Encounter with a Geometer, Part II

Marcel Berger

Editor's Note. *Part I of this article appeared in the February 1999 Notices. The article discusses Mikhael Gromov's extraordinary mathematics and its impact from the point of view of the author, Marcel Berger. It is partially based on three interviews of Gromov by Berger, and it first appeared in French in the* Gazette des Mathématiciens *in 1998, issues 76 and 77. It was translated into English by Ilan Vardi and adapted by the author. The resulting article is reproduced here with the permission of the* Gazette *and the author.*

Mikhael Gromov is professor of mathematics at the Institut des Hautes Études Scientifiques and, in addition, is the Jay Gould Professor of Mathematics at the Courant Institute of Mathematical Sciences, spending three months a year at Courant.

As the author said at the beginning of Part I, "The aim of this article is to communicate the work of Mikhael Gromov (MG) and its influence in almost all branches of contemporary mathematics and, with a leap of faith, of future mathematics. It is not meant to be a technical report, and, in order to make it accessible to a wide audience, I have made some difficult choices by highlighting only a few of the many subjects studied by MG. In this way, I can be more leisurely in my exposition and give full definitions, results, and even occasional hints of proofs."

The author's warning in Part I about bibliographical matters applies equally to Part II: "In order to shorten the text, I have omitted essential intermediate results of varying importance, and have therefore neglected to include numerous names and references. Although this practice might lead to some controversy, I hope to be forgiven for the choices."

Riemannian Geometry

Starting in the late 1970s, MG completely revolutionized Riemannian geometry. I mention in this section some results that reflect my taste. This article contains only a small bibliography; further references may be found in my 1998 survey article in *Jahresbericht der Deutschen Mathematiker-Vereinigung.*[1] Except for obvious cases, every Riemannian manifold will be compact; in any case it will always be assumed complete. In (M, g) the letter M stands for the manifold and the letter g its Riemannian metric. This by definition means that at every point m of M there is an inner-product structure $g(\cdot, \cdot)$ on the tangent space T_mM at this point. We begin by describing the various notions of "curvature".

Marcel Berger is emeritus director of research at the Centre National de la Recherche Scientifique (CNRS) and was director of the Institut des Hautes Études Scientifiques (IHÉS) from 1985 to 1994. His e-mail address is `berger@ihes.fr`.

The author expresses his immense debt to Ilan Vardi and Anthony Knapp—to Vardi for translating the article into English and for improving the clarity of the mathematical exposition, and to Knapp for editing the article into its current form.

[1]*Riemannian geometry during the second half of the twentieth century,* Jahresbericht ***100*** *(1998), 45–208; reprinted with the same title as volume 17 of the University Lecture Series, Amer. Math. Soc., Providence, RI, 2000, ISBN 0-8218-2052-4.*

The curvature tensor is the basic invariant of a Riemannian manifold. Some of its power comes from the fact it has three equivalent definitions. Two of these are in terms of the associated "Levi-Civita connection".

Informally a "connection" on a smooth manifold is a way of computing directional derivatives of vector fields. These directional derivatives, which do not exist in general, will be called "covariant". More precisely, a *connection* is an operator D that assigns to each pair of vector fields x and y on M a vector field D_xy on M, the *covariant derivative* of y with respect to x, in a fashion that is $\mathbb{R}$ linear in y, is $C^\infty(M)$ linear in x, and satisfies $D_x(fy) = x(f)Y + fD_xy$ for all $f \in C^\infty(M)$. The vector $(D_xy)_m$ at a point $m \in M$ depends only on x_m and the values of y on any curve whose velocity vector at m equals x_m. Consequently it is meaningful to speak of a vector field on a curve that is "parallel" along the curve: If σ is the curve and u is its tangent, then a vector field y on σ is *parallel* along σ if $D_uy = 0$ on σ. If σ has domain $[a, b]$, one knows that for each $y \in M_{\sigma(a)}$ there is a unique vector field $Y(t)$ on σ such that $y(a) = y$ and the field $y(t)$ is parallel along σ. The passage from $M_{\sigma(a)}$ to $M_{\sigma(t)}$ in this way is called *parallel transport.* Thus a connection yields a notion of parallel transport along curves. It yields also a notion of absolute (intrinsic) derivatives of all orders for all tensors on the manifold, in particular for functions.

Such intrinsic derivatives, apart from those of first order, do not exist on differentiable manifolds without additional structures.

A Riemannian manifold (M, g) has a unique connection D such that $D_x y - D_y x = [x, y]$ and

$$z(g(x, y)) = g(D_z x, y) + g(x, D_z y)$$

for all vector fields x, y, and z. This is called the *Levi-Civita connection* and will be understood throughout. At every $m \in M$ the *curvature tensor*, for every pair x, y of tangent vectors, is denoted by $R(x, y)$ and is an endomorphism of $T_m M$. There are three equivalent definitions of curvature; the first two are given in terms of the Levi-Civita connection D:

• The curvature can be computed explicitly using the two first derivatives of the metric g, namely,

$$R(x, y)z = (D_y D_x z - D_x D_y z - D_{[y,x]} z).$$

• Geometrically, the value of $R(x, y)$ is the defect from the identity of the parallel transport around an infinitesimal parallelogram with sides generated by x and y. To this tensor of type $(3, 1)$, it is useful to associate the 4-linear differential form $R(x, y, z, t) = g(R(x, y)z, t)$. For numerical functions f, the absolute second derivatives are still symmetric; a special case is the commutativity of partial derivatives in classical differential calculus. The third derivatives are no longer symmetric 3-forms, and the defect is represented exactly by the curvature tensor:

$$D^3 f(x, y, z) - D^3 f(x, z, y) = R(y, z, x, \operatorname{grad} f).$$

• One looks at the defect of (M, g) from being locally Euclidean. This can be achieved, for example, by computing the length of an arc of a small circle $\Gamma(\varepsilon)$ as in Figure 6. This arc, say of angle α, is obtained from ε and from a pair (x, y) of unit vectors in $T_m M$ by going a length ε along all geodesics whose initial tangent vector is contained in the angular sector of angle α determined by x and y. The truncated expansion of this length is given by the formula

$$\operatorname{length}(\Gamma(\varepsilon)) = \alpha\varepsilon\left(1 - \frac{R(x, y, x, y)}{3\sin^2\alpha}\,\varepsilon^2 + o(\varepsilon^2)\right).$$

The symmetries of R show that the second term depends only on the tangent plane P in $T_m M$ that is determined by x and y; its name is the *sectional curvature* of P and is denoted by $K(P)$. Knowledge of $K(P)$ on the complete Grassmannian manifold of tangent planes is equivalent to knowing the curvature tensor.

The real power of the curvature tensor R and the sectional curvature K is that they measure how (M, g) fails to be locally Euclidean. That is, (M, g) is locally Euclidean (i.e., locally isometric to Euclidean space of equal dimension; one says *flat*) if and only if R (or K) vanishes identically.

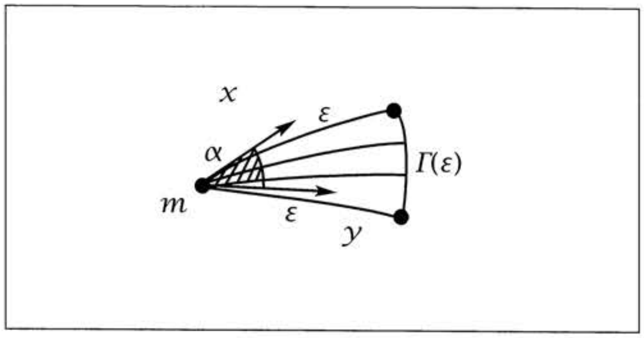


Figure 6. Curvature measures the defect of the manifold from being locally Euclidean. Sectional curvature operates at the two-dimensional level, appearing in the second term of the formula for the length of an arc of a small circle $\Gamma(\varepsilon)$.

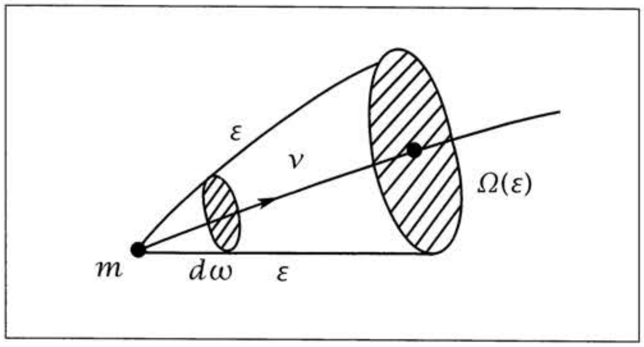


Figure 7. Ricci curvature operates directionally at the d-dimensional level in measuring the defect of the manifold from being locally Euclidean in various tangent directions. Specifically, it appears in the second term of the formula for the $(d-1)$-volume $\Omega(\varepsilon)$ generated within a solid angle.

Moreover, if K is constant everywhere and equal to k, then (M, g) is locally isometric to the standard simply connected space of constant sectional curvature k, namely, a sphere (of radius $1/\sqrt{k}$) if $k > 0$ and a hyperbolic space if $k < 0$ (the canonical hyperbolic space has curvature -1).

Something that is not emphasized in the Riemannian geometry literature is that despite its power, the curvature tensor does not in general determine the metric up to local isomorphism. There is room for strange examples, the reason being that, because of its symmetries, R depends only on $d^2(d^2 - 1)/12$ parameters, where d is the dimension of M. At present, knowledge of g requires knowing all its second derivatives, but these depend on more parameters, namely, $d^2(d + 1)^2/4$ parameters.

