Chapter 17

Geometry and Gravitation

17.1 Introduction to Geometry

Geometry is one of the oldest branches of mathematics, competing with number theory for historical primacy. Like all good science, its origins were based in observation and, with historical hindsight, we realize that the evident truths discovered by early geometers were really a result of limited perspective. But like them, for our discussion, we will take certain ideas as evident and as the basis for what we understand. The idea of the point and connected sets of points and particularly the idea of the straight line. As is evident from our discussion of Special Relativity, see Sections 11.3.5 on page 265 and 12.3 on page 280, we take the straight line to be the shortest distance between two points in space and the longest distance between two events in space-time.

Geometry developed from the need to measure land surfaces for agricultural purposes. The geometry that developed was what we now call plane geometry and the basis for it was first clearly articulated by Euclid and thus the name Euclidean geometry. Euclid set the foundation for plane geometry by means of a set of axioms, evident truths. Modern formulations of geometry realize that there are consistent systems that do not have the same set of axioms. The question then becomes one of choice or appropriateness. In fact, if the early geometers had considered the geometry that is appropriate to large distances on the earth, they would have developed a geometry that was not Euclidean. This alternative geometry is well known and is called spherical geometry. It differs from the Euclidean with the replacement of one axiom, the axiom of parallels. In Euclidean geometry, the axiom of parallels states that given a straight line and a point not on that line that there

is one and only one straight line through that point that never touches the original line no matter how far the lines are extended and that line is called parallel. In spherical geometry, the straight lines are the arcs of great circles, circles on the surface whose center is the center of the sphere. A point to note is that the center of the sphere is not on the surface. In the case of the sphere, all straight lines through a point not on the original line meet the original line, in fact twice. There is a line through a point not on the original line that requires the greatest distance to the nearest intersection of extension before meeting. This line at that point is said to be locally parallel to the original line and this line is unique.

Because in spherical geometry, the axiom of parallels is no longer valid; many of the usual rules of Euclidean geometry no longer hold. The sum of the interior angles of a triangle do not add to π but is always greater than π. Think of a triangle on the sphere of the earth formed by the equator and two lines of longitude. At the equator the two lines are locally parallel and the angle between them and the equator is $\frac{\pi}{2}$. They will meet at the north or south pole at some non-zero angle and thus the sum of all three angles is greater than π. Make a square, a four sided figure of equal length sides with all sides meeting at right angles, on the surface. In contrast to the Euclidean case, it does not stop and start at the same point but over-closes, two of the legs of the square meet before the full side length is achieved. A third test is to make a circle, a set of points that are equidistant from some point, on the earth. The ratio $\frac{\text{circumference}}{2r}$, where r is the distance from the point to the circle defined as the radius, is less than π. To most people this is trivial. The problem is that we are measuring on the surface of the sphere. In the underlying three dimensional space in which the sphere is imbedded, the geometry is Euclidean and the world makes sense. For instance, if, instead of the distance as measured from the center on the sphere, the distance used, r', is the distance to the axis that is perpendicular to the plane of the circle passing through the center, the usual result that the ratio $\frac{\text{circumference}}{2r'}$ is π. Because this first identification of a non-Euclidean geometry was on an imbedded sphere, these non-Euclidean geometries are now called curved spaces. This is an unfortunate accident of history as we will discuss shortly but it is so prevalent that everyone uses these terms and we will continue to use this nomenclature. Geometries are flat, Euclidean, or curved, non-Euclidean, with an example being a two dimensional spherical surface imbedded in a flat, Euclidean, three space.

17.2 Gaussian Curvature*

The next significant step in the development of modern geometry was taken by the great mathematical physicist Carl Friedrich Gauss. Gauss was interested in the general problem of the shape of a two dimensional surfaces in our three dimensional space. Instead of a plane, the basis for Euclidean geometry, or a sphere the basis for spherical geometry, consider a two dimensional surface in the shape of a pear imbedded in three space. At a point on the surface there are various curvatures, using an intuitive idea that will be articulated with greater care shortly. At the points near the bottom or top of the pear the surface is much like that of a sphere while in the neck region there is a another type of bend. Also at any point, if the region of examination is small enough, the geometry acts as if it is Euclidean or flat, i. e. for a small enough triangle, the sum of the interior angles of triangles is π.

