constitute the main theme that we develop in this book.

The work of Hadamard

The theory of spaces of nonpositive curvature has a long and interesting history, and it is good to look at its sources. We start with a brief review of the pioneering paper by Hadamard, "Les surfaces à courbures opposées et leurs lignes géodésiques" [62]. This paper, which was written at the end of the nineteenth century, can be considered as the foundational paper for the study of global properties of nonpositively curved spaces.

Hadamard was one of the first mathematicians (and may have been the first) who strongly emphasized the importance of topological methods in the study of spaces of nonpositive curvature. In the introduction to the paper cited above, he writes : "The only theory which has to be studied profoundly, as a basis to the current work, is the *Analysis situs*, which, as one can expect after reading the work of Poincaré, plays an essential role in everything that will follow."

Let us look more closely at the content of that paper.

Hadamard considers a surface S equipped with a Riemannian metric of nonpositive curvature. The surface is smoothly embedded in $\mathbb{R}^3$, the Riemannian metric is induced from that inclusion and the set of points at which the curvature is zero is finite.

Hadamard starts by noting that this surface is necessarily unbounded since if one of the coordinates assumes a maximum or a minimum, then, in a neighborhood of that point, the surface would be situated at one side of its tangent plane and therefore the curvature would be positive. He then proves that such a surface can always be decomposed into the union of a compact region and of a collection of *infinite sheets* (we are translating the term "nappes infinies" used by Hadamard). Each infinite sheet is homeomorphic to a cylinder, and is connected to the compact region along a boundary curve. Hadamard constructs examples of surfaces with nonpositive curvature having an arbitrary number of infinite sheets. He classifies the infinite sheets into two types: the *flared* infinite sheets (“nappes infinies évasée”), which we call “funnels”, and the *unflared* infinite sheets. An unflared infinite sheet is characterized by the fact that one can continuously push to infinity a homotopically non-trivial closed curve on such a sheet while keeping its length bounded from above. To each unflared infinite sheet, Hadamard associates an *asymptotic direction*. The simple closed curves that connect the funnels to the compact region can be taken to be closed geodesics that are pairwise disjoint.

Let us now suppose that the fundamental group of S is finitely generated (Hadamard says that S has “finite connectivity”). For such a surface, Hadamard describes a finite collection of homotopically non-trivial and pairwise non-homotopic simple closed curves $C_1, \dots, C_n$ such that any non-trivial homotopy class of closed curves on the surface can be uniquely represented by a finite word written in the letters $C_1, \dots, C_n, C_1^{-1}, \dots, C_n^{-1}$. Hadamard pushes forward this analysis to make it include not only representations of closed curves, but also representations of long segments of open infinite curves. In fact, with these elements, Hadamard initiates the approximation of long segments of bi-infinite geodesic lines by closed geodesics, and therefore he also initiates the theory of symbolic representation of bi-infinite geodesics on the surface, that is, the theory of representation of elements of the geodesic flow associated to S by bi-infinite words in the finite alphabet $\{C_1, \dots, C_n, C_1^{-1}, \dots, C_n^{-1}\}$. We note in passing that the theory of symbolic dynamics for the geodesic flow associated to a nonpositively curved surface was thoroughly developed about twenty years later by Marston Morse, and that Morse, in his paper [113], refers to Hadamard as his source of inspiration. Hadamard calls a closed curve a “contour”, and the curves $C_1, \dots, C_n$ the “elementary contours”. There are two sorts of closed curves in the collection $C_1, \dots, C_n$ which Hadamard considers: the curves that are the boundaries of the infinite sheets, and those that correspond to the genus of the surface. (These curves come in pairs and they correspond to the “holes” of the compact part of the surface S, as Hadamard calls them.) He notices that there is (up to circular permutation) one and only one relation in the representation he obtains that allows us to consider any of these contours as a product of the elementary contours, and therefore, by eliminating one of the contours, he obtains a symbolic representation that is unique up to circular permutation.

In the same paper, Hadamard proves that in each homotopy class of paths on S with fixed endpoints, there is a unique geodesic.[1] He also obtains an analogous result for the free homotopy classes of closed curves on S that are neither homotopic to a point nor to the core curve of a flared infinite sheet.

Hadamard then makes the fundamental observation that the distance function from a fixed point in S to a point that moves along a global geodesic in this surface is convex. This observation is at the basis of the definition of nonpositive curvature in the sense

[1] Here, the word geodesic is used, as in Riemannian geometry, in the sense of local geodesic. We warn the reader that this is not the definition that we are adopting in the rest of these notes, where a geodesic is a distance-minimizing map between its endpoints (and therefore it is a *global* geodesic).

of Busemann, which is the main topic of this book.

Next, Hadamard introduces the notion of *asymptotic geodesic rays*, and he constructs such rays as follows. Given a geodesic ray, he considers a sequence of geodesic segments joining a fixed point on the surface to a sequence of points on that ray that tends to infinity. He proves that this sequence of geodesic segments converges to a geodesic ray, which he calls *asymptotic* to the initial ray. He studies the asymptoticity relation and introduces through an analogous construction the notion of (local) geodesics that are asymptotic to a given closed geodesic.

Using an argument that is based on the convexity of the distance function from a point in a funnel to the closed geodesic that connects this funnel to the compact region in S, Hadamard proves that a geodesic that penetrates a funnel cannot get out of it.[2]

Finally, in the case where all the infinite sheets of S are funnels, Hadamard investigates the distribution of geodesics that stay in the compact part of the surface. These geodesics are the flowlines of the geodesic flow associated with the compact nonpositively curved surface with geodesic boundary, obtained from the surface S by deleting the funnels. For a given point x in S, Hadamard considers the set of initial directions of geodesic rays that start at x and stay in that compact surface. He proves that this is a perfect set with empty interior, whereas the set of directions of geodesic rays that tend to infinity is open.

We already mentioned that the setting of the paper [62] by Hadamard is that of a differentiable surface S embedded in $\mathbb{R}^3$, such that at each point of S, the Gaussian curvature is negative except for a finite set of points where it is zero. A theory of *metric spaces with nonpositive curvature*, that is, a theory that does not make any differentiability assumption and whose methods use the distance function alone, without the local coordinates provided by an embedding in Euclidean space or by another Riemannian metric structure, was developed several decades after the paper [62]. It is good to remember, in this respect, that the theory of metric spaces itself was developed long after the theory of spaces equipped with differentiable structures. For instance, Gauss's treatise on the differential geometry of surfaces [51], in which he defines (Gaussian) curvature and proves that this curvature depends only on the intrinsic geometry of the surface and not on its embedding in $\mathbb{R}^3$, was published in 1827,[3] whereas the axioms for metric spaces were set down by Fréchet, some 90 years later.

Of course, the setting of surfaces embedded in $\mathbb{R}^3$ has the advantage of providing visual characteristics for the sign of the curvature. For instance, it is well-known that for such a surface, we have the following:

- if the Gaussian curvature at some point is > 0, then the surface, in the neighborhood of that point, is situated on one side of the tangent plane;

- if the Gaussian curvature at some point is < 0, then the surface, at that point, crosses its tangent plane.

---

[2] Hadamard already obtained a similar result in [60].

[3] We recall that the surfaces considered by Gauss were always embedded in $\mathbb{R}^3$. It is only 30 years after Gauss's paper was written that Riemann introduced the concept of spaces equipped with (Riemannian) metrics with no embedding in $\mathbb{R}^3$ involved in the definition.

We also recall that if the Gaussian curvature at some point is 0, then any one of the above configurations can occur.