However, since g depends only on $d(d + 1)/2$ parameters, one could expect strong results with an invariant weaker than R. The natural one is the "Ricci curvature" Ricci, which is a quadratic form that assigns a real number $\operatorname{Ricci}(v)$ to every unit tangent vector v. This time it measures the defect from Euclidian at the level of a solid angle $d\omega$ in the direction of v, as in Figure 7. For this one looks

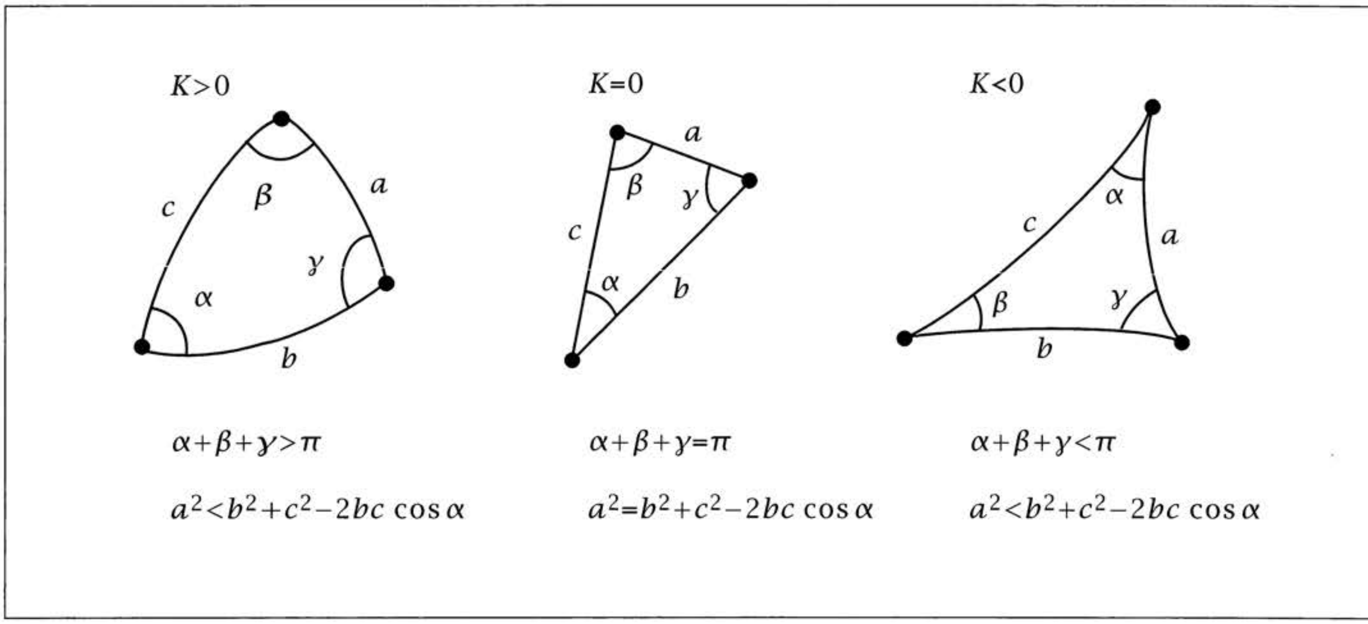


Figure 8. The manifolds with identically zero sectional curvature K are the locally Euclidean ones, and the ones with positive sectional curvature are those for which the sum of the three angles of any triangle is always larger than π, as in spherical geometry. In the negative case the sum of the angles of every sufficiently small triangle is less than π, as in hyperbolic geometry.

at the $(d-1)$-volume $\Omega(\varepsilon)$ generated by the geodesics of length ε starting in $d\omega$. The formula for the volume is

$$\mathrm{Vol}(\Omega(\varepsilon)) = d\omega \cdot \varepsilon^{d-1}\left(1 - \frac{\mathrm{Ricci}(v)}{3}\,\varepsilon^2 + o(\varepsilon^2)\right).$$

Algebraically, as a function of R or K, Ricci(v) is nothing but the trace of the sectional curvatures of all planes containing v. This expresses the fact that volumes are determinants and that derivatives of determinants are traces. One also has

$$\mathrm{Ricci}(v) = \sum_{i=2}^{d} K(v, x_i) = \sum_{i=2}^{d} R(v, x_i, v, x_i)$$

for any collection of x_i completing v to an orthonormal basis.

Finally, the *scalar curvature* scal(m) is the mean of the numbers Ricci(v) as v runs through the unit tangent vectors at m. To interpret it geometrically, we look at the limiting expansion of the volumes of small balls of radius ε centered at m. Since we have only to integrate over the unit ball of the v's, the expansion will clearly start with the volume of the Euclidean ball of radius ε; the next order term will be the scalar curvature, multiplied by a suitable coefficient.

When the metric g is multiplied by a scaling factor k, the curvatures are multiplied by k^{-1}. Scaling can thus make the curvature as small as desired. To block this effect of scaling, we can bound the diameter, which is scaled by $\sqrt{k}$.

Sectional Curvatures of Constant Sign

From the work of Élie Cartan, a Riemannian manifold has negative sectional curvature everywhere if and only if the sum of the angles of every sufficiently small triangle is smaller than π. This is true for any triangle when the manifold is simply connected. Similarly, it is not hard to see, via Cartan's argument, that there is positive curvature everywhere if and only if the sum of the angles of every sufficiently small triangle is larger than π. The surprising and basic point in many Riemannian geometry results is that this result about the angles in the positive case holds for any triangle without any extra assumption on the manifold. This was discovered by Alexandrov for surfaces and extended by Toponogov to all abstract Riemannian manifolds. The inequality between the sum of the angles and π translates into an inequality for the sides, as in Figure 8.

We say that a manifold is *negatively curved* if the sectional curvature of every tangent plane is negative, *positively curved* if the sectional curvature of every tangent plane is positive. It is therefore natural to ask the global question: Which compact manifolds enjoy such a property?

For positive sign, the striking fact is that the only examples known today are spheres; projective spaces $\mathbb{K}P^n$ over the real numbers $\mathbb{R}$, the complex numbers $\mathbb{C}$, the quaternions $\mathbb{H}$, and the Cayley numbers $\mathbb{O}$ (only $\mathbb{O}P^1$ and $\mathbb{O}P^2$ exist); and some sporadic examples in dimensions 6, 7, and 13. These last examples are also homogeneous spaces, or almost.

Apart from the very weak topological restriction coming from the positivity of scalar curvature, which is just the nullity of a single topological invariant, there was not a single restriction known for positively curved manifolds before MG's paper

[6], "Curvature, diameter and Betti numbers". In [6] MG showed that positive curvature forces the sum of the Betti numbers over any field to be universally bounded for all dimensions. In fact the proof works also for nonnegative sectional curvature. The proof is a marvelous juxtaposition of subtle algebraic topology and a type of Morse theory for the distance function. Grove and Shiohama in 1977 succeeded in extending the notion of a *critical point* to the distance function even though it is not smooth. MG showed for positively curved manifolds that if a sequence of critical points for the distance to a given point has its distances in geometric progression, then the sequence has to be finite and its length depends only on the dimension. This follows from the Toponogov comparison theorem for triangles. The result about the Betti numbers follows by using various initial points and using Mayer-Vietoris sequences. As MG remarks in [6], the proof works without much change when the condition $K \geq 0$ is replaced by

$$\text{Diameter}(g)^2 \cdot \inf K \geq k,$$

for any $k < 0$. Consequently a lower bound for the sectional curvature and an upper bound for the diameter are enough to control the Betti numbers of the manifold under consideration.

The case of negatively curved manifolds was also completely mysterious before the publication by MG in 1978 of [3], "Manifolds of negative curvature". The content of this paper and more are in book form in [1], *Manifolds of Nonpositive Curvature*, which is a detailed account of a series of lectures given by MG at the Collège de France.

From the work of Hadamard (1898) and Cartan (1926), one knew that if (M^d, g) is negatively curved, then its (simply connected) universal cover has to be diffeomorphic to $\mathbb{R}^d$. This cover can be constructed by considering the geodesics originating at any given point (by means of the exponential map). For a naive person, such as I, one considers the classification as finished, since "everything is in $\pi_1(M^d)$." So the question is reduced to an algebraic problem. Before [1] one did not know much concerning the algebraic structure of such a $\pi_1(M^d)$, seen as a group. This takes us back to the first section of Part I and the notion of hyperbolic group invented by MG to solve this negative curvature problem. Although [1] is a basic advance, one still does not know if the fundamental group of a manifold of negative curvature is any more special than an arbitrary hyperbolic group.

Before quoting results of [1], let me explain why it is so important to study manifolds of negative curvature. I quote MG on a philosophical point: "Almost all geometries are of negative curvature." Negatively curved geometries (either Riemannian or more general) are the kind that one is most likely to encounter in nature. For the moment this affirmation is only heuristic. Except in dimension two! In fact, the most natural way to construct geometries in dimension two is to glue Euclidean regular polygons (the simplest geometric objects we know) along their sides. Then the result is not a smooth surface but a locally Euclidean object with distributional curvatures at the vertices equal to 2π minus the sum of the angles of the polygons that meet there, hence negative as soon as most polygons have more than six sides. By contrast, in larger dimensions, when we glue polyhedra along their faces, we run into difficulty in knowing how to recognize at the vertices whether the distribution curvature can truly be said to be negative. Despite a number of papers on this subject, this question remains mysterious.

In [3] appears the important result of topological finiteness when the total volume of the negatively curved manifold (normalized by the condition $K \leq -1$) is bounded. There is also a finiteness result in the real analytic case. The latter result is obviously false in general: just take larger and larger connected sums. Results in Riemannian geometry using real analyticity are extremely scarce.

In [1], besides the negative case, the nonpositive situation is treated at length. For those manifolds MG introduces on the sphere at infinity the notion of *Tits metric*. This is a basic tool for studying the fine structure of negativity versus nonpositivity. To visualize the situation, the reader should think of *space forms* as basic examples. These were defined in Part I as any compact quotient of a Riemannian globally symmetric space of noncompact type. One has negativity if the rank is 1, but only nonpositivity if the rank is 2 or larger. We will return later to MG's work in the negative-curvature realm and why he thinks this study is so important. For the moment one can reduce things to what he calls a vague conjecture: "*In high dimensions every hyperbolic manifold is arithmetic.*"

We now consider the positive case, but this time for scalar curvature. At present the only classification that geometers have been able to solve in Riemannian geometry is that of manifolds with positive *scalar* curvature. Apart from a few remaining questions when the manifold is not simply connected, one has a complete classification of compact manifolds that can admit a metric with positive scalar curvature. As was mentioned at the end of Part I, a basic result in the field is a 1980 joint paper of MG with Lawson, "The classification of simply connected manifolds of positive curvature". Its basic tool is a geometrically controlled surgery.

MG wanted to find the right conceptual tools to explain existing results concerning positive scalar curvature. He managed to find one and called it "K-area". Its definition is bafflingly simple. Take all nontrivial vector bundles over M, and look for the minimum of the inverse of their largest curvature. More precisely, consider all complex vector bundles over M, put on them Riemannian bundle metrics, and endow them with a connection preserving this metric. They then have a sectional curvature. As soon as such a bundle has a nonzero Chern class, its curvatures cannot all vanish. If $\|R(X)\|$ is the supremum of this curvature over all tangent planes, then the K-area of M is the minimum of $\|R(X)\|^{-1}$, the minimum taken over all nontrivial bundles and all possible Riemannian bundle metrics on them.

In MG's 1996 paper [12], "Positive curvature, macroscopic dimension, spectral gaps and higher signatures", the central statement is the following relation between K-area and the scalar curvature Scal of a Riemannian manifold: there is a universal constant $c(d)$ such that, for every complete spin manifold M^d, $\mathrm{Scal}(M^d) \geq \epsilon^{-2}$ implies $\text{K-area}_{\mathrm{st}}(M^d) \leq c(d)\epsilon^2$. The index "st" means one works on vector bundles stabilized by products with trivial ones. Its importance should not be underestimated; it captures the essence of MG's joint work with Lawson. It has a beautiful corollary: on a torus a Riemannian metric with nonnegative scalar curvature must be flat. This result is very strong because scalar curvature is a very weak invariant, just a numerical function on the manifold. Moreover, this sheds geometric light on scalar curvature; such a light was missing before. In fact, we have seen that one can completely classify the compact manifolds admitting a metric with positive scalar curvature, but the proof is completely nongeometric. In particular, there has not yet been given a single local interpretation of the positivity of scalar curvature.

Finally, we are still far from a complete classification of manifolds with positive sectional curvature, negative sectional curvature, positive Ricci curvature, and similarly for nonnegative and nonpositive curvature.

The Space of All Riemannian Structures: $d_{G\text{-}H}$ and Collapsing

In his ICM address [4], MG launched a whole program of *synthetic Riemannian geometry*. His aim was nothing less than the study of the space of all Riemannian structures in order to give some structure to this space and to study completeness, possible convergences, compact subspaces, compactification, etc. MG says he was inspired by the history text of Klein, *Development of Mathematics in the 19th Century*. The questions as they stand are too general, so it was necessary to look at the subsets consisting of manifolds satisfying various conditions on their curvature, diameter, volume, injectivity radius, etc. The need was to understand in depth the 1969 paper of Cheeger, "Finiteness theorems for Riemannian manifolds", in which it is proved that, apart from $d = 4$, there are only finitely many diffeomorphism types for d-manifolds satisfying the following three conditions: the sectional curvature stays in $[-1, 1]$, the diameter is bounded above, and the volume is bounded below and has a positive lower bound.

MG's starting point is completely elementary: He defines a metric on the set of all complete separable metric spaces. More precisely, there are two possible definitions: one for the set $\mathcal{Z}$ of all compact metric spaces and a variant for the set of all complete separable metric spaces with base point specified. The latter is what was needed for Gromov's proof of Milnor's conjecture about groups of finite type, which was discussed in Part I. In Part II we stick to the metric on the set $\mathcal{Z}$ of all compact metric spaces, since interest will be in only that. The definition begins from the "Hausdorff distance" between compact subsets X and Y of a metric space Z, namely,

$$d_Z(X, Y) = \max\left\{ \sup_{x\in X}\inf_{y\in Y} d(x, y),\ \sup_{y\in Y}\inf_{x\in X} d(x, y)\right\}.$$

If X and Y are now no longer subsets of the same space, consider all possible pairs $\{f, g\}$ of isometric embeddings $f : X \to Z$, $g : Y \to Z$ of them into a third metric space Z. Compactness ensures that $f(X)$ and $g(Y)$ are closed. Then define $d_{G\text{-}H}(X, Y)$ as the infimum of the $d_Z(f(X), g(Y))$ over all possible Z, f, and g. In effect this distance measures the best possible simultaneous approximation of X by Y and of Y by X. This $(\mathcal{Z}, d_{G\text{-}H})$ is a complete metric space. In this space, compact Riemannian manifolds can be approximated as well as desired by finite (metric) subsets. This approximation property will be basic in the next subsection, on Ricci curvature. The metric $d_{G\text{-}H}$ is called the *Gromov-Hausdorff metric*.

The main question now is to look for compact, or precompact, or even finite, subsets of $\mathcal{Z}$. In this direction Cheeger's finiteness result looks promising. Page 74 of [9], "Volume and bounded cohomology", already written in 1978, implicitly uses a compactness result in order to prove an existence result for extremal metrics under the simplicial volume, as discussed in Part I. This existence result, as well as page 63 of the Filling paper [8], were never to my knowledge written up in detail. About the compactness result, MG says:

> I always took it for granted, but since people asked for more, I wrote it in the 1981 paper *Structures Métriques pour les Variétés Riemanniennes*. [This is [7] in the present article.]

This compactness result has its germ in Cheeger's work, but stating it explicitly had to wait until

1978. The convergence can be only $C^{1,\alpha}$. To see that one cannot do better, just consider a cylinder closed by two hemispheres. The compactness result holds under Cheeger's conditions stated above (K bounded above and below, diameter bounded above, and volume bounded below), so the natural question is whether one can suppress some of these hypotheses. The least natural is the positive lower bound for the volume, but many examples show that, if K is kept inside [-1,1] and the diameter is bounded, a manifold can "collapse" on manifolds (or more general objects) of smaller dimension. The simplest case to visualize is that of flat tori, which can collapse on tori of any dimension, and also to a point. This matter is related to the vanishing of minimal volume, which was defined in Part I.

What we have here is a whole program, sketched out in the ICM 1978 address "Synthetic Riemannian geometry": consider the limit of collapsing and also study the collapsing itself by looking at the inverse images of the collapsed points. This program was essentially completed in 1992 with Cheeger and Fukaya in "Nilpotent structures and invariant metrics on collapsed manifolds", the precise details being too involved to quote here. This work has the following interesting corollary: any Riemannian manifold admits a canonical decomposition into two sets, one where large balls remain diffeomorphic to $\mathbb{R}^d$ and the other where balls admit generalized circle fibrations, i.e., nilpotent structures.

I owe the reader a technical but extremely important notion, that of the *injectivity radius.* It is the largest number r such that all balls of radius r are diffeomorphic to $\mathbb{R}^d$ under the exponential map, namely, the spray made up of geodesics issuing from the center of this ball and of length r. Roughly speaking, collapsing can happen only when the local injectivity radius is small. A key lemma in Cheeger's dissertation was how to prevent collapsing by giving a lower bound for the injectivity radius as a function of the volume, the diameter, and the supremum and infimum of the sectional curvature.

Universality of Ricci Curvature Bounded Below

The most spectacular result of [7], *Structures Métriques pour les Variétés Riemanniennes*, is one that asserts that the two hypothesis Ricci $\geq k$ and diameter bounded above imply precompactness in $\mathcal{Z}$. Roughly put, in this class of Riemannian manifolds there are only finitely many "metric types". The proof is not too hard once one has the right framework, because, from the work of Bishop in 1963, one knows that in dimension d the condition Ricci $\geq (d-1)k$ gives complete control over the volume of balls of a given radius: for any point x the function $\frac{\mathrm{Vol}(B(x,r))}{\mathrm{Vol}(B_k(r))}$ is nonincreasing in the radius r, where $\mathrm{Vol}(B_k(r))$ denotes the volume of a ball of radius r in the simply connnected reference space of constant sectional curvature equal to k.[2] One has only to use a counting argument and a classical metric trick: when disjoint balls of a given radius r are packed as tightly as possible, then the set of balls with the same centers and with radius $2r$ is a covering. In the metric on $\mathcal{Z}$, the finite subset of the manifold made up of the centers of these balls is a good approximation to the manifold as the common radius r goes to zero.