In order to proceed, Gauss needed a definition of curvature. It had to be local, at a point, and agree with our intuitive notions about curvature. The basic idea is that, on a curved surface, as you move through nearby points on the surface, the normal to the surface changes direction. Thus he produced the following construction: as you move over an element of area on the surface, the tip of the unit normal will paint an area on the unit sphere, see Figure 17.1 on page 365. the curvature at a point on the surface

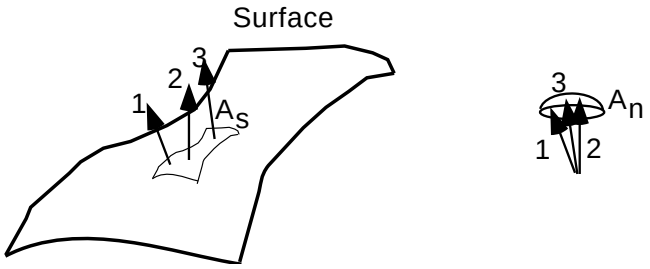

Figure 17.1: **Gauss's Definition of Curvature** Gauss defined curvature as the ratio of the area generated by the tips of the unit normals, $\mathrm{A_n}$, for an element of area, $\mathrm{A_s}$, on the surface as the area on the surface, $\mathrm{A_s}$, goes to zero, $K_{\mathrm{G}} \equiv \lim_{\mathrm{A_s}\to 0} \frac{\mathrm{A_n}}{\mathrm{A_s}}$.

is the ratio of the area generated by the tips of the unit normals, $\mathrm{Area_n}$, for an element of area, $\mathrm{Area_s}$, on the surface as the area on the surface goes to

zero,

$$K_{\rm G} \equiv \lim_{{\rm Area_s}\to 0} \frac{{\rm Area_n}}{{\rm Area_s}}. \tag{17.1}$$

In order to appreciate the subtlety of this construction, let's consider several examples. A flat surface has no curvature since the normal is always the same and thus the $\rm Area_n$ that is generated is that of a point and thus the $\rm Area_n$ is zero. On a sphere of radius r, using the usual spherical coordinates, θ and ϕ, a patch of $\rm Area_s = r^2\delta\theta\delta\phi$ and the normal which is the radius vector generates an $\rm Area_n = \delta\theta\delta\phi$. Thus the curvature is $\frac{1}{r^2}$. This construction shows that this idea of curvature makes sense and that the limit defining it exists for reasonably shaped surfaces[1]. Also note that in the limit of large r the curvature is zero. Now consider a point on the neck of the pear mentioned above. Another example and probably easier to visualize is a Pringle potato chip, see Figure 17.2 on page 366.

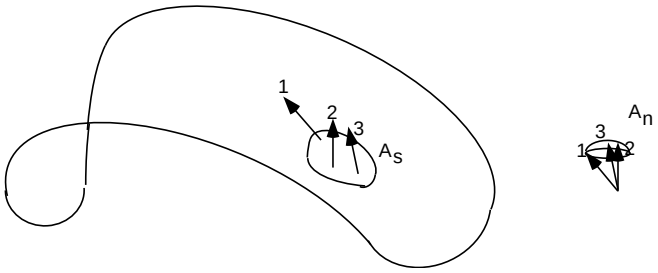

Figure 17.2: **Curvature of a Pringle** A Pringle is an example of a negatively curved surface. The area, $\rm A_n$, generated by the normals to the surface, $\rm A_s$, at any point is not zero. The difference between this case and the sphere though is that the area, $\rm A_n$, is oppositely oriented from that of the area on the surface, $\rm A_s$, i. e. a right hand coordinate plane on $\rm A_s$ generates a left handed coordinate system on $\rm A_n$, see Section 17.3 on page 366.

17.3 Example of negative curvature: the Pringle*

I have no idea how Pringles are manufactured, but I will construct my Pringle-like surface by taking a circle of radius R_1 centered on the origin in the two plane, (x, z), displacing it by R_2, $R_2 > R_1$, and then making this circle a surface of revolution about the z axis. This generates a torus

[1]The definition of reasonably shaped is stated as 'the surface is locally flat', flat in the sense of the usual idea of the existence of the derivative.

or donut shape. We can take a segment of the inner surface, the surface toward the z axis, as our Pringle.

The advantage of this construction is that the labeling of points on the surface and the properties of the normal vector can be determined easily. For example, a point on the surface can be determined from the angle around the original circle as measured from the top most point, θ, and the angle of rotation of the circle around the z axis, ϕ both ranging from zero to 2π.