Hadamard indicated in the note [61] how to extend some of the results he proved for surfaces to higher dimensions. The development of these ideas in the general setting of Riemannian manifolds of nonpositive curvature has been carried out by Elie Cartan, in particular in his famous "Leçons sur la géométrie des espaces de Riemann" [36].[4]

The works of Menger and Wald

A few years after the introduction by Fréchet of the axioms for metric spaces, Karl Menger initiated a theory of geodesics in these spaces. A *geodesic* in a metric space is a path whose length is equal to the distance between its endpoints. Menger generalized several results of classical geometry to this new setting and he introduced new methods that did not make any use of local coordinates or of differentials, but only of equalities involving the distance function, and of the triangle inequality. Menger wrote several important papers involving this new notion of geodesic, and he also introduced a notion of "discrete geodesic", which is based on his definition of *betweenness*: a point z in a metric space is said to lie between two distinct points x and y if z is distinct from x and from y and if we have

$$d(x, y) = d(x, z) + d(z, y).$$

In a complete metric space, the existence of a geodesic joining any two points is equivalent to the existence, for any distinct points, of a point that lies between them. We refer the reader to the commented edition of Menger's papers in the volume [108], published on the one hundredth anniversary of Menger's birth. We shall have several opportunities to mention Menger's work in the following chapters, but here, we mention an important notion that he introduced, which we shall not consider further in this book. This notion contains an idea that is at the basis of the various definitions of curvature that make sense in general metric spaces. The idea is to construct "comparison configurations" for sets of points (say of finite cardinality) in a given metric space X. The comparison configurations are built in a model space, which is generally one of the complete simply connected surfaces M_κ of constant Gaussian curvature κ, that is, either the Euclidean plane (of curvature $\kappa = 0$), or a sphere of curvature $\kappa > 0$, or a hyperbolic plane of curvature $\kappa < 0$. The comparison configuration for a given subset $F \subset X$ is a subset $F^\star$ of M_κ, equipped with a map from F to $F^\star$, which is generally taken to be distance-preserving and which is called a "comparison map". Then, one can define notions like curvature at a given point x in X by requiring the *ad hoc* property for the comparison configurations associated to certain classes of subsets contained in a neighborhood of x. Of course, a comparison

[4]Note that the later editions of Cartan's book [36] contains additional material, in particular the development of the work of Hadamard to higher dimensions.

configuration does not always exist, but in the case where F is of cardinality 3, one can always construct a comparison configuration $F^\star$ of F in any one of the surfaces M_κ with $\kappa \leq 0$. In fact, the axioms for a metric space X are equivalent to the fact that one can construct a comparison configuration in the Euclidean plane for any triple of points in X.[5] Let us now consider an example.

Given three pairwise distinct points a, b and c in the Euclidean plane M_0, either they lie on a unique circle (the circumscribed circle), or they lie on a Euclidean straight line. It is useful to consider here such a line as being a circle of radius ∞, in order to avoid taking subcases. Now for any triple of pairwise distinct points in a metric space X, Menger defined its "curvature" as being equal to $1/R$, where R is the radius of a circle in the Euclidean plane which is circumscribed to a comparison configuration associated to that triple. With this definition, the three points in X are aligned (that is, they satisfy a degenerate triangle inequality) if and only if their curvature is zero. Given an arc (or, say, a one-dimensional object) A contained in X, Menger says that the curvature of A at a point $a \in A$ is equal to κ if for any triple of pairwise distinct points in A that are sufficiently close to a, the curvature of this triple is close to κ.

With this definition, Menger introduced the notion of *curvature* for one-dimensional objects in an arbitrary metric space, and he posed the problem of the definition of curvature for higher-dimensional objects.

Now, we must mention the work of Abraham Wald, a student of Menger, who introduced a notion of two-dimensional curvature in an arbitrary metric space.[6] The definition again uses a limiting process, but now it involves quadruples of points in that space. The problem is that in general, a quadruple of points in a metric space does not necessarily possess a comparison configuration in the Euclidean plane. However, Wald starts by proving that if the metric space X is a differentiable surface, then, for every point x in X, there exists a real number $\kappa(x)$ satisfying the following property:

($\star$) for each $\epsilon > 0$, there exists a neighborhood $V(x)$ of x such that for every quadruple of points Q in $V(x)$, there is an associated real number $\kappa(Q)$ satisfying $|\kappa(x) - \kappa(Q)| < \epsilon$ such that the quadruple of points Q possesses a comparison configuration in the model surface $M_{\kappa(Q)}$.

Wald then proves that the quantity $\kappa(x)$ is equal to the Gaussian curvature of the surface at the given point x.

Now let X be a metric space that is "Menger convex", that is, a metric space X in which for every pair of distinct points x and y, there exists a point z that lies between them. Suppose furthermore that the Wald two-dimensional curvature exists at each point of X. In other words, suppose that one can associate to each point x in X a real number $\kappa(x)$ satisfying property ($\star$) that is stated above. Under these assumptions, Wald shows that the space X has the structure of a differentiable surface embedded in $\mathbb{R}^3$ that induces the same *length structure* as that of the original metric on X. In

[5]We also mention that an early version of the idea of a comparison map is already contained in the very definition of the Gauss map.

[6]Wald gave this 2-dimensional curvature the name "surface curvature", and Menger refers to it as "Wald curvature".

other words, the lengths of an arbitrary curve in X, on the one hand measured using the original metric and on the other hand measured using this differentiable surface structure, coincide. Furthermore, at each point a in X, the quantity $\kappa(a)$ is equal to the Gaussian curvature induced by the differentiable embedding of the surface in $\mathbb{R}^3$. With this result, Wald solved a problem that had been posed by Menger in [105], which asked for a metric characterization of Gauss surfaces among Menger convex metric spaces. We refer the reader to the Comptes Rendus Note [141] by Wald, presented by Elie Cartan, which describes this work.

We note in passing that for extra-mathematical reasons, Wald stopped working on this subject soon after he published the solution to Menger's problem, and his research interests switched to statistics and econometry. The story is interesting and it is told by Menger in [109]. It seems that there was no direct continuation to Wald's work.

Now, after the notion of curvature in metric spaces, we pass to the notion of nonpositive curvature.

The works of Busemann and Alexandrov

For the development of the theory of nonpositively curved metric spaces, we shall consider works that have been carried out in two different directions: the works of H. Busemann and the works of A. D. Alexandrov and his collaborators. Both Busemann and Alexandrov started their works in the 1940s, and the two approaches gave rise to rich and fruitful developments, with no real interaction between the two. The ramifications of these two theories continue to grow today, especially since the rekindling of interest that was given to nonpositive curvature by M. Gromov in the 1970s.

Let us briefly describe the basic underlying ideas of these works. First we need to recall a few definitions. Consider a metric space X in which each point x possesses a neighborhood U such that any two points in U can be joined by a geodesic path in U. A metric space X with such a property is said to be a *locally geodesic* space. We say that such a neighborhood U is a *geodesically convex* neighborhood of x. A *geodesic segment* $[a,b]$ in X is, by definition, the image of a geodesic path in X joining a and b.[7] A *triangle* in U is the union of three geodesic segments $[a,b]$, $[a,c]$ and $[b,c]$ contained in U. The segments $[a,b]$, $[a,c]$ and $[b,c]$ are called the *sides* of this triangle.

We start by presenting Busemann's definition of nonpositive curvature, which has the advantage of being the simplest to describe, and on which we shall focus in these notes.

We say that the space X has *nonpositive curvature in the sense of Busemann* if every point x in X possesses a geodesically convex neighborhood U such that for any

[7] From now on, we use the term "geodesic" in the sense of "global geodesic", that is, a path whose length is equal to the distance between its endpoints.

geodesic triangle with sides $[a, b]$, $[a, c]$ and $[b, c]$ contained in U, we have

$$\mathrm{dist}(m, m') \leq (1/2)\mathrm{dist}(b, c),$$

where m and m' are respectively the midpoints of $[a, b]$ and $[a, c]$. This property can be stated in terms of a convexity property of the distance function, defined on the product of any two geodesic segments $[a, b] \times [a, c]$ in X equipped with their natural (barycentric) coordinates. We shall develop this point of view in later chapters.

A nonpositively curved space in the sense of Busemann is sometimes referred to as a "locally convex metric space", or "local Busemann metric space". The terminology "nonpositively curved space", in this sense, is due to Busemann.[8]

A complete Riemannian manifold of nonpositive sectional curvature is an example of a metric space of nonpositive curvature in the sense of Busemann.

The most important writings of Busemann on this subject are certainly the paper [26] and the book [28].

Now we consider the point of view of A. D. Alexandrov.

Before giving the definition of nonpositive curvature in the sense of Alexandrov, we must recall the notion of angle which two geodesic segments (or more generally two paths) in a metric space, that start at a common point, make at that point.

The notion of angle in a metric space has been introduced by Alexandrov as a generalization of the notion of angle in a surface. The first papers by A. D. Alexandrov deal with the intrinsic geometry of surfaces, and the notion of angle is certainly the most important notion in that theory, after the notion of distance and that of length of a path.[9] In fact, Alexandrov introduced several notions of angle, and these notions coincide provided some reasonable hypotheses on the ambient metric spaces are satisfied. For our needs, it suffices to consider the notion of *upper angle* that two geodesic segments with a common initial point make at that point.

Let X be a locally geodesic metric space, let U be a geodesically convex open subset of X and let us consider a geodesic triangle Δ in U, with sides $[a, b]$, $[a, c]$ and $[b, c]$. Let us define the *upper angle* α of the triangle Δ at the vertex a. For every point x on the segment $[a, b]$ and for every point y on the segment $[a, c]$, we consider a triangle $\Delta_{x,y}$ in X with sides $[a, x]$, $[a, y]$ and $[x, y]$, where $[a, x]$ and $[a, y]$ are subsegments of $[a, b]$ and $[a, c]$ respectively and where $[x, y]$ is a geodesic segment joining the points x and y. Let $\Delta^{\star}_{x,y}$ be an associated comparison triangle in the Euclidean plane and let $\alpha_{x,y}$ be the angle of $\Delta^{\star}_{x,y}$ at the vertex that corresponds to the vertex a of $\Delta_{x,y}$. Then, the *upper angle* of the triangle Δ at the vertex a is defined as

$$\alpha = \limsup_{x,y \to a} \alpha_{x,y}.$$

[8] Busemann has some additional hypotheses on the metric spaces that he considers. One such hypothesis is the uniqueness of the prolongation of geodesics. There are interesting examples of spaces that do not satisfy this hypothesis and for which the results that we are interested in are valid, and for that reason we have tried to avoid this hypothesis in this book. We recall the precise definition of the spaces considered by Busemann (which he calls G-spaces) in the Notes on Chapter 2 below.

[9] One of the basic papers of Alexandrov on that theory is [2]. We also refer the reader to [3] and [4].

The *angular excess* of the triangle Δ is then defined as

$$\delta(\Delta) = \alpha + \beta + \gamma - \pi,$$

where α, β and γ are the upper angles of Δ at the vertices a, b and c.

Finally, the space X is said to be *nonpositively curved* in the sense of Alexandrov if every point in X possesses a geodesically convex neighborhood U such that the angular excess of any triangle in U is ≤ 0.

Let us note that in the case where the space X is a differentiable surface, the angular excess $\delta(\Delta)$ is the classical *total curvature*, that is, the integral of the Gaussian curvature over the region $R \subset S$ which is bounded by the triangle Δ. Indeed, the Gauss-Bonnet theorem applied to a disk D with piecewise geodesic boundary embedded in a differentiable surface $S \subset \mathbb{R}^3$ says that

$$\tau + \omega = 2\pi\chi,$$

where $\omega = \int_D dS$ is the total curvature of the disk D, that is, the integral of the Gaussian curvature with respect to the area element of that disk. In the special case considered, χ is the Euler characteristic of the disk, which is equal to 1, and τ is the sum of the *rotations* of the boundary of the disk at the vertices. At a vertex whose angle is α, the rotation is equal to $\pi - \alpha$. This gives

$$\omega = 2\pi - (\pi - \alpha) - (\pi - \beta) - (\pi - \gamma) = \delta(\Delta).$$

Thus, the notion of nonpositive curvature in the sense of Alexandrov generalizes the classical notion of nonpositive curvature for differentiable surfaces.

Complete Riemannian manifolds with nonpositive sectional curvature are also examples of nonpositively curved metric spaces in the sense of Alexandrov.

It should be noted that a metric space which is nonpositively curved in the sense of Alexandrov is also nonpositively curved in the sense of Busemann, but that the converse is not true. For instance, any finite-dimensional normed vector space whose unit ball is strictly convex is nonpositively curved in the sense of Busemann, but if the norm of such a space is not associated to an inner product, then this space is not nonpositively curved in the sense of Alexandrov. Alexandrov mentions this example in [3], p. 197.

Let us note finally that the techniques that are used by Alexandrov in all his works rely heavily on the notion of angle in a metric space, whereas the techniques of Busemann seldom use this notion.

Convexity

To end this introduction, we would like to make a few comments on convexity theory in relation with nonpositive curvature, but before that, we mention a particular link between the study of convex polyhedra and that of the differential geometry of surfaces.

In his historical report [46], Werner Fenchel traces back the origin of Alexandrov's work on the intrinsic geometry of convex surfaces to some early work on convex polyhedra. He first recalls Cauchy's rigidity result of 1812 stating that if two combinatorially equivalent convex polyhedra in $\mathbb{R}^3$ have congruent corresponding faces, then the polyhedra are themselves congruent.[10] Fenchel then reports that Cauchy, in a note he made for the French Academy of Sciences, wrongly announced that this result on polyhedra immediately implies that there is no closed convex surface that admits isometric deformations, a result that had already been claimed, also with a false proof, by Newton, around the year 1770.[11] Still, the problem was posed, and the relation between the rigidity of convex polyhedra and the rigidity of convex surfaces was clear: by replacing the condition on the isometry between the faces by a local (infinitesimal) condition, one is naturally led to the problem of finding local conditions on two closed convex surfaces under which these surfaces are isometric. The list of mathematicians who worked on this problem includes the names of J. H. Jellett, H. Liebmann, H. Weyl, S. Cohn-Vossen, A. D. Alexandrov, A. V. Pogorelov and others. We refer the reader to the paper by Fenchel for this fascinating story.

Of course, there are also relations between convexity and nonnegative curvature. We mention as an example and without further comment the following result of Alexandrov and Pogorelov: if X is a length metric space homeomorphic to the 2-sphere, then X has nonnegative curvature if and only if X is isometric to a convex surface S in $\mathbb{R}^3$, and in this case the surface S is unique up to rigid motions of $\mathbb{R}^3$.

We turn back to the relation between convexity theory and the theory of spaces of nonpositive curvature. First of all, as we have already said, it had already been realized by Hadamard that the convexity of the distance function in a nonpositively curved space is responsible for many of the global properties of that space. We also mentioned that this idea has been extensively explored by Busemann, who defined nonpositive curvature precisely by a convexity property of the distance function, and who showed, using this new definition, that most of the important properties of a nonpositively curved Riemannian manifolds are valid in a setting which is much wider than that of Riemannian geometry. Secondly, many of the basic results in convexity theory have the flavour of nonpositive curvature, and we mention as an important class of examples the "local-implies-global" properties, such as the fact that a locally convex function is globally convex, or the fact that a local geodesic in a Busemann space (*i.e.* in a simply connected metric space that is nonpositively curved in the sense of Busemann) is a global geodesic, and there are many others.

The aim of the chapters that follow is to describe these facts in some detail.

[10] Cauchy, in his paper *Sur les polygones et les polyèdres*, cf. [37], traces back this work on polyhedra to Euclid. He says that this rigidity statement is contained in Definition 9, of Book XI of the *Elements*.

[11] Cauchy's arguments were corrected later on by H. Lebesgue among others.