With such a result in hand what can we now hope for? Not everything, as examples show that Ricci ≥ 0 permits an infinite number of homology types. However, in this program one now has many strong results, typically from work of Cheeger and Colding; an informative text is Gallot's lecture at the Séminaire Bourbaki in November 1997. Let us mention one result: in every dimension d there exists an $\eta(d) > 0$ such that if a manifold satisfies

$$\mathrm{Diameter}^2(g) \cdot \mathrm{Inf\ Ricci}(g) > -\eta(d),$$

then its fundamental group is nilpotent up to a subgroup of finite index. Besides many new ideas of "Ricci-synthetic geometry", essential use is made of the basic technique introduced in [5], "Groups of polynomial growth and expanding maps": extract a suitable limit from the sequence $(M^d, \varepsilon^{-1} \cdot g)$ as ε goes to zero. One step consists in showing that this limit, which captures the structure at infinity of (M, g), is pleasant—in fact, a cone. By contrast, for general Riemannian manifolds this limit can be awful. An essential idea of [7] for studying $\pi_1(M)$ is to use geometrically chosen generators with ad hoc loops. MG used this technique in [3] and also in a 1978 paper for his main theorem on almost-flat manifolds; a detailed exposition appears in the 1981 book *Gromov's Almost Flat Manifolds* by Buser and Karcher. This result answered the long-standing question: Which manifolds have almost zero curvature? The first fact is that such manifolds need not be tori, because nilpotent manifolds can be collapsed to a point, or equivalently are the manifolds obtained by successive circle fibrations starting from a point. But there are essentially no more such manifolds, and this is MG's result: manifolds with almost zero curvature have to be almost-nilpotent, i.e., the quotient, up to a subgroup of finite index, of a nilpotent Lie group. The appropriate hypothesis is

$$\mathrm{Diameter}^2(g) \cdot \sup(|K|) \leq \varepsilon(d)$$

for a universal positive $\varepsilon(d)$.

MG's work, "Paul Lévy's isoperimetric inequality", published in 1980, also deals with a lower bound for the Ricci curvature and is just as spectacular; this text, only an IHÉS preprint, appears now as one of the appendices to [13]. It shows that a lower bound on Ricci curvature is enough to completely control the "isoperimetric profile" of

[2] *The volume of the balls in the reference space does not depend on the point, as the space is a homogeneous space.*

a manifold. In (M^d, g) the *isoperimetric profile* is the function of τ given by the lower bound of the $(d-1)$-volume of the boundary ∂D over all domains D whose volume is equal to τ. The proof uses in an essential way geometric measure theory, which has been available to mathematicians since the end of the 1960s. This theory provides absolute minimal objects—here a domain of given volume whose boundary has smallest possible volume—having reasonable singularities; the set of singularities is of codimension at most 8. This control is needed in order to study PDEs on a Riemannian manifold from a geometric point of view.

The Spectrum of Riemannian Manifolds

A Riemannian manifold has a Laplacian, which is the canonical elliptic linear second-order operator given by $\Delta f = -\mathrm{Trace}_g D^2 f$. For compact manifolds this operator has a discrete spectral decomposition: the equation $\Delta\phi = \lambda\phi$ has nonzero solutions for a discrete unbounded set $\{\lambda_i\}$ of eigenvalues in $\mathbb{R}^+$, including 0. This set is called the *spectrum* of the manifold. Moreover, the λ_i always have finite multiplicity, and every eigenfunction is C^∞. Let us list each λ_i as often as its multiplicity and denote by $\{\phi_i\}$ a corresponding orthonormal set of eigenfunctions. Then every L^2 function can be written uniquely as the sum of an L^2-convergent series $\sum_i a_i\phi_i$, where $a_i = \int_M f\phi_i$. For smooth functions f, the series converges in the C^∞ topology. Consequently, in terms of this expansion one can solve various differential equations on the manifold, such as the heat equation, the wave equation, the Schrödinger equation, etc.

The main problem is to try to analyze the spectrum as a subset of $\mathbb{R}^+$. One introduces the *counting function* $N(\lambda) = \#\{\lambda_i \le \lambda\}$. Its asymptotic behavior was found in 1949 by Minakshisundaram and Pleijel:

$$N(\lambda) \underset{\lambda\to\infty}{\sim} \frac{\beta(d)}{(2\pi)^d}\,\mathrm{Vol}(g)\lambda^{d/2}.$$

In this formula $\beta(d)$ is the volume of the unit ball in $\mathbb{R}^d$, and thus $\mathrm{Vol}(g)$ and the dimension d are the only Riemannian invariants that play a role. This is the Riemannian generalization of Weyl's famous asymptotic formula for bounded domains of the plane. One wants to control the gaps, and this amounts to finding the next term in the asymptotic expansion of $N(\lambda)$. In 1968 Hörmander showed that

$$N(\lambda) = \frac{\beta(d)}{(2\pi)^d}\,\mathrm{Vol}(g)\lambda^{d/2} + O(\lambda^{(d-1)/2})$$

as $\lambda \to \infty$. The exponent $(d-1)/2$ is optimal, as is shown by the standard sphere, where the gaps are huge, expressing the fact that the eigenfunctions, which are the spherical harmonics, have large multiplicity.

However, this estimate is somewhat unsatisfactory for a Riemannian geometer, since the constant in the O term is not explicit. One would prefer an error term controlled by Riemannian invariants such as curvature, diameter, injectivity radius, etc. Moreover, one would like to control the gaps from the beginning and not only asymptotically. These two questions were settled in [12], "Positive curvature, macroscopic dimension, spectral gaps and higher signatures".

The control is given by the supremum and infimum of the sectional curvature and by the injectivity radius, which was defined above. The proof is very involved and has the striking feature that it is opposite to standard spectral arguments. Normally, when working with elliptic operators, one uses control of the spectrum for numerical functions on the manifold to control the spectrum for sections of various fiber bundles over the manifold by means of the "Kac-Kato-Feynman inequality". Here MG controls the low eigenvalues by controlling the spectrum of suitable bundles via the topology, the Atiyah-Singer index theorem, and techniques of Vafa-Witten and Bochner-Lichnerowicz. Then he applies the Kac-Kato-Feynmann inequality in reverse.

The second major contribution to the spectrum came much earlier, in the above-mentioned 1980 preprint "Paul Lévy's isoperimetric inequality"; it was made more precise by Bérard, Besson, and

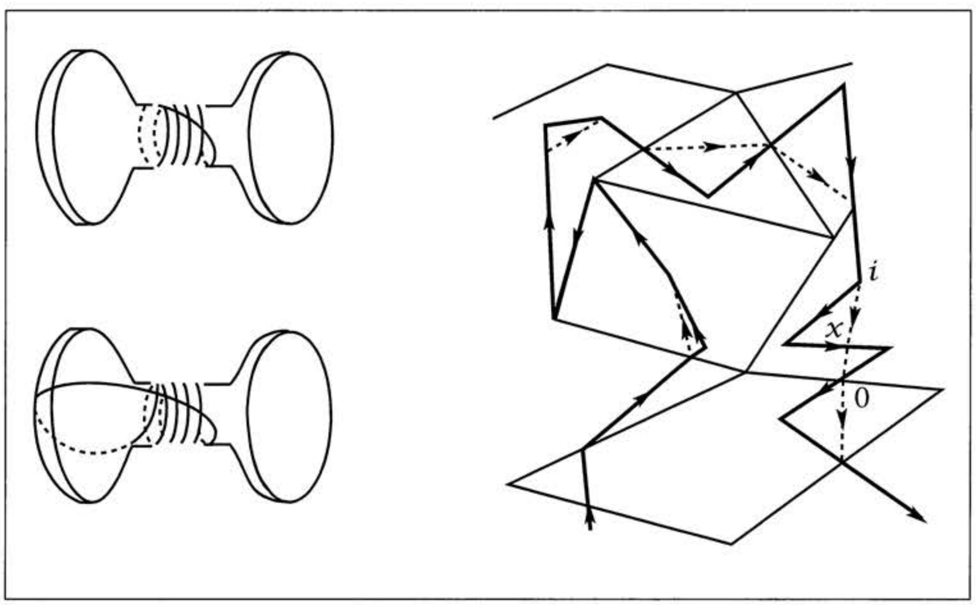


Figure 9. The diagrams on the left show how to unwrap carefully a many-times-twisted rope around a dumbbell without increasing its length too much. The diagram on the right indicates how to deduce from a triangulation of the manifold an efficient triangulation of its loop space.

Gallot in 1985. In the spirit of the preceding subsection, a control of the isoperimetric profile enabled MG to completely squeeze the ith eigenvalue within universal bounds. For every i one has:

$$\text{univ}(\text{Inf Ricci}, d, \text{Diam}(g))\, i^{2/d} \leq \lambda_i$$
$$\leq \text{univ}(\text{Inf Ricci}, d, \text{Volume}(g))\, i^{2/d},$$

where "univ" refers to some function that depends only on the specified arguments. This estimate agrees with the exponent in the Minakshisundaram-Pleijel asymptotic.

Periodic Geodesics

Having obtained a nice distribution for the spectral counting function $N(\lambda)$, one can hope that there is also a nice distribution for the length of periodic geodesics. We have in mind the following analogy: The Laplacian controls the quantum mechanics of the manifold via Schrödinger's equation, whereas the geodesic flow controls its classical Hamiltonian mechanics. One draws a parallel between ϕ_i as eigenfunction—in terms of stationary modes and pure vibrations—and the periodic geodesics γ, which are stationary motions. One is therefore tempted to draw a parallel between the eigenvalues λ_i, giving the frequencies of the vibrations, and the lengths $L(\gamma)$ of the periodic geodesics γ. In that case the geodesic flow will be completely described by the set of periodic geodesics and their lengths. In particular, it is reasonable to hope for an infinite number of such periodic geodesics, their lengths making up a discrete subset of $\mathbb{R}^+$, and to have an asymptotic expansion for the *counting function* $CF(L)$ defined by

$$CF(L) = \begin{Bmatrix} \text{number of periodic geodesics} \\ \text{of length smaller than } L \end{Bmatrix}.$$

Under this analogy the growth corresponding to $\lambda^{d/2}$ for the spectrum would be exponential in L.

One can take a suitable surface of revolution to see that the set of lengths need not be discrete. These examples show also that there are continuous bands of periodic geodesics. On the other hand, one still does not know whether, for any compact manifold and any Riemannian metric on it, there exist infinitely many periodic geodesics. One does know the existence of infinitely many periodic geodesics in dimension two; the first unsettled case is the three-dimensional sphere. In the counting process, running two or more times along a given periodic geodesic is not to be considered as different from going once; this phenomenon is the major difficulty in getting an infinite number of truly geometrically different periodic geodesics. Let us now examine MG's contributions to the subject.

We begin by briefly describing the extreme case of negatively curved manifolds. In this case there always exists at least one periodic geodesic in any free homotopy class of curves: just take the minimum length curve in the given class. Thus huge fundamental groups yield many periodic geodesics, and one gets almost trivially an exponential growth for the counting function CF. Optimal results are obtained if the manifold is the underlying one of a space form as defined earlier; in this case the factor in the exponential is no less than the one in the constant curvature case, with equality only for isometry with the space-form structure. This was proved by Katok in 1982 for surfaces and for all dimensions by Besson, Courtois, and Gallot in 1996; we quote this result because MG's notion of simplicial volume enters in a fundamental way in the proof.

Let us consider the opposite situation—simply connected manifolds. Before [2], "Homotopical effects of dilatation", there was an almost complete paralysis. Why was this? Morse theory started with Birkhoff's 1913 result yielding at least one periodic geodesic on any convex surface. The basic idea applies to any compact manifold: One considers the set $\Omega(M)$ of all closed curves on the manifold that are homotopic to zero, together with the function on this space given by length. The critical point will be exactly the periodic geodesics. So Morse theory apparently yields as many periodic geodesics as the Betti numbers of this space of curves. There are three difficulties: The first is that this space is infinite dimensional, a difficulty that is overcome by taking a finite-dimensional approximation (Birkhoff already knew how to do this by using the injectivity radius) and by replacing every curve by an approximating one made up of geodesic pieces, a so-called *broken geodesic*. The second difficulty is the computation of the Betti numbers of $\Omega(M)$. The fact that we take curves without a fixed base point makes things quite difficult, but this algebraic-topology difficulty was overcome quite successfully by various topologists. There remains the third difficulty: Morse theory gives only the existence of (many) critical points, but does not say anything about the value of the function—here the length of periodic geodesics. A typical example: If we have an infinite number of critical points (as when the Betti numbers are nonzero for an infinite sequence of dimensions of $\Omega(M)$) and if all the lengths are multiples of a given one, then the geodesics so obtained could be only the covering of a single one.

So the aim is to quantize Morse theory at the level of the values at critical points. Trivial examples show that one cannot expect any results for general functions. But in the Riemannian case and the space $\Omega(M)$ of [2], MG managed to quantify things as in Figure 9. Two ideas were used. The first pertains to the left-hand diagrams in Figure 9, where one has a curve homotopic to zero turning many times around a "thin" part; one contracts it very carefully to a point by contracting each "turn" along the big part of the manifold. Doing this for every turn yields a nice control over the length. If all turns pass through the big part at once, the length will become too large. The subtle part of [2] consists in deducing from a triangulation of M a triangulation of $\Omega(M)$; it is indicated in the right-hand diagram of Figure 9. The dimension of this triangulation equals the sum of the dimensions of the simplices of the triangulation of M that are crossed by a given broken geodesic. The final trick is to use the simple connectivity in order to contract the whole 1-skeleton to a point so that the pieces of broken geodesics running through edges no longer count.

The conclusion is this: For any compact (M, g), there exist two positive constants a and b such that for every length L one has

$$CF(L) \geq a \sum_{k \leq bL} \beta_k(\Omega(M)),$$

where β_k denotes the Betti numbers. The proof consists in contracting the curves of $\Omega(M)$ to a point while permanently controlling, with an eye on the length, the topology in degree i of $\Omega(M)$. This result applies only in the case where all periodic geodesics are nondegenerate; nondegeneracy is essential in order to apply Morse theory.

The above inequality solves the problem of periodic geodesics for "doubly generic" Riemannian manifolds: their counting function $CF(L)$ grows exponentially with L. To explain this statement, we need some algebraic topology. Compact manifolds fall into two classes: those called *rationally elliptic*, and those called *rationally hyperbolic*, most manifolds being rationally hyperbolic. The elliptic ones are those having all their homotopy groups $\pi_k(M^d)$ finite for every $k > 2d - 1$. On the other hand, from work of Felix and Halperin in 1982 one knows that rationally hyperbolic manifolds have Betti numbers $\beta_i(\Omega(M))$ that grow exponentially with i. There remain two difficulties. The first was noted above: how to get periodic geodesics geometrically, i.e., how to eliminate the ones that are multiples of others. But the iterates of a periodic geodesic have their lengths in arithmetic progression, so this difficulty goes away under the exponential. The second is that Morse theory applies only when all the critical points are nondegenerate. Here we will have the notion of nondegenerate periodic geodesic. In full generality a Riemannian manifold might have degenerate periodic geodesics, but not for a "generic metric". One knows by results of Klingenberg, Takens, Anosov, and Rademacher that "bumpy" metrics—those for which all periodic geodesics are nondegenerate—are generic; this genericity is made precise in the Baire category sense. MG's result is that for generic metrics on most manifolds one has exponential growth for the counting function of periodic geodesics. This result addresses the problem stated at the beginning of this subsection.

Which Spaces for Geometry? Gromov's Program

We now summarize Chapter $3\frac{1}{2}$ of [13], *Metric Structures for Riemannian and Non-Riemannian Spaces*. This chapter ends with: "We humbly hope that the general ambiance of X can provide a friendly environment for treating asymptotics of many interesting spaces of configurations and maps." Present models of geometry, even if quite numerous, are not able to answer various essential questions. For example: among all possible configurations of a living organism, describe its trajectory (life) in time; give as a function of time the mean diameter of planar self-avoiding Brownian motion; improve results of statistical mechanics; create a geometric theory of probability, say by quantifying geometrically the law of large numbers. Some of what we come to now is more or less known in probability theory and in statistical mechanics either formally or heuristically. The aim here is to lay a foundation, an axiomatic theory powerful enough to handle the above problems, solved or unsolved.

We now sketch the answer given in Chapter $3\frac{1}{2}$, referring the reader to the book [13] for details and many more results. One thing to realize is that in geometry the notion of measure is ultimately more important than that of metric. Measure arises first in probability theory, since it is needed to make any statistical assertion. In the Riemannian case the measure comes automatically from the metric. This order of events does not preclude the Riemannian measure from being basic. In Riemannian geometry the innovation of Riemann's was to dissociate the metric from the vector-space structure in Euclidean geometry and to replace the vector space by a differentiable manifold. Here MG dissociates the metric and the measure in a Riemannian manifold by introducing the notion of *mm-space*. This is a triple (X, d, μ) in which (X, d) is any complete separable metric space and μ is initially a finite measure on the σ-algebra of Borel sets, i.e., the smallest σ-algebra containing the open sets. Then the measure space is completed by adjoining to the σ-algebra all subsets of Borel sets of measure 0. It is assumed that every one-point set has measure 0. Let $m = \mu(M) > 0$. It is known[3] that any mm-space with $m = \mu(M)$ always admits a measure-preserving parametrization $\phi : [0, m) \to X$, i.e., a one-one onto function ϕ from the complement of a set of measure 0 to the complement of a set of measure 0 such that ϕ preserves measurable sets and the measure.

[3] *As a result of (§43, IX) in Hausdorff's* Set Theory, *Theorem 2 of the 1942* Annals *paper by Halmos and von Neumann, and an easy supplementary argument. For an exposition, see §2 of V. A. Rohlin, On the fundamental ideas of measure theory,* Translations Amer. Math. Soc. *(1)* ***10*** *(1962), 1–54.*

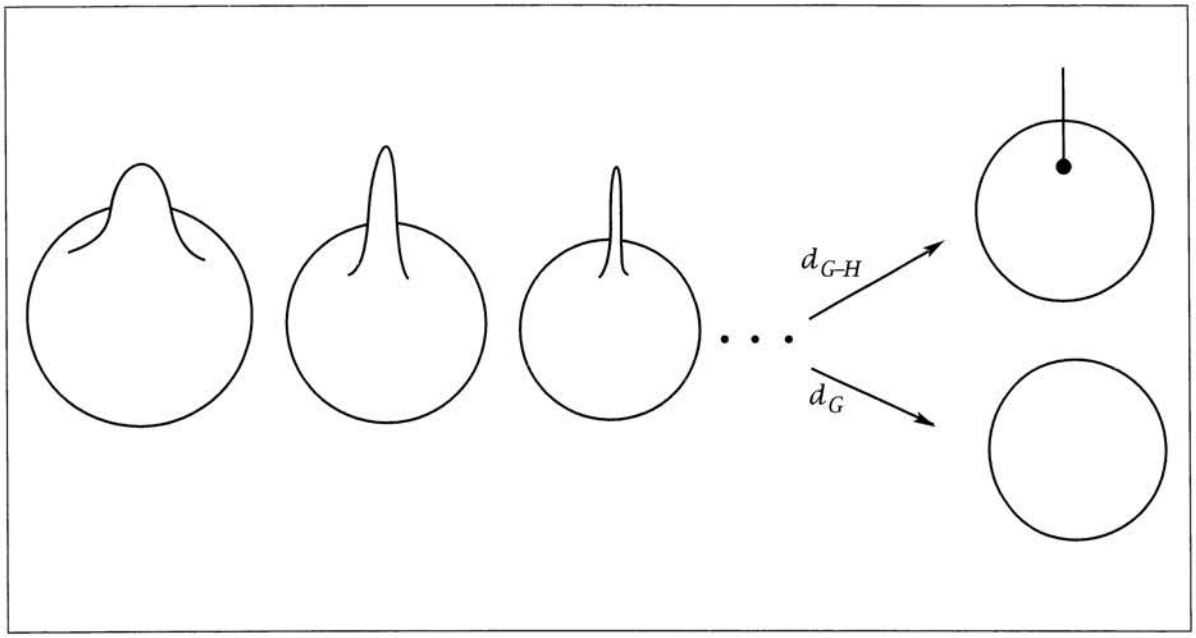


Figure 10. The limiting space of a sequence of spheres equipped with caps that are converging to a "hair" is a sphere with this hair in the metric $d_{G\text{-}H}$ but is only the sphere (with no hair) in d_G. The distinction arises because a hair has zero measure.

The first thing MG does is to consider the set $\mathcal{X}$ of all mm-spaces and to endow it with various structures. It seems that no reasonable measure can exist on $\mathcal{X}$, but MG defines a metric on $\mathcal{X}$, denoted here by d_G. For the sake of simplicity, we define $d_G(X, X')$ only when $\mu(X) = \mu(X') = m$. Unlike the Gromov-Hausdorff metric $d_{G\text{-}H}$ on $\mathcal{Z}$ defined in the section on Riemannian geometry, the metric d_G is not very geometric and is very hard to visualize since already a measure-preserving bijection between the interval $[0,1]$ and the square $[0,1]^2$ is hard to visualize. We first consider measure-preserving parametrizations of (X, d, μ) and (X', d', μ'), say $\phi : [0, m) \to X$ and $\phi' : [0, m) \to X'$. We pull back d and d' to real-valued functions on the square $[0, m)^2$, namely, $d \circ (\phi \times \phi)$ and $d' \circ (\phi' \times \phi')$. We introduce the ε *almost-distance* $\varepsilon(\phi, \phi')$ between d and d' as the smallest ε such that the set of $t \in [0, m)^2$ with

$$|d((\phi \times \phi)(t)) - d'((\phi' \times \phi')(t))| \geq \varepsilon$$

is of measure smaller than ε. Then $d_G(X, X')$ is defined as the infimum of $\varepsilon(\phi, \phi')$ over all possible parametrizations ϕ and ϕ' of X and X' of the above type. It is easy to check that this d_G is ≥ 0 and is symmetric and transitive, but it is hard to show that $d_G(X, X') = 0$ implies that X and X' are suitably isomorphic; we return to this point in a moment. The metric space $(\mathcal{X}, d_G)$ so constructed is complete. Heuristically speaking, the Gromov distance d_G is similar to the Gromov-Hausdorff distance $d_{G\text{-}H}$ in the purely metric case, but this time we are asking questions whose answers apply only almost everywhere.

One can compare the nature of $d_{G\text{-}H}$ and d_G somewhat by using a picture (Figure 10). We take a sequence of spheres with hats, the hats converging toward a segment (a hair). For $d_{G\text{-}H}$ the limit will be a sphere with a hair, but for d_G it will be only a sphere. This is satisfactory, since hairs have measure zero and thus can be neglected.

We return to comment on the proof that $d_G(X, X') = 0$ implies that X and X' are suitably isomorphic. There are two things to say. One is that the proof uses MG's notion generalizing sectional curvature to mm-spaces. The other is that the proof gives a typical example of how control of the volume (measure) of metric balls can have strong metric applications. This technique captures some of the essence of MG's precompactness in the section above. The notion of curvature used here for a given (X, d, μ) is a collection $\mu^{X;k}$ of measures, k being any natural number. The measure $\mu^{X;k}$ is defined on the space M_k of all symmetric $k \times k$ real matrices: this is the measure pushed forward from $\mu \times \cdots \times \mu$ on $X \times \cdots \times X$ by the natural map $X \times \cdots \times X \to M_k$ that assigns to k-tuples of points in X the set of their mutual distances. For example,

$$\int_{M_2} f \, d\mu^{X;2} = \int_{X \times X} f \begin{pmatrix} d(x,x) & d(x,y) \\ d(x,y) & d(y,y) \end{pmatrix} d\mu(x) \, d\mu(y).$$

It turns out that knowledge of all these measures, as k varies, allows one to reconstruct the metric d up to isomorphism of metric spaces.

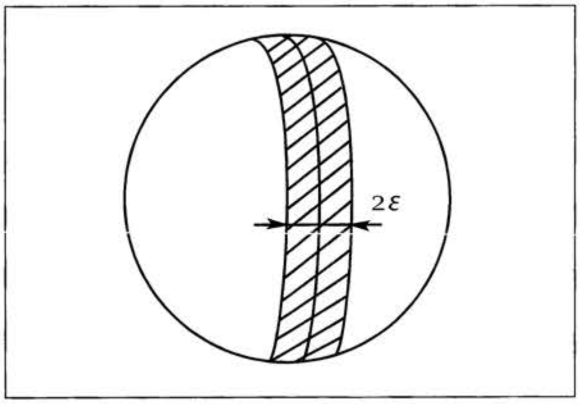


Figure 11. The complement of the ε-neighborhood of an equator of an n-dimensional unit-volume sphere has a volume that, for ε fixed, goes exponentially to zero as the dimension goes to infinity. More precisely, this volume is smaller than $2\exp(-n\varepsilon^2/2)$). Applying the isoperimetric inequality to domains of the sphere, one deduces from this that, as the dimension goes to infinity, any function tends to its central value on a set of measure arbitrarily close to 1.

Next in Chapter $3\frac{1}{2}$, the author introduces the notion of *observable diameter*—more precisely, the notion of κ-observable diameter with magnification λ. We normalize the magnification λ to be 1, since the asymptotic behavior of observable diameter as the dimension goes to infinity does not depend much on λ. The idea is to introduce notions corresponding to physical reality and physical experiments. Physical reality is taken to be a metric space (X, d). An object can be observed only by signals we perceive from it. The signals are Lipschitz functions, and we restrict ourselves to Lipschitz functions with Lipschitz constant 1, i.e., those satisfying $|f(x) - f(y)| \leq d(x, y)$ for all x and y. What we perceive, due to the lack of accuracy of our instruments, holds up only to a small error, and observable diameter is intended to capture this variability. The notion of observable diameter can be defined for any geometric concept such as the central radius (the minimal radius of a ball covering the whole metric space), the center of mass, etc. A metric and a measure are enough to define such notions.

The κ-observable diameter of (X, d, μ), denoted by $\mathrm{ObsDiam}(X, \kappa)$, is the smallest real number δ such that, for every Lipschitz numerical-valued function f on X with Lipschitz constant 1, there exists an $A \subset \mathbb{R}$ of $\mathbb{R}$-diameter smaller than δ such that

$$\mu(f^{-1}(A)) \geq \mu(X) - \kappa.$$

From the observer's point of view, this says that if $\mu_* = \mu \circ f^{-1}$ is the pushed-forward measure by f, then

$$\mu_*(A) \geq \mu_*(f(X)) - \kappa.$$

In what follows one will see that $\mathrm{ObsDiam}(X, \kappa)$ is not very sensitive to the parameter κ, so MG suggests taking $\kappa = 10^{-10}$ once and for all; then we can write simply $\mathrm{ObsDiam}(X)$. The geometric law of large numbers consists in studying in various contexts the asymptotic behavior of ObsDiam as the dimension goes to infinity. Here are some results about ObsDiam illustrating the subtlety of this notion. In particular, the topology of the space is not the important feature; its metric and its measure are the determining factors.

Historically, the first estimation of observable diameter was for standard spheres. As early as 1919 Paul Lévy studied the so-called *concentration* phenomenon of spheres S^n: most of the measure of the sphere is concentrated around an equator, and this effect becomes more pronounced as the dimension gets large. This is because $\int_0^{\pi/2} \sin^n t\, dt$ is concentrated at $\pi/2$ as $n \to \infty$. In other words, consider the tubular neigborhood $U_\varepsilon(D)$ of a hemisphere D, namely, the set of points whose distance to D is $\leq \varepsilon$; then Paul Lévy proved that

$$\mathrm{Vol}(S^n \setminus U_\varepsilon(D)) < 2\exp(-n\varepsilon^2/2).$$

The isoperimetric inequality for spheres can be seen at the level of tubular neighborhoods, and we will use it shortly for domains whose measure is, as with the hemisphere, half of the total volume. For now, consider any function f on the sphere, and look at it close to its central value c. If the isoperimetric inequality is applied to the domain where f takes values smaller than its central value, then the above two facts yield: the set of values of f that are outside the interval $[c - \varepsilon, c + \varepsilon]$ has measure $< 2\exp(-n\varepsilon^2/2)$. In the language of observable diameter,

$$\mathrm{ObsDiam}(S^n) = O(1/\sqrt{n}).$$

Of course, the exact diameter of S^n equals π for every dimension.

In the preceding section we discussed the control that MG obtains over the isoperimetric inequality for manifolds with positive Ricci curvature. MG's result can be immediately translated to the estimate $\mathrm{ObsDiam} = O(1/\sqrt{n})$ for any such manifold. But this condition is necessary, as examples show that even nonnegative Ricci curvature will lead to manifolds with $\mathrm{ObsDiam} = O(1)$ but not $= o(1)$. Topology is not the factor producing these larger estimates; other examples lead to metrics on S^n close to the standard one and whose observable diameter can be as large as desired. The hardest result on observable diameter given in Chapter $3\frac{1}{2}$ is the following: for complex algebraic submanifolds $X \subset \mathbb{C}P^n$ of degree d and codimension k, one has

$$\mathrm{ObsDiam}(X) = O\left(\frac{\log n}{n}\right)^{1/2d} \text{ as } n \to \infty \text{ with } k \text{ and } d$$

fixed. One can say that algebraicity takes the place of positive Ricci curvature (strictly speaking, there is no such relation). The metric we are considering on X is the one induced in the Riemannian sense, not just the distance on $\mathbb{C}P^n$ restricted to X in the trivial sense. This means the distance between two points is the infimum of the length in $\mathbb{C}P^n$ of all curves joining them. Metric length structures on algebraic manifolds are an extremely difficult subject which is very rarely tackled. MG likes to call this topic "the muddy waters of metric algebraic geometry." This is reflected by the fact that the behavior of the (intrinsic) diameter is still not understood. For a given degree (in a fixed $\mathbb{C}P^n$, of course) the diameter is bounded using general machinery, but does the diameter become infinite as the degree gets large? This is known only for curves in $\mathbb{C}P^2$ and only since Bogomolov's work in 1994. But the question remains open starting with surfaces in $\mathbb{C}P^3$. The proof of the above estimate for the observable diameter of algebraic manifolds needs more than twenty pages and is very involved.

A geometric law of large numbers consists in studying observable geometric quantities on the products $X^n = X \times \cdots \times X$ as $n \to \infty$. First, we have to specify which mm-structure we are considering on X^n, given a fixed mm-space (X, d, μ). For the measure one always takes the product measure, but for the metric there is a choice according to the situation. The extension of the Riemannian case consists in taking the l_2-product metric. This is a special case of the l_p-product metric, given by $(\sum_i \mathrm{dist}_i^p)^{1/p}$. For general mm-spaces one cannot do better than $\mathrm{ObsDiam}(X^n) = O(n^{1/2p})$, but this is better than the real diameter, namely, $O(n^{1/p})$. One more example due to MG is the discrete cube $\{0,1\}^n$, for which $\mathrm{ObsDiam} \approx n^{1/4}$, whereas the full diameter of the cube is $O(1)$. For the case of the regular simplex, the observable diameter is $O(1/n)$.

Now we turn to the spectrum. MG succeeds in defining a spectrum $\{\lambda_i\}$ for any mm-space. We will work with only the first eigenvalue λ_1, defining it below. Of course, in general we do not have a differential operator like the Laplacian. In the special case of Riemannian manifolds, $\lambda_1(X)$ is characterized as the minimum of

$$\frac{\int_X ||\mathrm{grad} f||^2 \, d\mu}{\int_X f^2 \, d\mu}$$

over all functions with $\int_X f \, d\mu = 0$. On an mm-space all of the above ingredients are defined except for the gradient. But one has only to define $||(\mathrm{grad} f)(p)||$ for a function f on a metric space (X, d) as $\limsup_{\varepsilon \to 0} \frac{|f(x)-f(y)|}{d(x,y)}$ for x and y in the ball of center p and radius ε. MG proves an inequality valid for any mm-space connecting λ_1 and the observable diameter:

$$\mathrm{ObsDiam}(X, \kappa) \leq \log \kappa^{-1} / 2\sqrt{\lambda_1(X)}.$$

Glances at Other Important Results

We present here, even more briefly than above, a series of results whose omission would not do justice to our geometer.

Space Forms

There is nothing more natural for a geometer than to look for geometries that generalize Euclidean geometry. Some names associated to this quest are Clifford, Klein, and Killing. One starts with *space forms* in the strict sense, namely, the geometries that enjoy the basic property of Euclidean space: two triangles with corresponding sides equal are *congruent*, which means that there is an isometry of the space (a local one in general) sending one onto the other. One can also speak of 3-*point transitivity*. If one imposes this condition as well as simple connectivity, only three geometries are possible: Euclidean, spherical, and hyperbolic. This class coincides with simply connected Riemannian manifolds of constant sectional curvature.

We shall take the spaces in question to be compact manifolds that are quotients of the three standard ones. We stick to the compact case and to manifolds for simplicity. In the Euclidean and spherical cases, examples are easy to construct; moreover, one had a complete classification by the end of the 1960s. The hyperbolic case is a completely different story. In dimension two, examples are easy, but a classification is much harder and in fact is the basic content of Teichmüller theory for Riemann surfaces. But starting in dimension three, one had to wait until 1931 to have some examples, and in higher dimensions until Armand Borel in 1963. Borel's construction is based in an essential way on number theory, the corresponding space forms being called *arithmetic*; his construction is valid for all symmetric spaces of arbitrary rank. In a joint 1988 paper with Piatetski-Shapiro, "Nonarithmetic groups in Lobachevsky spaces", MG managed to construct some nonarithmetic examples in all dimensions (even if number theory was always present at the start).

It remains open whether arithmetic examples are more numerous or less numerous than nonarithmetic ones. MG has a program to try to solve this. His idea is to mix suitably the notion of hyperbolic polyhedron with number theory. Part of the difficulty is that the known construction of Riemann surfaces by taking triangles or other polygons in the hyperbolic plane and reflecting them about their sides—this kind of construction was shown by Vinberg in 1984 to be impossible in large dimensions (around 40). The construction works in dimension three, but which dimensions permit such constructions by reflections is an open problem.

One next looks at *space forms of rank one*, the simply connected ones being the symmetric spaces of rank one. The compact simply connected ones are the generalized projective spaces $\mathbb{K}P^n$ met in the Riemannian geometry section. For $\mathbb{K} \neq \mathbb{R}$, there is no classification problem for manifolds that are compact quotients of these because even dimension and positive curvature force simple connectivity up to a two-element group by a theorem of Synge.

The analogous simply connected negative curvature spaces, denoted here by $\mathrm{Hyp}^n(\mathbb{K})$, offer more of a challenge. (For $\mathbb{K} = \mathbb{R}$ one has the standard hyperbolic geometry.) The geometric characterization of these simply connected spaces is that they are 2-*point transitive* in the sense that pairs of points with the same mutual distance can be carried to each other by a global isometry. We again look for compact quotient manifolds of these spaces; they will be the locally 2-point transitive geometries. Such quotients exist, as shown by Borel, but their classification is not finished: For $\mathbb{K} = \mathbb{C}$ one knows only the existence of some nonarithmetic examples. For $\mathbb{H}$ and $\mathbb{O}$, it is shown in a 1992 joint work of MG and Schoen, "Harmonic maps into singular spaces and p-adic superrigidity for lattices in groups of rank one", that all such quotients must be arithmetic. Of this result, MG says that the most important thing about the paper is not this corollary, but the introduction and use of "harmonic maps" with values in singular spaces (Tits buildings in this case). This technique is now widely used.

In this area the work of [11], "Foliated Plateau problem", has not received much attention. However, MG believes that this pair of papers is important. He says:

> One of the essential ideas of this text is that, in treating the solutions of elliptic equations, the right framework is that of foliations. But, if one excepts the trivial case where the tangent bundle is enough, in general one has to go to infinite dimensions to get the space of solutions. There are some holes in this text, but it does several things: it furnishes this general framework and therefore serves to make the problems well posed, and afterward it contains also some things in the spirit of Nevanlinna theory.

Kählerian Manifolds

In much of his work MG examines Kähler manifolds with a vengeance. He absolutely wants to find *robust* results. For example, integrability of an almost-complex structure—the condition that the structure come from a complex structure—is fragile. On the other hand, invariants such as the fundamental group and the spectrum are robust. In the 1989 paper by MG, "Sur le groupe fondamental d'une variété kählérienne", one finds the first known strong restriction for the algebraic structure of possible fundamental groups of Kähler manifolds. The final theorem has not yet been obtained. Classical methods are useless for this result. MG used transcendental methods, namely, L^2-cohomology and the index theorem. He used L^2-cohomology also in other instances, such as in his 1991 paper "Kähler hyperbolicity and L^2-Hodge theory", in which he showed that the two conditions "negative sectional curvature" and "Kähler" determine the expected sign of the Euler-Poincaré characteristic. This is a conjecture of H. Hopf from the 1930s, stated for the general Riemannian case and sectional curvature of constant sign. The case of dimension two follows immediately from the Gauss-Bonnet theorem. The case of dimension four is proved. The conjecture is open starting in dimension six, and examples indicate that the proof cannot follow directly from the higher-dimensional generalization by Allendoerfer and Weil of the Gauss-Bonnet theorem.

Building Examples in Riemannian Geometry

The construction of subtle examples is an important aspect of mathematics. MG has produced a number of these.

In the section on Riemannian geometry, we saw results going beyond the condition "sectional curvature is positive, or nonnegative". It is natural to ask the same kind of question for negative sectional curvature. This cannot be done, as MG constructs for every $\varepsilon > 0$ on the sphere S^3 a Riemannian metric whose diameter is equal to 1 while the curvature satisfies $K < \varepsilon$.

In Part I we met the notion of systolic softness. The basic example furnished by MG is incredibly simple: consider $S^1 \times S^3$, as obtained from $[0,1] \times S^3$ by gluing the two copies $\{0\} \times S^3$ and $\{1\} \times S^3$ with a Clifford translation (i.e., along Hopf fibers) of greater and greater length.

According to MG, his most subtle and important constructions are in the realm of negative sectional curvature. Recall from the beginning of Part II that it is difficult to construct compact manifolds of negative curvature. Borel's examples of space forms have sectional curvature in $[-1, -\frac{1}{4}]$. Of course, one can just deform (not too much) those examples, and sectional curvature will remain negative. However, this approach leaves untouched the question of finding a classification of the set of negatively curved manifolds. The class of these manifolds is very interesting, as it supplies us with objects worthy of study for themselves but also very subtle to deal with, since products automatically yield many vanishing curvatures. Even the case of polyhedra is not simple (except as we saw in dimension two). Finally, these manifolds are linked with the hyperbolic groups seen at the beginning of Part I.

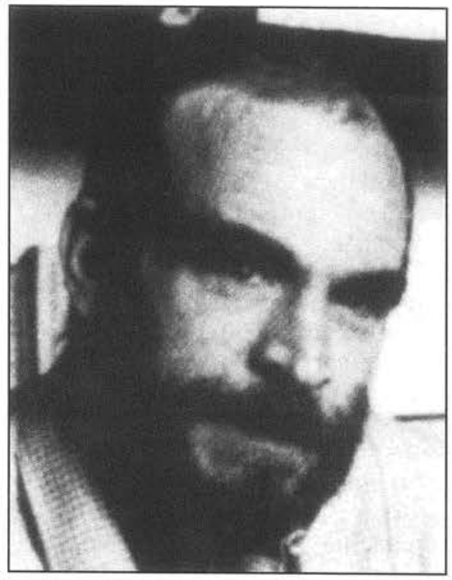
In a 1987 joint paper with Thurston, "Pinching constants for hyperbolic manifolds", one finds two essential constructions for negative curvature that work in every dimension ≥ 4. In both constructions one starts with a compact space form M of hyperbolic type, i.e., the sectional curvature is constant and is equal to -1. Consider in M a totally geodesic submanifold of codimension 2 (i.e., a submanifold N in which geodesics starting in M and tangent to N remain in N). Look now at cyclic coverings of M ramified along N. It is not too hard to endow such a covering with negative curvature, and one can even control the *pinching*, the ratio $\sup K / \inf K$. MG studies the volumes of these objects. A major result of the book [1], *Manifolds of Nonpositive Curvature*, furnishes bounds for the volume as a function of the pinching. This construction yields manifolds whose topology can differ strongly from that of a space form. In the first type of example one can show that the pinching can be as close as desired to 1. Hence the conclusion: for any ε there exist manifolds with curvature in $[-1-\varepsilon, -1+\varepsilon]$, of bounded diameter, that do not admit a metric of constant negative curvature.

A second construction enables MG to obtain examples of a complementary type: for every ε with $0 < \varepsilon < 1$, there exist manifolds of negative curvature that do not admit a metric with curvature in the range $[-1, -1+\varepsilon]$. This result is hard to prove but essential to the understanding of negative curvature. The proof uses the technique of diffusion of cycles discussed in Part I.

Conclusion

If MG has a muse, it is not the axiomatic one of Euclid. MG is instead guided by concepts such as softness versus rigidity, computability, physical reality of objects, etc. In particular, when talking about results, he is concerned with the robustness of the invariants used. His other principle is to avoid empty generalization: "Many theorems are not interesting if one cannot produce examples where the result is not already there". From this point of view the Filling paper [8] is exemplary. In case some of his results do not meet the above criterion, he adds, "Then put them in what is now called foundations."

We have seen time and again that MG's papers are like icebergs: most of the results lie under the surface and are accessible only to exceptional mathematicians who are willing to devote their time to them. So why does MG not write his results in detail? We think that the best way to answer this and other questions is to let MG speak for himself: "Checking in full detail the proof in my head was already so painful that I was left with no energy for more." Let us also quote what he says in a expository paper of 1992, "Stability and pinching":



Mikhael Gromov

> The results we present are, for the most part, not new and we do not provide detailed proofs (these can be found in the papers cited in our list of references). What may be new and interesting for non-experts is an exposition of the stability/pinching philosophy which lies behind the basic results and methods in the field and which is rarely (if ever) presented in print (this common and unfortunate fact of the lack of an adequate presentation of basic ideas and motivations of almost any mathematical theory is, probably, due to the binary nature of mathematical perception: either you have no inkling of an idea or, once you have understood it, this very idea appears so embarrassingly obvious that you feel reluctant to say it aloud; moreover, once your mind switches from the state of darkness to the light, all memory of the dark state is erased and it becomes impossible to conceive the existence of another mind for which the idea appears nonobvious).

Finally, for those who want to know more about MG's process of discovery, we end with the following quotation of his response to his being awarded the AMS Steele Prize in 1997 (*Notices*, March 1997). The response analyzes the results of [10], "Pseudo-holomorphic curves in symplectic manifolds", a paper that was discussed in Part I:

> I saw the light when struggling with Pogorelov's proof of rigidity of convex surfaces where he appeals to the Bers-Vekua theory of quasi-analytic functions. There was nothing seemingly complex-analytic in the linearized system written down by Pogorelov, and

then it struck me that *every* first order elliptic equation or quasilinear system of two equations in two variables has the same principal symbol as Cauchy-Riemann and then the solutions appear as (pseudo) holomorphic curves for the almost complex structure defined by the field of the principal symbols. Now the surface rigidity trivially followed from positivity of the intersections of holomorphic curves. What fascinated me even more was the familiar web of algebraic curves in a surface emerging in its full beauty in the softish environment of general (nonintegrable!) almost complex structures. (Integrability had always made me feel claustrophobic.) And my mind was ready for the miracle; Donaldson's ideas were in the air. So I tried to replay Yang-Mills on my holomorphic curves (strings?) and reluctantly abandoned the idea, being convinced by Pierre Deligne that the area of curves cannot be controlled without a symplectic structure. Everything went smoothly with the symplectic structure, and I even came to understand the definition of quasianalytic functions and of the nonlinear Riemann-mapping theorem of Schapiro-Lavrentiev (albeit I am still unable to read a single line of this style of analysis).

I was happy to see my friends using holomorphic curves immediately after birth: Eliashberg, Floer, McDuff. Eliashberg came across them independently in the contact framework but was unable to publish (staying in the USSR). Floer has morsified them by breaking the symmetry, and I still cannot forgive him for this. (Alas, prejudice does not pay in science.) McDuff started the systematic hunt for them which goes on till present day. And what goes on today goes beyond these lines and the pen behind them.

References

[1] W. Ballmann, M. Gromov, and V. Schroeder, *Manifolds of Nonpositive Curvature*, Birkhäuser, Boston, 1985.

[2] M. Gromov, Homotopical effects of dilatation, *J. Differential Geom.* **13** (1978), 303–310.

[3] ____, Manifolds of negative curvature, *J. Differential Geom.* **13** (1978), 223–230.

[4] ____, Synthetic geometry in Riemannian manifolds, *Proceedings of the International Congress of Mathematicians* (Helsinki 1978), Acad. Sci. Fennica, Helsinki, 1980, pp. 415–419.

[5] ____, Groups of polynomial growth and expanding maps., *Publ. Math. IHÉS* **53** (1981), 53–73.

[6] ____, Curvature, diameter and Betti numbers, *Comment. Math. Helv.* **56** (1981), 179–195.

[7] ____, *Structures Métriques pour les Variétés Riemanniennes* (J. Lafontaine and P. Pansu, eds.), Textes Mathématiques, 1, CEDIC, Paris, 1981.

[8] ____, Filling Riemannian manifolds, *J. Differential Geom.* **18** (1983), 1–147.

[9] ____, Volume and bounded cohomology, *Publ. Math. IHÉS* **56** (1983), 5–99.

[10] ____, Pseudoholomorphic curves in symplectic manifolds, *Invent. Math.* **82** (1985), 307–347.

[11] ____, Foliated Plateau problem, I and II, *Geom. Funct. Anal.* **1** (1991), 14–79 and 253–320.

[12] ____, Positive curvature, macroscopic dimension, spectral gaps and higher signatures, *Functional Analysis on the Eve of the 21st Century II* (New Brunswick, NJ, 1993), (S. Gindikin, J. Lepowsky, and R. Wilson, eds.), Birkhäuser, Boston, 1996, pp. 1–213.

[13] ____, *Metric Structures for Riemannian and Non-Riemannian Spaces*, Birkhäuser, Boston, 1999.

About the Cover

While in Euclidean 3-space there is only one regular dodecahedron and its dihedral angle is approximately 116°, in hyperbolic 3-space there exists a continuous family with varying dihedral angle.

In the limiting case for small distances, hyperbolic 3-space looks like Euclidean 3-space, and small hyperbolic dodecahedra therefore have dihedral angles close to 116°. On the other hand, for the largest possible dodecahedron with the vertices on the sphere at infinity, the dihedral angle is precisely 60°.

Of special interest are the three dodecahedra with the intermediate dihedral angles 60°, 72°, and 90°, because they tessellate hyperbolic 3-space. These dodecahedra are shown on the front cover, each at the center of its own Poincaré ball model of hyperbolic space.

Examples of tessellations of hyperbolic 3-space by a bounded region were unknown before 1931.

—*Matthias Weber*