Using these coordinates, a point on the surface is at

$$\begin{aligned} x &= [R_2 - R_1 \sin\theta]\cos\phi \\ y &= [R_2 - R_1 \sin\theta]\sin\phi \\ z &= R_1 \cos\theta, \end{aligned} \tag{17.2}$$

and the area, A_s, generated by incrementing the two coordinates which are orthogonal is $[R_2 - R_1 \sin\theta]R_1\delta\theta\delta\phi$. The unit normal vector is along the line from the center of the circle at ϕ and the point on the surface or $\hat{n} = (-\sin\theta\cos\phi)\hat{x} + (-\sin\theta\sin\phi)\hat{y} + \cos\theta\hat{z}$. As the area A_s is swept out, the change in the unit normal is $\delta\hat{n} = (-\cos\theta\cos\phi\delta\theta + \sin\theta\sin\phi\delta\phi)\hat{x} + (-\cos\theta\sin\phi\delta\theta - \sin\theta\cos\phi\delta\phi)\hat{y} + (-\sin\theta\delta\theta)\hat{z}$. Again the lines swept out by the coordinate increments are orthogonal and the area, A_n, generated is $\sin\theta\delta\theta\phi$. The Gaussian curvature is $|K_G| = \frac{\sin\theta}{(R_2 - R_1\sin\theta)R_1}$.

I have put absolute value signs on this result because the curvature in this case is actually negative. You should realize that, if we choose the coordinate directions in A_s to be right handed in the sense that the normal is outward and generated by rotating directed lines at constant θ into lines of constant ϕ, then the area A_n is left handed in the sense that the image traces of constant θ and ϕ are now left handed. This change in orientation of the areas is the indicator that this curvature is negative and thus

$$K_G = -\frac{\sin\theta}{(R_2 - R_1\sin\theta)R_1}. \tag{17.3}$$

There are other features of this result that are worth commenting on. The obvious result that the curvature is independent of ϕ is expected. More intriguing is the θ dependence, $K_G(\theta)$. Note that, had we done the analysis for the region $\pi < \theta < 2\pi$, the orientation of the image plane would have been the same as the original element of surface and thus, as given by Equation 17.3 on page 367, the curvature is positive. At $\theta = \frac{\pi}{2}$, the curvature is $K_G(\frac{\pi}{2}) = \frac{1}{(R_2 - R_1)R_1}$. The square root of the inverse of the curvature is the geometric mean radius of the two circles that make up the surface at this point, the radius of our original circle and the radius of the surface from

the axis of symmetry, the z axis. This same observation is also valid for the $\theta = \frac{3\pi}{2}$. This is a general result that we will deal with in more detail in the next section, Section 17.4 on page 368. The other interesting set of points is at $\theta = 0$ and $\theta = \pi$. Here, the curvature is zero. This can be looked at in two ways. These points are the transition points from the region of negative curvature, the inside of the torus, and the region of positive curvature, the outside of the torus. Since we expect the curvature to smooth, it is required that the curvature vanish at these points. More significantly, This region really is flat in the sense that it is Euclidean.

Think of a cylinder. The curvature of a cylinder is zero – the normal moves along a line as you move around the cylinder but does not change as you move along the axis of the cylinder. Thus, the area, A_n is zero. It is also important to note that the geometry of the cylinder is the same as that of a flat plane; you can unroll the cylinder onto a flat plane. You can do your geometry in the flat plane with the straight lines being the same as usual and the geometry is Euclidean, interior angles of triangles add to π. Thus the cylinder can be covered entirely by a single flat map. You cannot cover a curved surface entirely with a single flat map. You can cover it locally but at some places the distortion caused by the mapping becomes so severe that points are mapped to lines and visa versa. Think of a map of the earth. The usual atlas projection treats the poles, points, as lines. If you exclude the anomalous points by restricting the range of the coordinates you do not cover the earth with a single map but need more than one flat map. This is also a general property of non-Euclidean spaces. Is a cone flat or curved?

17.4 Curvature and Geodesics

In order to proceed further, we will have to examine the general issue of curves in the surface. An arbitrary path connecting two points in the surface can have lots of turns and bends. There are two sources of these, the bends of the surface and the bends of the path within the surface. We can eliminate the bends within the surface by considering only straight line paths between the points. These, by definition, are the shortest distance paths between the points. Since these may be very curved instead of calling them straight lines a better name is geodesic. One of many theorems of the theory of surfaces is that these are unique. These geodesic paths thus contain the bends of the surface and only those bends. In Section 17.5 on page 370, we will develop a specific differential condition for geodesics that is valid in any coordinate system. For now, we will continue with the more intuitive notions of their

properties.

Remembering that our two surface is imbedded in a flat three space, we can identify three directions at any point on the path, the direction along the local tangent to the path, the direction in the surface perpendicular to that direction (Don't forget that, at a point on the surface for a small enough region, the surface is flat and thus this direction is known. To find it, pick another point on the surface not on the original straight line and draw another geodesic through it. These two paths determine a plane, the tangent plane. All geodesics through p share this tangent plane.), and the direction that is perpendicular to these two. This last direction is locally perpendicular to the surface in the sense that the two other directions have generated the tangent plane at the point. This direction is called the normal direction. We already took advantage of these ideas in the identification of the normal to the surface in the previous section, Section 17.2 on page 365, in which we constructed the Gaussian curvature.

In the neighborhood of the point, the original geodesic is contained in the plane formed from the normal direction and the tangent direction of the geodesic. In the neighborhood of the point p, pick two other points on the original geodesic on opposite sides of p but near p, which will all be in that plane. As is well known from analytic geometry, three points determine a circle. This circle is called the osculating circle. Osculating is from the latin word for "kissing." In some sense, the idea of the osculating circle is the next step up from the tangent. The tangent is determined by two nearby points, determines a magnitude and a direction, and in the limit leads to the concept of the derivative. The osculating circle is determined by three nearby points and utilizes the second derivative, the difference in two tangents, the tangents formed from the original point and the other two points. The inverse of the radius of this osculating circle is called the curvature of the original geodesic. Remember that by using geodesics, there is no bending in the surface. All the bending is due to the surface. There is another geodesic through p that is orthogonal. On that geodesic, construct an osculating circle. Thus at p, for a pair of orthogonal geodesics, there are two osculating circles, one for each of the mutually orthogonal geodesics. As the orientation of this orthogonal pair of geodesics is varied, there will be a direction in which the curvature for each of the orthogonal geodesics will be an extremum. There is no other orientation of the geodesics that have extremum curvatures except trivial variations on this orientation. This last result is called Euler's Theorem. Gauss showed that the Gaussian curvature of the surface as defined in Section 17.2 on page 365 is the product of these

two extremum curvatures,

$$K_G = \frac{1}{R_1 \times R_2}, \tag{17.4}$$

where R_1 and R_2 are the radii of the osculating circles. In addition, the sign of the curvature is determined by the relationship of the two osculating circles. The curvature is positive if both the osculating circles are on the same side of the surface. This is the case for the sphere as discussed earlier. For the Pringle, Section 17.3 on page 366, on the inner edge, the osculating circles are on opposite sides of the surface and this is the signature of negative curvature.

As is always the case with the Gaussian curvature, this curvature is a basic property of the surface and does not depend on the coordinate system that we used to make the construction. Granted that the construction of the curvature is most readily done in a coordinate system that is based on a system of orthogonal geodesics, it is still clear from the nature of the Gauss map and Equation 17.1 on page 366 that the coordinates make the construction possible by staking out the grid but that the local value of the curvature is the same regardless of the coordinate system used. In fact the coordinate system that was used for the torus, Equation 17.2 on page 367, are not geodesic coordinates; the lines of constant ϕ are geodesics but the lines of constant θ are not. This issue will be discussed in much greater detail later, see Sections 17.5 on page 370, 17.6 on page 371.

17.5 The Theorema Egregium and the Line Element

As is clear from Section 17.4 on page 368, Gauss made an extensive study of the nature of surfaces imbedded in a Euclidean three space. He is responsible for many of the insights and theorems that govern understanding of these surfaces. He was, of course, interested in two surfaces imbedded into the larger three space. He recognized the important role of curvature in defining the nature of the surface; to within an orientation and a translation, the surface is determined by its curvature. His most famous theorem in the theory of surfaces was so striking to him that, when he recognized its implications, he gave it the title of the Theorema Egregium. A direct translation of the latin would call this the egregius theorem. The modern sense of egregius: outstandingly bad is not the original meaning. The original use of the word was in the sense of outstandingly good and is what is intended in

the latin. It was later usage that lead to the current interpretation of egregious as outstandingly bad, see [OED 1971]. It seems that modern young people are not the first ones to reverse the meaning of bad and good when describing things. Regardless, the point of Gauss' name for the theorem was in the sense of outstandingly good. Maybe a better translation would be the Extraordinary Theorem.

This theorem proved that all the important properties of the surface could be developed from information that is intrinsic to the surface and did not need to use properties that were determined by the imbedding of the surface in a Euclidean three space or the coordinate system that was used to do the construction of the Gauss map. The only element that is needed to construct curvature is the length of the line element in whatever coordinate system is being used to identify places on the two plane.

The other amazing fact is the realization of Riemann that these techniques developed by Gauss carry over to manifolds of any number of dimensions, Section 17.6 on page 371. The theorem's proof is rather tedious and not really enlightening except in its use of intermediate elements that are very important in our later study of geometry in higher dimensions. We will proceed to look at the situation in higher dimensions directly introducing the concepts as required.

In other words, if when you begin to label points on the surface with some set of coordinate labels and, if at the same time, you determined the actual lengths separating nearby coordinate points, you would have all the information that you need to determine the curvature. For an n-dimensional space, each place is designated by an n-vector, his object is called the metric and is designated g_{ij} and the length of a line element the two nearby places $(x_{10}, x_{20}, \cdots, x_{n0})$ on the surface

17.6 Geometry in Four or More Dimensions*

Riemann showed that the methods of analysis used by Gauss for two planes embedded in a three space could be extended to higher dimensional spaces and, in particular, that a space of any dimension could possess non-Euclidean features and that, like the case of the two plane, the non-Euclidean aspects can be manifest by measurements contained to the two planes that constitute the higher dimensional space.

The fundamental construction element in the analysis of the higher dimensional case is called the Riemann tensor. For our purposes it is sufficient to define

The metric is g_{ij} is called the first fundamental form and is defined as

$$g_{\mu\nu} \equiv \frac{\partial x^\gamma}{\partial q^\mu}\frac{\partial x_\gamma}{\partial q^\nu} \tag{17.5}$$

The second fundamental form is $b_{\mu\nu}$ is

$$b_{\mu\nu} = \frac{\partial^2 x^\gamma}{\partial q^\mu \partial q^\nu} n_\gamma \tag{17.6}$$

The mixed Riemann curvature tensor is

$$R^\mu_{\nu\gamma\rho} = \frac{\partial \Gamma^\mu_{\nu\rho}}{\partial q^\gamma} - \frac{\partial \Gamma^\mu_{\nu\gamma}}{\partial q^\rho} + \Gamma^\eta_{\nu\rho}\Gamma^\mu_{\eta\gamma} - \Gamma^\eta_{\nu\gamma}\Gamma^\mu_{\eta\rho} \tag{17.7}$$

where

$$\Gamma^\mu_{\nu\gamma} = g^{\mu\eta}\Gamma_{\nu\gamma,\eta} \tag{17.8}$$

where

$$\Gamma_{\nu\gamma,\eta} = \frac{1}{2}\left\{\frac{\partial g_{\eta\gamma}}{\partial q^\nu} + \frac{\partial g_{\eta\nu}}{\partial q^\gamma} - \frac{\partial g_{\nu\gamma}}{\partial q^\eta}\right\} \tag{17.9}$$

and the Riemann curvature tensor is

$$R_{\mu\nu\rho\gamma} = R^\eta_{\nu\rho\gamma} g_{\eta\mu} \tag{17.10}$$

The Ricci tensor is

$$R_{\mu\nu} \equiv R^\rho_{\mu\rho\nu} \tag{17.11}$$

The symmetries of the Riemann curvature tensor are:

$$R_{\lambda\mu\nu\kappa} = R_{\nu\kappa\lambda\mu} \tag{17.12}$$

$$R_{\lambda\mu\nu\kappa} = -R_{\mu\lambda\nu\kappa} = -R_{\lambda\mu\kappa\nu} = R_{\mu\lambda\kappa\nu} \tag{17.13}$$

The cyclicity condition:

$$R_{\lambda\mu\nu\kappa} + R_{\lambda\kappa\mu\nu} + R_{\lambda\nu\kappa\mu} = 0 \tag{17.14}$$

17.7 Coordinate Labels in General Relativity

17.8 Einstein Equations

$$R^{\mu\nu} - \frac{1}{2}g^{\mu\nu}R = -\frac{8\pi \mathrm{G}}{c^4}T^{\mu\nu} \tag{17.15}$$