Differential Geometry
Lia Vas

Surfaces. Normal and Gaussian curvatures

When considering curves, we have seen that the curvature measures the extent of twisting and turning of a curve in space. We now turn to surfaces, two-dimensional objects in three-dimensional space and examine how the concept of curvature translates to surfaces.

In Calculus 3, you have encounter surfaces defined as graphs of real valued functions of two variables $z = f(x, y)$. This function also can take the form $x = f(y, z)$ or $y = f(x, z)$. In some cases, this function is given implicitly as $F(x, y, z) = 0$. For example, a sphere of radius a is given by $x^2 + y^2 + z^2 = a^2$ and it is impossible to get a single two variable function that would describe the whole sphere. Cylinder $x^2 + y^2 = a^2$ is another such example. Representing a surface using **parametric equations** encompasses all of the above scenarios. Parametric equations of a surface have the form

$$x = x(u, v) \quad y = y(u, v) \quad z = z(u, v).$$

The variables u and v are **parameters** of the above equations.

Thus, a parametric surface is represented as a vector function of two variables, i.e. the **domain** D consisting of all possible values of parameters u and v is contained in $\mathbb{R}^2$. The **range** of the surfaces is contained in the three dimensional space $\mathbb{R}^3$. Thus, a surface $\mathbf{x}$ is a mapping of D into $\mathbb{R}^3$. This is denoted by $\mathbf{x} : D \to \mathbb{R}^3$. The vector function $\mathbf{x}$ can also be represented as

$$\mathbf{x}(u, v) = (x(u, v),\ y(u, v),\ z(u, v)).$$

Notice an analogy with curves. We can think of curves as one-dimensional objects in three-dimensional space and surfaces as two-dimensional objects in three dimensional space. Thus, a curve can be described using a **single** parameter t. Surface, on the other hand, is described using **two** parameters u and v.

	Mapping	Dimension	Parameter(s)	Equations
Curve	$\boldsymbol{\gamma} : (a, b) \subseteq \mathbb{R} \to \mathbb{R}^3$	1	t	$\boldsymbol{\gamma}(t) = (x(t), y(t), z(t))$
Surface	$\mathbf{x} : D \subseteq \mathbb{R}^2 \to \mathbb{R}^3$	2	u, v	$\mathbf{x}(u, v) = (x(u, v),\ y(u, v),\ z(u, v))$

For example, the surface given in the form $z = f(x, y)$ can always be parametrized as $\mathbf{x} = (x, y, f(x, y))$. We review an important example from Calculus 3.

Planes. The general equation of a plane is

$$ax + by + cz + d = 0.$$

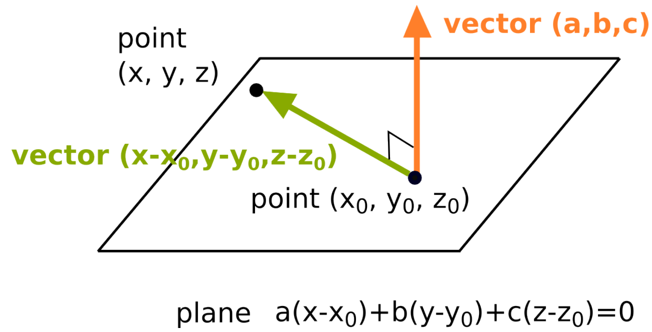
A plane is uniquely determined by a point in it and a vector perpendicular to it. The equation that describes any point $\mathbf{x} = (x, y, z)$ in the plane through a point

$\mathbf{x}_0 = (x_0, y_0, z_0)$ perpendicular to a vector $\mathbf{a} = (a, b, c)$ is

$$\mathbf{a} \cdot (\mathbf{x} - \mathbf{x}_0) = 0$$

The above vector equation of the plane has the following scalar form.

$$a(x - x_0) + b(y - y_0) + c(z - z_0) = 0.$$



If $c \neq 0$, the plane $ax+by+cz = d$ can be parametrized by $(x, y, \frac{1}{c}(d-ax-by))$. For example, the plane $x - 3y + z = 2$ can be parametrized by $(x, y, 2 - x + 3y)$. If $a \neq 0$, one can solve for x and use y and z as parameters. For example, the plane from the previous example can also be parametrized by $(2 + 3y - z, y, z)$. If $b \neq 0$ one can solve for y and use x and z as parameters. For example, the plane $2x + y = 4$ can be parametrized by $(x, 4 - 2x, z)$.

The tangent plane of a surface at a point.

For a parametric surface

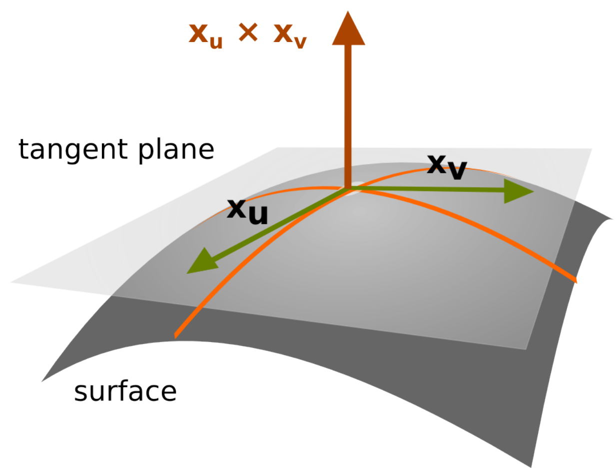
$$\mathbf{x} = (x(u, v), y(u, v), z(u, v)),$$

the derivatives $\mathbf{x}_u$ and $\mathbf{x}_v$ are vectors in the tangent plane. Thus, their cross product

$$\frac{\partial \mathbf{x}}{\partial u} \times \frac{\partial \mathbf{x}}{\partial v} = (x_u, y_u, z_u) \times (x_v, y_v, z_v)$$

is perpendicular to the tangent plane.

If a surface is given by implicit function $F(x, y, z) = 0$, then this cross product also corresponds to the **gradient** ∇F of F.



A convenient parametrization of some surfaces requires a change of coordinates. We review cylindrical and spherical coordinates next.

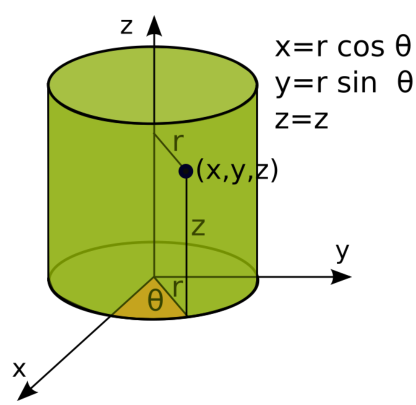
Cylindrical coordinates.

$$x = r\cos\theta, \quad y = r\sin\theta, \quad z = z.$$

Here x and y are converted using polar coordinates and the only change in z may come just from changes in x and y. Note that

$$x^2 + y^2 = r^2$$

in these coordinates.

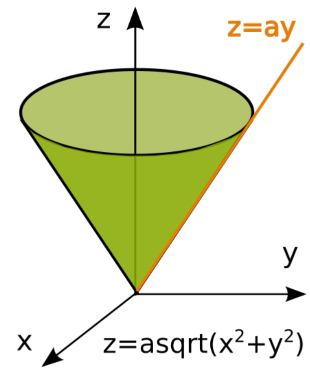


Cylindrical coordinates can be used for parametrization of surfaces of revolution and cylindrical surfaces.

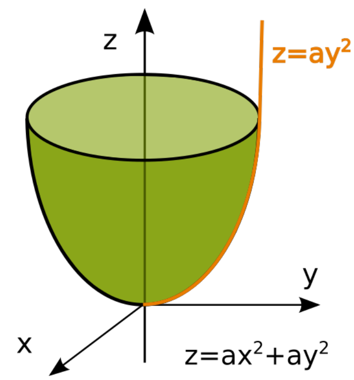
Surfaces of revolution $z = f(\sqrt{x^2 + y^2})$. To parametrize such a surface, note that $\sqrt{x^2 + y^2}$ is r in cylindrical coordinates. Thus, you can use

$$\mathbf{x}(r, \theta) = (r\cos\theta, r\sin\theta, f(r))$$

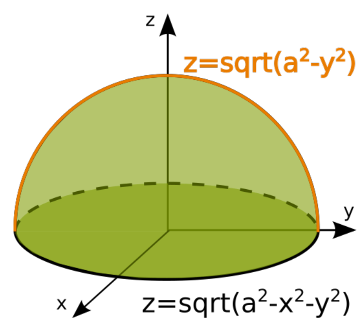
as your parametrization. We consider some examples.



1. The cone $z = a\sqrt{x^2+y^2} = ar$ is obtained by rotating the line $z = ay$ about the z-axis. It can be parametrized by
$$\mathbf{x} = (r\cos\theta, r\sin\theta, ar).$$
Note that this cone can also be parametrized by $(x, y, a\sqrt{x^2+y^2})$. However, the first parametrization has much simpler derivatives as well as much nicer bounds for the parameters r and θ when we integrate this cone over some natural regions.

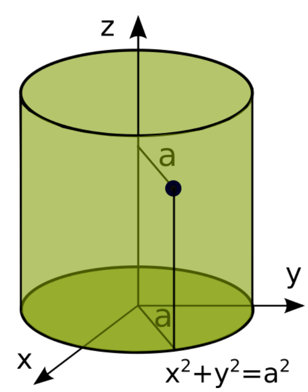


2. A paraboloid $z = ax^2 + ay^2$ is obtained by rotating the parabola $z = ay^2$ about the z-axis. It can be parametrized by
$$\mathbf{x} = (r\cos\theta, r\sin\theta, ar^2).$$
Note that it can also be parametrized by $(x, y, ax^2 + ay^2)$.



3. The upper half-sphere $z = \sqrt{a^2-x^2-y^2}$ is obtained by rotating the half-circle $z = \sqrt{a^2-y^2}$ about the z-axis. It can be parametrized by
$$\mathbf{x} = (r\cos\theta, r\sin\theta, \sqrt{a^2-r^2}).$$

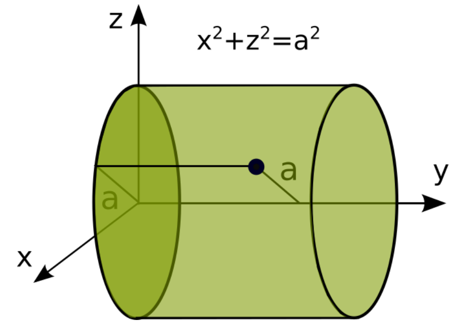
Cylindrical surfaces. Cylindrical surfaces are given by an equation in which at least one of the three variables x, y or z is not present. For example, $F(x, y) = 0$. To graph this surface, graph the curve with the implicit equation $F(x, y) = 0$ in the xy-plane and translate the graph along the z-axis. Similarly, to graph a surface given by $F(x, z) = 0$, translate the curve $F(x, z) = 0$ in the xz-plane along the y-axis. We consider two such examples.



1. The surface given by the equation $x^2 + y^2 = a^2$ is the cylinder of radius a with a circular base and it is parallel with the z-axis. If using cylindrical coordinates, the value of r is constant and it is equal to a and the z-coordinate corresponds to the second parameter which we call h to indicate that it corresponds to the height of the point. Thus, this cylinder can be parametrized by
$$\mathbf{x} = (a\cos t, a\sin t, h).$$

2. The surface given by the equation $y^2 + z^2 = a^2$ is the cylinder of radius a with a circular base. It is parallel with the y-axis. Analogously as in previous example, this cylinder can be parametrized by

$$\mathbf{x} = (h, a\cos\theta, a\sin\theta).$$



Examples.

1. Parametrize the following surfaces. Describe the surfaces or sketch their graphs.

   (a) $2x + 3y + z = 6$ (b) $z = 9 - \sqrt{x^2 + y^2}$ (c) $x^2 + z^2 = 4$ (d) $z = y^2$

2. Find an equation of the tangent plane to $x^2 + z^2 = 4$ at $(0, 3, 2)$.

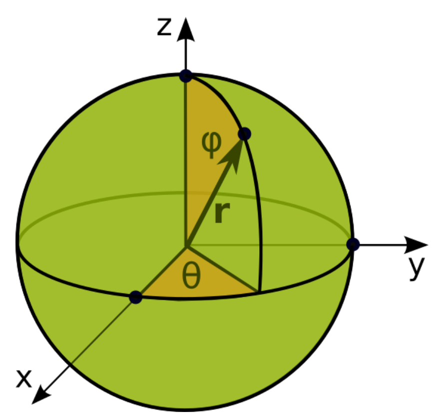
Solutions. 1. (a) This is the plane passing (3,0,0), (0,2,0), and (0,0,6). It can be parametrized as $\mathbf{x} = (x, y, 6 - 2x - 3y)$.
(b) This is the downwards cone with vertex (0,0,9). It can be parametrized as $\mathbf{x} = (r\cos\theta, r\sin\theta, 9 - r)$.
(c) This is the cylinder of radius 2, parallel with the y-axis. It can be parametrized as $\mathbf{x} = (2\cos\theta, h, 2\sin\theta)$.
(d) This is a cylindrical surface which can be obtained by translating the parabola $z = y^2$ in yz-plane along the x-axis. It can be parametrized by $\mathbf{x} = (x, y, y^2)$.
2. Use the parametrization $\mathbf{x} = (2\cos\theta, h, 2\sin\theta)$ from the previous problem and calculate $\mathbf{x}_\theta \times \mathbf{x}_h$ to be $(-2\cos\theta, 0, -2\sin\theta)$. At $(0, 3, 2)$, the values of parameters are $\theta = \frac{\pi}{2}$ and $h = 3$ so the normal vector is $(0, 0, -2)$. The equation of the plane is $0(x-0) + 0(y-3) - 2(z-2) = 0 \Rightarrow z = 2$.

Spherical coordinates. If $P = (x, y, z)$ is a point in space and O denotes the origin, let

- r denotes the distance of the point $P = (x, y, z)$ from the origin O. Thus,

$$x^2 + y^2 + z^2 = r^2;$$

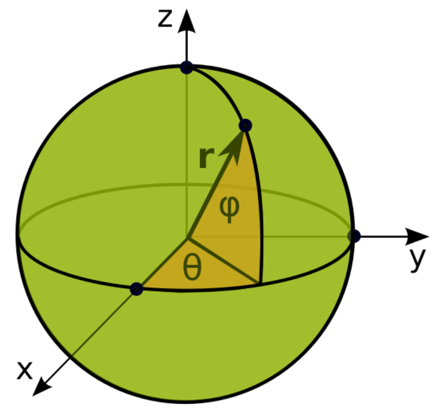
- θ is the angle between the projection of vector $\overrightarrow{OP} = (x, y, z)$ on the xy-plane and the vector $\overrightarrow{i} = (1, 0, 0)$ (positive x axis); and
- ϕ is the angle between the vector $\overrightarrow{OP}$ and the vector $\overrightarrow{k} = (0, 0, 1)$ (positive z-axis).



With this notation, spherical coordinates are (r, θ, ϕ). The conversion equations are

$$x = r\cos\theta\sin\phi \quad y = r\sin\theta\sin\phi \quad z = r\cos\phi.$$

In this parametrization, the north pole of a sphere centered at the origin corresponds to value $\phi = 0$, the equator to $\phi = \frac{\pi}{2}$ and the south pole to $\phi = \pi$. To match the geographical latitude (for which north and south pole correspond to $\frac{\pi}{2}$ and $\frac{-\pi}{2}$ and equator to $\phi = 0$), the angle ϕ is often considered to be the angle between the equator and the vector $\overrightarrow{OP}$. In this case, $\cos\phi$ and $\sin\phi$ are switched in the equations of the spherical coordinates and we obtain



$$x = r\cos\theta\cos\phi \quad y = r\sin\theta\cos\phi \quad z = r\sin\phi.$$

The angle θ corresponds to the geographical longitude and the angle ϕ corresponds to the geographical latitude.

The values of ϕ from the interval $[0, \frac{\pi}{2}]$ correspond to the points on the northern hemisphere, and the values of ϕ from $[-\frac{\pi}{2}, 0]$ correspond to the points on the southern hemisphere. The θ-interval $[0, \pi]$ corresponds to the eastern hemisphere, and the θ-interval $[-\pi, 0]$ to the western hemisphere. For example, the longitude of Philadelphia $75°10'$ west corresponds to $\theta = -75°10'$ and the latitude of Philadelphia $39°57'$ north corresponds to $\phi = 39°57'$.

Regular surfaces. Coordinate patches

In order to have a unit-length tangent vector at every point of a curve $\boldsymbol{\gamma} = \boldsymbol{\gamma}(t)$, we need to require that $\frac{d\gamma}{dt} \neq 0$. This condition also ensures that the curvature is well-defined and nontrivial. Analogously, for surfaces, we want to ensure that the tangent plane at every point is defined (i.e. that is not collapsed into a line or a point). Since the normal vector to the tangent plane of a parametric surface $\mathbf{x} = \mathbf{x}(u, v)$ is given by $\frac{\partial \mathbf{x}}{\partial u} \times \frac{\partial \mathbf{x}}{\partial v}$, we want to require a condition that guarantees that this vector is non-zero i.e.,

$$\frac{\partial \mathbf{x}}{\partial u} \times \frac{\partial \mathbf{x}}{\partial v} \neq \mathbf{0}.$$

The condition $\frac{\partial \mathbf{x}}{\partial u} \times \frac{\partial \mathbf{x}}{\partial v} \neq \mathbf{0}$ guarantees that the vectors $\frac{\partial \mathbf{x}}{\partial u}$ and $\frac{\partial \mathbf{x}}{\partial v}$ are not on the same line. Thus, they are linearly independent and they constitute a **basis of the tangent plane** and every other vector in the tangent plane can be represented as a linear combination of these two vectors. [1]

[1]**Linear Algebra background.** Consider two vectors $\mathbf{v}_1$ and $\mathbf{v}_2$ that do not lie on the same line. In this case, we say that $\mathbf{v}_1$ and $\mathbf{v}_2$ are **linearly independent**. Consider also the plane determined by these two vectors. For arbitrary vector $\mathbf{v}$ in the plane, we can consider the projection of $\mathbf{v}$ in direction of $\mathbf{v}_1$. This projection is a multiple of $\mathbf{v}_1$. Let a denote the multiplication factor. Consider also the projection of $\mathbf{v}$ in direction of $\mathbf{v}_2$ and let b denote the the multiplication factor. Thus,

$$\mathbf{v} = a\mathbf{v}_1 + b\mathbf{v}_2$$

The sum $a\mathbf{v}_1 + b\mathbf{v}_2$ is called a **linear combination** of $\mathbf{v}_1$ and $\mathbf{v}_2$. This shows that every vector in the plane that we consider can be expressed as a linear combination of $\mathbf{v}_1$ and $\mathbf{v}_2$. In this case, we say that $\mathbf{v}_1$ and $\mathbf{v}_2$ **generate** the plane.

Linearly independent vectors that generate a plane are called **a basis**. Consideration of projections above demonstrates that any two linearly independent vectors in a plane constitute a basis of the plane.

For example, vectors $(1, 0)$ and $(0, 1)$ are a basis of xy-plane (space $\mathbb{R}^2$): these two vectors are not colinear and every vector (x, y) is the linear combination $x(1, 0) + y(0, 1)$.

Thus, we consider just surfaces such that one can parametrize a region on the surface around every point with parametric equations $x = x(u, v)$, $y = y(u, v)$, $z = z(u, v)$ with the following properties.

- The functions $x = x(u, v)$, $y = y(u, v)$, $z = z(u, v)$ are continuous in both variables (thus, there are no gaps or holes) and one-to-one (no self-overlaps).
- The partial derivatives of $x = x(u, v)$, $y = y(u, v)$, $z = z(u, v)$ are continuous (thus, there are no corners or sharp turns and the surface is "smooth").
- The cross product $\frac{\partial \mathbf{x}}{\partial u} \times \frac{\partial \mathbf{x}}{\partial v}$ is not equal to $\mathbf{0}$ (thus, the tangent plane at each point is not collapsed into a line or a point).

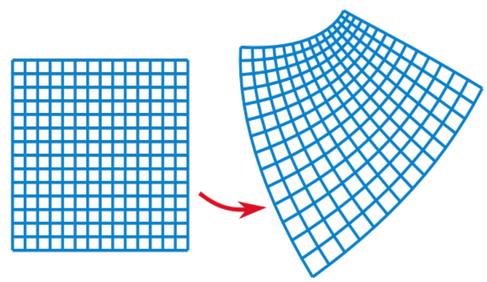
We refer to those surfaces as **regular** surfaces. Note that it may not be possible to describe the whole surface with a single set of equations which make it regular, but, in all the relevant cases, it will be possible to cover the entire surface by "patching" several different regular parametrizations together. This brings us to the concept of **coordinate patches**. You can think of them as

> *different regular parametrization which agree on the overlaps and, combined, cover the entire surface including any non-regular ("problematic") regions or points.*

More formally, the condition that the parametrizations agree on the overlaps is expressed as follows: if $\mathbf{x} = \mathbf{x}(u, v)$ is one coordinate patch defined on domain D and $\bar{\mathbf{x}} = (\bar{u}, \bar{v})$ is another defined on domain $\bar{D}$, that then the composite functions $\mathbf{x}^{-1} \circ \bar{\mathbf{x}}$ and $\bar{\mathbf{x}}^{-1} \circ \mathbf{x}$ are one-to-one and onto continuous functions with continuous derivatives on the intersection of D and $\bar{D}$. If this condition is satisfied, we say that the patches **overlap smoothly**.

This leads us to a more formal definition of a surface. We say that M is a **surface** if there is a collection of coordinate patches such that: (1) The coordinate patches cover every point of M and they overlap smoothly. (2) Every two different points on M can be covered by two different patches. (3) The collection of patches is maximal with respect to conditions (1) and (2). This means that if a patch overlaps smoothly with every patch in collection, then it is itself in the collection.

You can think of a coordinate patch as a way to have a coordinate system on the portion of the surface covered by the patch. The portion is regular so any bending caused by the bending of the surface is not relevant – the coordinate system functions just like having one in a flat uv-plane.



A coordinate patch

The same concepts can be defined in three-dimensional space. any three vectors $\mathbf{v}_1$, $\mathbf{v}_2$ and $\mathbf{v}_3$ that do not lie on the same plane are said to be **linearly independent**. Any other vector $\mathbf{v}$ can be expressed as a sum of its projections in directions of $\mathbf{v}_1$, $\mathbf{v}_2$ and $\mathbf{v}_3$

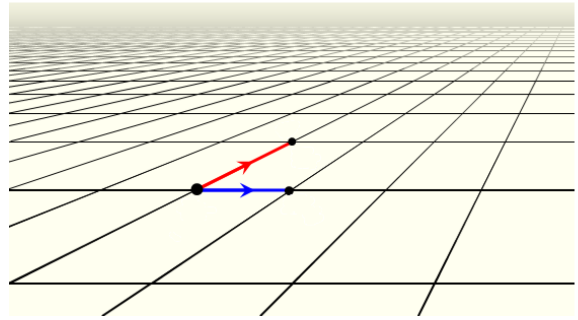
$$\mathbf{v} = a\mathbf{v}_1 + b\mathbf{v}_2 + c\mathbf{v}_3$$

i.e. as a **linear combination** of $\mathbf{v}_1$, $\mathbf{v}_2$ and $\mathbf{v}_3$. Thus, $\mathbf{v}_1$, $\mathbf{v}_2$ and $\mathbf{v}_3$ **generate** the space. Linearly independent vectors that generate the space are called **a basis**. By considering projections, any three linearly independent vectors in space are a basis of the space.

For example, vectors $(1, 0, 0)$, $(0, 1, 0)$ and $(0, 0, 1)$ are a basis of $\mathbb{R}^3$ because they are not in the same plane, and every vector (x, y, z) is the linear combination $x(1, 0, 0) + y(0, 1, 0) + z(0, 0, 1)$.

Another example of a basis of $\mathbb{R}^3$ are the vectors $\mathbf{T}$, $\mathbf{N}$ and $\mathbf{B}$ of the moving frame of a curve at any of its points.

In analogy to people, small in comparison to the size of the Earth, you can imagine small creatures inhabiting a surface. We shall refer to these creatures as "locals" and you can think that each patch is a neighborhood for a group of such locals. In the perspective of locals, the surface appears to be completely flat because they are too small to perceive any bending and their neighborhood looks like a flat uv-plane to them.



Locals' view of their home coordinate patch

Example. Let us consider the following parametrizations of the sphere $x^2 + y^2 + z^2 = a^2$ with radius $a > 0$.

1. Using x, y as parameters, one has to use at least two patches, $(x, y, \sqrt{a^2 - x^2 - y^2})$ for the upper hemisphere and $(x, y, -\sqrt{a^2 - x^2 - y^2})$ for the lower hemisphere. The values of the parameters x, y are $a \leq x \leq a$ and $-\sqrt{a^2 - x^2} \leq y \leq \sqrt{a^2 - x^2}$.

   Since the magnitude of $\mathbf{x}_x \times \mathbf{x}_y$ for both patches is $\frac{a}{\sqrt{a^2 - x^2 - y^2}}$, this magnitude becomes undefined when $y = \pm\sqrt{a^2 - x^2}$ because the denominator becomes zero. Because of this, these two patches are not regular on the equator and more patches are needed to patch the equator points. Solving the equation of the sphere for y (so $y = \pm\sqrt{a^2 - x^2 - z^2}$), we obtain another two patches $(x, \pm\sqrt{a^2 - x^2 - z^2}, z)$. They regularly cover the front and the back of the sphere but do not cover the circle $x^2 + z^2 = a^2$. Combining these four patches, we covered everything regularly except the intersections of two circles, the points $(0, a, 0)$ and $(0, -a, 0)$. Finally, to completely cover the sphere, we can patch the two holes with $(\pm\sqrt{a^2 - y^2 - z^2}, y, z)$. Thus, we can cover the entire sphere in six proper coordinate patches.

2. In cylindrical coordinates, one can parametrize the sphere by two patches, $(r\cos\theta, r\sin\theta, \pm\sqrt{9 - r^2})$ with $0 \leq \theta \leq 2\pi$ and $0 \leq r \leq a$. The positive sign of the last coordinate corresponds to the upper and the negative sign to the lower hemisphere. Just like for the first two patches in the previous example, these patches are regular everywhere except on the equator where they overlap. Thus, more than these two patches are needed to patch the sphere.

3. Using the "Calc 3" spherical coordinates, we can parametrize the sphere by

   $$\mathbf{x} = (a\cos\theta\sin\phi, a\sin\theta\sin\phi, a\cos\phi) \text{ with } 0 \leq \theta \leq 2\pi \text{ and } 0 \leq \phi \leq \pi.$$

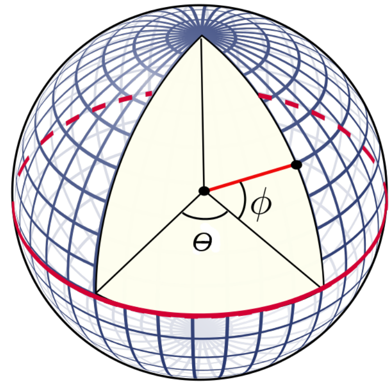
   In this parametrization, for a given ϕ value, both 0 and 2π values of θ correspond to the same point on the sphere. Because of this, you can consider $\theta = 0$ and $\theta = 2\pi$ as a self-overlap making this patch non-regular for these values of θ. Thus, we need more than this coordinate patch. For another patch, one can take the same formulas for x, y, and z, but with $-\pi \leq \theta \leq \pi$ for example. This patch is regular at the meridian which was problematic on the first patch.

   In addition to the meridian with $\theta = 0$ and 2π, something "fishy" is going on at the poles also. Indeed, note that the poles do not have a single value of θ corresponding to them – only the values of ϕ are fixed, $\phi = 0$ for the north pole and $\pi = \pi$ for the south pole. Because of this also, more parches are needed for the sphere to patch up these trouble spots.

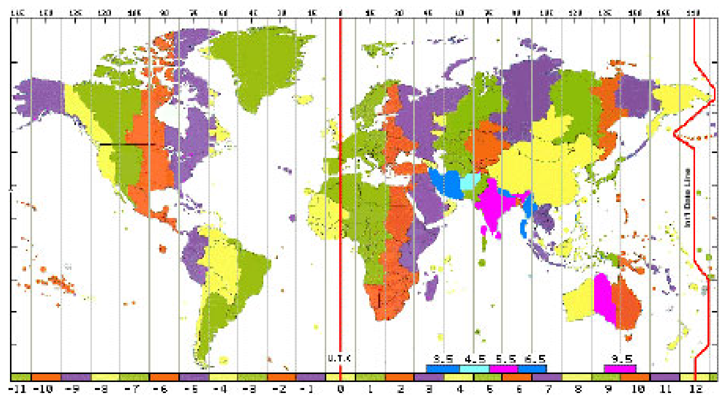
4. In the "Earth" spherical coordinates, we have that

$\mathbf{x} = (a\cos\theta\cos\phi, a\sin\theta\cos\phi, a\sin\phi)$ with

$-\pi \leq \theta \leq \pi$ and $\frac{-\pi}{2} \leq \phi \leq \frac{\pi}{2}$. Recall that the θ-interval $[0,\pi]$ corresponds to the eastern and $[-\pi, 0]$ to the western hemisphere.



The value $\theta = 0$ corresponds to the meridian passing the Royal Observatory in Greenwich, London. The time zones to the east are ahead of the time in London, and the time zones to the west are behind the time in London.

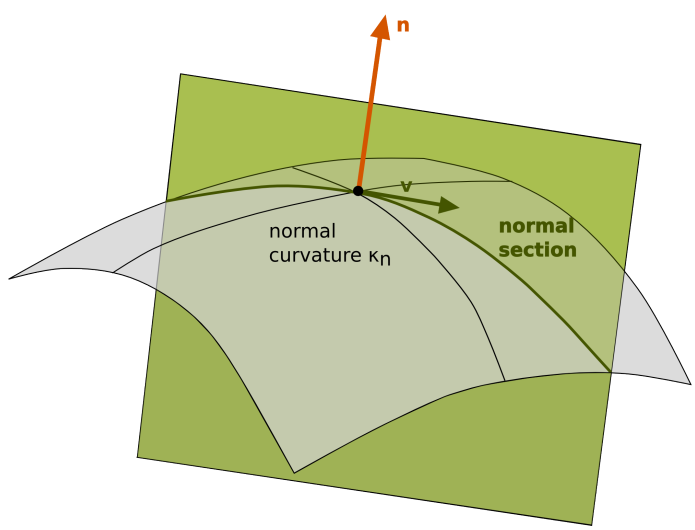


The self-overlap of this patch at $\theta = \pi$ and $\theta = -\pi$. This corresponds to the international time where the time is both 12 hours ahead and 12 hours behind the time in London (so it is "today and tomorrow at the same time").

Since the θ values are not uniquely determined at the poles, the time on the poles is not well-determined: it is every hour of the day at the poles.

This lack of regularity at the poles is reflected also in the values of the magnitude of the normal vector $\mathbf{x}_\theta \times \mathbf{x}_\phi$. The length of this vector can be computed to be $a^2|\cos\phi|$. Thus, the normal vector is zero if $\cos\phi = 0$. So, at point at which $\phi = \frac{\pi}{2}$ (which is the north pole) and at point at which $\phi = \frac{-\pi}{2}$ (which is the south pole), the normal vector of this patch is not well defined and, as a consequence, at these two points the geographical longitude is not uniquely defined.

The normal curvature and Gaussian curvature



A surface may curve by different extents in different directions at a point. Because of this, we first introduce the curvature in the direction of a given vector. This leads us to the **normal curvature** at a point **in the direction** of a given vector.

(1) To start, pick a point P on the surface and pick a vector $\mathbf{v}$ in the tangent plane at P.

(2) Consider the plane determined by $\mathbf{v}$ (thus passing P) and the normal vector of the tangent plane (thus perpendicular to the tangent plane). Let $\boldsymbol{\gamma}$ be the curve in the intersection of the surface and this plane. The curve $\boldsymbol{\gamma}$ is called the **normal section** at P in the direction of $\mathbf{v}$.

(3) