Snapshots of modern mathematics
from Oberwolfach

№ 13/2022

Closed geodesics on surfaces

Benjamin Dozier

We consider surfaces of three types: the sphere, the torus, and many-holed tori. These surfaces naturally admit geometries of positive, zero, and negative curvature, respectively. It is interesting to study straight line paths, known as geodesics, in these geometries. We discuss the issue of counting closed geodesics; this is particularly rich for hyperbolic (negatively curved) surfaces.

1 Topological surfaces

Suppose you are an ant living in a two-dimensional world. What are the possibilities for the shape of this world? First, you would like to know some very basic information, without getting into intricacies of things like distances and angles. Suppose you only care about notions that would stay the same after continuous stretching and bending. Such properties are called *topological* (not to be confused with “topographical”, which is concerned with elevations). One possibility is that this world could be an infinite plane. Or it could be a sphere. These two are topologically different: you cannot get from one to the other without doing something violent (this is why making a flat map of the world is hard). The space could also have holes, for instance it could look like a *torus*, the surface of a doughnut. Let’s assume that the world has only a finite area (compact), no boundary, and is *orientable*. This last condition means that the surface has just one side, and rules out exotic spaces like the Klein bottle. It turns out that all such surfaces can be classified into three types, which are listed in Theorem 1.1 and illustrated in Figure 1.

Theorem 1.1 (Classification of surfaces). *Every compact, orientable surface without boundary has the same topology as either*

1. *a sphere,*
2. *a torus,*
3. *a surface of genus g, with some $g \geq 2$ (that is, a "g-holed torus").*

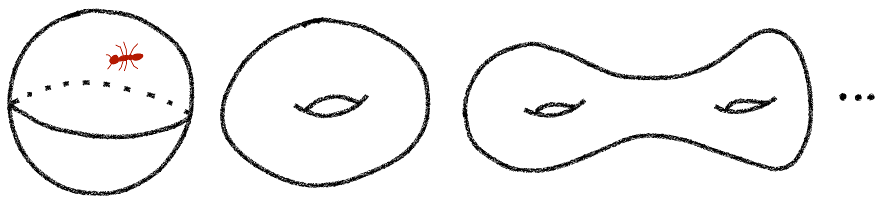


Figure 1: The sphere, the torus, and a surface of genus 2.

2 Geometry on surfaces

We've just seen that the topology of the surfaces that we're interested in is well understood and not so complicated. But of course we often do care about properties such as lengths and angles. The study of these properties is what we mean by *geometry*. There will now be many, many different possibilities for a fixed topology. For instance, consider the round sphere and an ellipsoid (see Figure 2). For each of these spaces, the way that we measure the distance between a pair of points p, q is to consider all paths in three-dimensional space that lie on the surface and connect p and q, and take the shortest length among these. The geometries we get are genuinely different, as can be seen by noting that all the points on the sphere behave exactly the same way, while the ellipsoid's shape is less symmetric. We could also take a round sphere, cut out a little piece, modify it somehow and glue it back in, to get yet a different geometry.

It is interesting to study the whole diverse zoo of geometries, but for our purposes this will be too many possibilities, so we will limit our scope somewhat. We will focus our attention on what are called *homogenous* geometries, those that locally look the same near any point of the surface.

2.1 The sphere

It turns out that when the topology is that of the sphere, there is really only one possibility for such a homogenous geometry, namely the normal round, perfectly

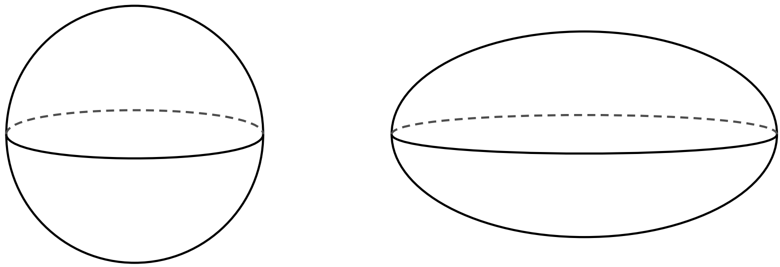


Figure 2: The round sphere and an ellipsoid. They have the same topology, but the ellipsoid's geometry is less symmetric.

symmetric sphere that we are familiar with.[1] (Actually, multiplying all the distances by some positive constant will lead to a new homogenous geometry, but this behaves so similarly to the original that we regard it as essentially the same.) So by imposing this condition of homogeneity, we have cut down from a diverse collection of possible geometries to a single special one.

One property of the round metric on the sphere is that it is everywhere *positively curved*. Roughly speaking, positive curvature at a point p means that the whole surface "bends inward" at p. If we were an ant confined to this sphere, we wouldn't be able to directly detect this inward bending, but we could do experiments to indirectly verify it. For instance, if we were to trace out a small circle on the sphere (that is, pick a point p and draw all the points that are some fixed distance r from it), we would notice that the circumference of the circle is a bit smaller than $2\pi r$, the circumference of a circle of radius r in standard Euclidean geometry. This "small circumference" property is a feature of positive curvature.

One can in fact quantify the curvature at a point p as a real number; it measures the difference between the circumference, $2\pi r$, of a Euclidean circle and the circumference of a circle of radius r centered at p on our geometric surface. All points on the ellipsoid will have positive curvature, but it won't be constant across the surface. For instance, at the extreme left and right points, the curvature is high, while at the extreme top and bottom points, it is low (but still positive). A precise definition of the concept of homogeneity discussed above is that the curvature is the same at every point of the surface.

[1] Proofs of this statement are somewhat involved, using either significant amounts of complex analysis or differential geometry.

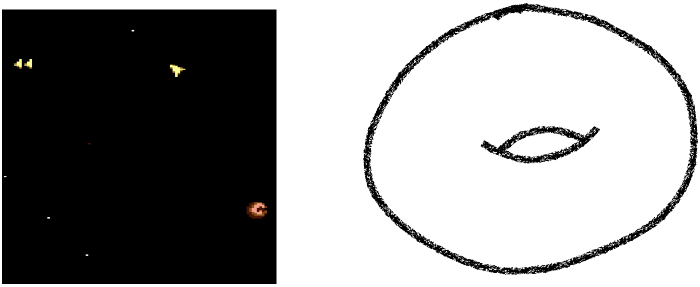


Figure 3: On the left, a screenshot of the *Asteroids* game. Bullets fired by the ship towards the top of the screen come out at the bottom. On the right, a torus results from taking the square and gluing the top to the bottom, and the left to the right.

2.2 The torus

One way to build a torus is to start with a square, and glue pairs of opposite sides together. This type of world is familiar from several classic arcade video games, including *Pac-Man* and *Asteroids* (as shown on the left in Figure 3). In these games, if you move off the top of the screen, you come back at the bottom screen, directly below where you went off the top. Similarly, if you move off the left side of the screen, you come back on the right side, directly to the right of where you went off.

The geometry that we put on this torus comes from the square: locally, the distance between two points is defined to be exactly the distance between the corresponding points on the square. (Note that this is a different type of construction than for the round sphere metric; although we can topologically embed a torus into three-dimensional space, the metric we are considering does not have anything to do with this embedding). This geometry on a torus has the homogeneity property that we want. Even though the edges and vertices of the square might initially seem to be special, after we've done the gluing, they look exactly like any other point on the torus.

How many different homogeneous geometries are there? Notice that we can generalize the construction above. Instead of starting with a square, we can start with any parallelogram and glue opposite sides together to get a torus. Some pairs of different parallelograms will actually result in the same geometric torus. But this construction still gives a continuous, infinite family of different homogeneous geometries on the torus (in contrast to the case of the sphere, where there was just one possibility). And it turns out this construction gives *all* the possible homogeneous geometries that one can put on a torus.

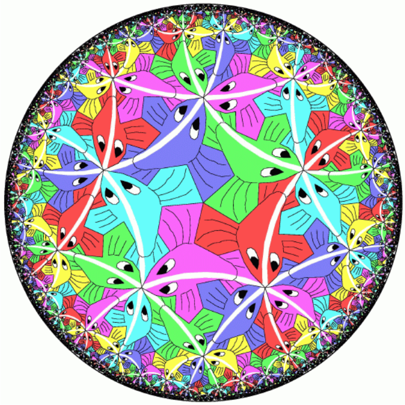


Figure 4: Fish in the hyperbolic disc. All the fish have the same size in the hyperbolic metric.

Notice that, in terms of curvature, our tori behave exactly like the Euclidean plane. If we draw a circle of some small radius r on such a torus, it looks exactly like a standard Euclidean circle; in particular, its circumference is $2\pi r$. This means that each torus constructed above has *zero* curvature at every point.

2.3 Higher genus surfaces

We saw above that on the sphere, the (unique) homogeneous geometry has positive curvature, while on the torus, there are many homogeneous geometries, but they all have zero curvature (modeled on the flat Euclidean plane). These geometries come up frequently in every day life. But what about surfaces of genus $g \geq 2$?

The answer is found in *hyperbolic geometry.* The hyperbolic disc is the model space for negative curvature. The topology of the hyperbolic disc is actually the same as that of the Euclidean plane, but its geometry is radically different. Locally it looks like a saddle at every point. Circles of small radius r have circumference *larger* than $2\pi r$. Does such a geometry really exist?

To construct the hyperbolic disc, we start with the open unit disc (all points of distance less than 1 from the origin in the Euclidean plane). Now very close to any point p, we define hyperbolic distance to be the Euclidean distance multiplied by the factor

$$\frac{1}{1 - \|p\|^2} \tag{1}$$

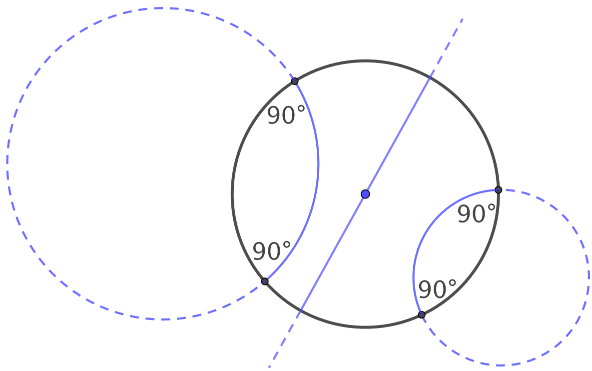
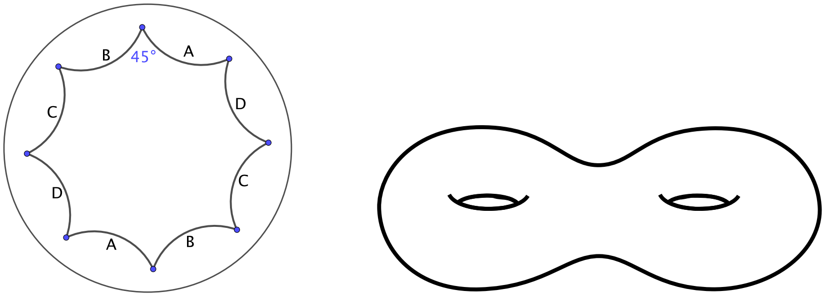


Figure 5: Some hyperbolic geodesics (the solid blue arcs and line segment) on the hyperbolic disc (the disc bounded by the black circle).

(here $\|p\|$ is the distance from p to the origin). Note that as p approaches the boundary unit circle, this factor becomes very large (since its denominator becomes very small). So hyperbolic distances near the unit circle are vastly greater than the Euclidean distance. On the other hand, angles in the hyperbolic sense are defined to be the same as they are in the Euclidean sense. Figure 4 represents the hyperbolic plane. In the hyperbolic metric, all the fish are exactly the same size, though the ones near the boundary unit circle look smaller to our Euclidean eyes.

A *geodesic* is the generalization of a straight line to an arbitrary geometry. The defining property is that if we take two points which are close together on a geodesic, the path along the geodesic will be the shortest path between those two points. So geodesics are the routes one should take to travel most efficiently. The geodesics in a flat plane are the usual straight lines between points. Knowing about geodesics tells us a lot about the geometry of the surface – it is akin to knowing all the best routes between towns in some province. Now let us discuss geodesics on the hyperbolic disc. It turns out that from the formula (1) for hyperbolic distance, one can prove that hyperbolic geodesics on the disc are circular arcs that meet the boundary circle at right angles. (Here, and in what follows, we understand the angle between two arcs of circles to be measured between their tangents at the point where they meet.) Well, that is what most of them look like, but lines through the center of the disc also count (a line can be profitably thought of as a "circle" that goes through infinity). This is illustrated in Figure 5.



(a) An equilateral hyperbolic octagon with all angles 45°.

(b) Genus 2 surface formed by gluing opposite sides of the octagon.

Figure 6

A brief historical aside. The existence of the hyperbolic disc is a result from the early 19th century. It gave a rather surprising answer to a question that geometers had been mulling over since the time of the ancient Greeks. In Euclid's axiomatization of plane geometry, the "parallel postulate" seemed less obvious than the others. An equivalent formulation of this postulate is that the sum of the angles in any triangle equals 180°. The ancient Greeks wondered if this statement could in fact be proved from the other postulates. The existence of the hyperbolic disc shows that this is not the case. This geometry does not satisfy the angle condition for triangles (for instance, if you further divide the octagon in Figure 6a into triangles, you can see that at least one must have angle sum strictly less than 180°), while it does satisfy the other postulates.

It takes a while to get a feel for this exotic geometry, but it turns out to be very well suited to higher genus surfaces. We can adapt our construction of geometric tori from the flat Euclidean plane, by replacing the Euclidean plane with the hyperbolic disc. We begin by finding a hyperbolic polygon (that is, a shape whose edges are hyperbolic geodesics) that has eight sides, and all eight angles are equal to 45°. This would not be possible in the Euclidean plane, but amazingly it is in the more flexible hyperbolic disc. We then glue together opposite sides, as indicated by the letters in Figure 6a. Because the angle has been chosen carefully (so that the sum of all eight angles is 360°), upon doing the gluing, all points, including the single point coming from the eight original vertices, look the same. And, with some imagination, one can see that after the gluing what we get is a genus 2 surface.

How flexible is the construction described above? If we carefully vary the sides of the octagon keeping the sum of the angles the same, and opposite sides

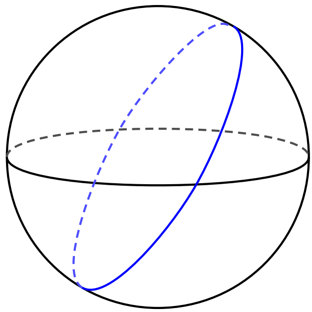


Figure 7: The sphere with its equator and another great circle.

of equal length, then we will still be able to glue them to get a hyperbolic genus 2 surface. This gives a continuous family of geometric surfaces. Furthermore, we can get a hyperbolic structure on any surface of genus $g \geq 2$ by gluing together sides of an appropriately chosen hyperbolic polygon. All hyperbolic structures are obtained from hyperbolic polygons in this way. And the hyperbolic structures are the only homogeneous geometries on surfaces of genus $g \geq 2$. (Note that one cannot get a surface of genus 1 in this way, since any hyperbolic rectangle has angle sum strictly less than 360°, and thus the four corners of the square cannot be glued together nicely.)

The hyperbolic plane, and thus every hyperbolic surface, has negative curvature at every point. This can be seen from a direct calculation of circumferences using the metric as defined in (1) above. It is also a reflection of the "fast growing" property of hyperbolic space. As we move away from the center of the hyperbolic disc, space grows very quickly, and in particular circles have larger circumference than circles of the same radius in Euclidean geometry.

3 Geodesics on surfaces

3.1 Geodesics on the sphere

On the round sphere, geodesics (the most efficient paths) are *great circles*. On the surface of the Earth, one example would be the equator. In general, a great circle arises by taking the intersection of the sphere with any plane through the center of the sphere, as shown in Figure 7. If you fly from New York to Beijing, your flight will likely take you over the North Pole, since this lies on the great circle connecting the two cities. This is surprising if you try to guess the best route by looking at a standard paper map of the world.

Notice that geodesics on the sphere have some interesting properties. First, every geodesic eventually returns to where it started, and it will be going in the same direction that it started off in when it returns. In general a geodesic with this property is called a *closed geodesic* (or a *periodic geodesic*, since the behavior eventually repeats as you move forward or backward). The *length* of the closed geodesic is just the distance traveled before it starts repeating. The fact that all its geodesics are closed is something extremely special about the sphere. Second, we observe that there are infinitely many closed geodesics, but any two closed geodesics are topologically the same, in that we can move one continuously to the other (they are *homotopic*).

3.2 Geodesics on the torus

Understanding geodesics on the flat tori we constructed above is in some sense easier than it was for the round sphere. We already understand straight lines in the plane very well. To get a geodesic on a torus, we just start with a straight line segment and extend it as far as possible in both directions. Most of the time, the geodesic will wind around the surface, never coming back to exactly where it started. But for special starting directions, we will get a closed geodesic. If our torus is obtained from the square, this will happen exactly when the slope of the line is a rational number. When we've found one closed geodesic, we can translate it a little to get another closed geodesic. So each is part of a continuous family of parallel closed geodesics of the same length. But if we take two closed geodesics that go in different directions, we cannot continuously deform one to the other; they are not homotopic.

Mathematicians like to count things, and geodesics are interesting things to count. Even if we only count families of parallel closed geodesics, there are still infinitely many. But the number of families of length less than any particular finite bound L is finite. Hence it makes sense to ask about this number. Computing it exactly is a bit too much to ask, but we can study how it behaves as the bound L approaches infinity. Here is the answer:

Theorem 3.1. *The number of different families of closed geodesics of length at most L on a torus of area 1 is asymptotic to*

$$\frac{6}{\pi^2} \cdot \pi L^2.$$

Here the term *asymptotic* means that the ratio of the actual count of geodesics to the quantity $\frac{6}{\pi^2} \cdot \pi L^2$ tends to 1 as L tends to infinity. We could cancel a factor of π, but writing as above reveals a number-theoretic connection. The probability that two randomly chosen large integers are relatively prime (do not have any prime factors in common) tends to $6/\pi^2$, which gives the first

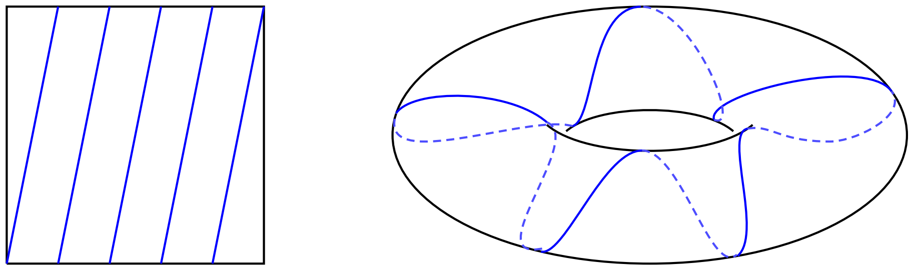


Figure 8: A closed geodesic on the torus.

term above. The area of a Euclidean ball of radius L is πL^2, which is the second term. If we simplify the discussion by considering just the square torus, then one can see that families of closed geodesics correspond to straight line segments connecting the origin in the Euclidean plane to a point of the form (m, n), where m and n are relatively prime integers. You can prove the above formula for the square torus using this observation.

But perhaps even more important than the precise constants is the appearance of the L^2 term in the above. We say that the number of geodesics grows *quadratically* in terms of length, since L^2 is a quadratic function of L.

3.3 Geodesics on higher genus surfaces

The story presented thus far has been well-understood for more than 100 years. But the situation is harder, and more interesting, for hyperbolic surfaces. There have been many recent advances, but there are still lots of open questions.

What do geodesics on hyperbolic surfaces of higher genus look like? There are many cases. Just as for the torus, a geodesic can ceaselessly wander around the surface. However, we will focus on the closed geodesics, those that eventually return to their starting point going in the same direction that was initially started off in. We note two phenomena on hyperbolic surfaces that do not have analogs on the flat torus. First, a closed geodesic can self-intersect (this happens when the geodesic returns to a point it has already hit, but going in a different direction). A closed geodesic that does not self-intersect is called *simple*; examples of a simple geodesic and a non-simple one are illustrated in Figure 9. Second, any two different closed geodesics are not homotopic. This implies that closed geodesics do not occur in families; each is "alone". We can ask, then, the same question that was answered for the torus by Theorem 3.1, namely, how many closed geodesics of length at most L are there?

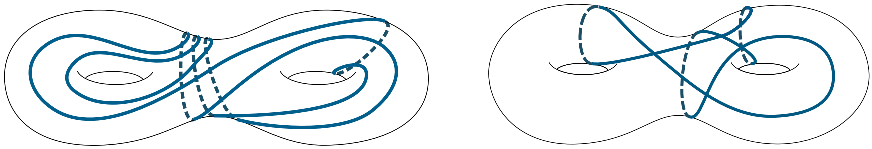


Figure 9: On the left, a simple geodesic. On the right, a non-simple geodesic (has self-intersections).

Theorem 3.2 (Huber [2])**.** *The number of closed geodesics of length at most L on a hyperbolic surface of genus $g \geq 2$ is asymptotic to $\frac{e^L}{L}$.*

There are several interesting things about this deep theorem. It tells us that the number of closed geodesics grows exponentially fast, which is much faster than the quadratic growth we saw for flat tori (there is a competing factor of L in the denominator, but this gets trounced by the e^L term in the numerator). Also, there are no constants that depend on the particular surface, or even on the genus of the surface. Asymptotically, the behavior is the same for any hyperbolic surface!

Now let's look at just the geodesics that don't self-intersect.

Theorem 3.3 (Mirzakhani [4])**.** *The number of simple closed geodesics of length at most L on a hyperbolic surface of genus $g \geq 2$ is asymptotic to*

$$c \cdot L^{6g-6}.$$

Here the constant c depends on the particular surface.

The growth rate here is much slower than the exponential rate for all closed geodesics (so most long closed geodesics must have self-intersections). But the rate is faster than the rate for (families of parallel) closed geodesics on the flat torus.

The proof of the above theorem leads one to study what is called the "moduli space" of all hyperbolic structures of a fixed genus. Each point of this space represents an individual hyperbolic surface. The moduli space itself has rich and intricate geometry.

4 Ongoing research

Researchers continue to explore generalizations and variations of the theorems above. New proofs of Mirzakhani's Theorem 3.3 have been developed, some of which lead to more precise counts [1]. Techniques from Mirzakhani's work can also be applied to the study of the "frequencies at which a hyperbolic surface wants to vibrate" (mathematicians call this the *spectrum of the Laplacian*, see [3]).

We saw that hyperbolic geometry is the natural geometry of surfaces of genus $g \geq 2$. But we can also force the flat geometry that was very natural for the torus onto higher genus surfaces. The price we pay is that the geometry will not be completely homogenous; there will be some points near which the space does not look flat. For instance, instead of taking a hyperbolic octagon as in Figure 6a, one could take an ordinary Euclidean regular octagon, and then glue opposite sides together. The result is a genus two surface that looks flat everywhere except at a single point that comes from the vertices of the octagon. Such an object is called a *translation surface*; the study of these forms an exciting active area of research.

Image credits

Figure 3 the left-hand image is a screenshot of gameplay from
`https://www.retrogames.cz/play_125-Atari7800.php?language=EN`

Figure 4 is from `https://www.d.umn.edu/~ddunham/mathhoriz/paper.html`
Reproduced with permission from Doug Dunham.

Figure 6b, Figure 9 created by Jenya Sapir. Used with permission.

References

[1] A. Eskin, M. Mirzakhani, and A. Mohammadi, *Effective counting of simple closed geodesics on hyperbolic surfaces*, Journal of the European Mathematical Society **24** (2022), no. 9, 3059–3108.

[2] H. Huber, *Über eine neue Klasse automorpher Funktionen und ein Gitterpunktproblem in der hyperbolischen Ebene. I*, Commentarii Mathematici Helvetici **30** (1956), 20–62.

[3] M. Lipnowski and A. Wright, *Towards optimal spectral gaps in large genus*, 2021, arXiv:2103.07496.

[4] M. Mirzakhani, *Growth of the number of simple closed geodesics on hyperbolic surfaces*, Annals of Mathematics Second Series **168** (2008), no. 1, 97–125.

Benjamin Dozier *is an assistant professor of mathematics at Cornell University.*

Mathematical subjects
Geometry and Topology

DOI
10.14760/SNAP-2022-013-EN

Snapshots of modern mathematics from Oberwolfach provide exciting insights into current mathematical research. They are written by participants in the scientific program of the Mathematisches Forschungsinstitut Oberwolfach (MFO). The snapshot project is designed to promote the understanding and appreciation of modern mathematics and mathematical research in the interested public worldwide. All snapshots are published in cooperation with the IMAGINARY platform and can be found on www.imaginary.org/snapshots and on www.mfo.de/snapshots.

ISSN 2626-1995

Mathematisches Forschungsinstitut
Oberwolfach gGmbH
Schwarzwaldstr. 9 – 11
77709 Oberwolfach
Germany

Director
Gerhard Huisken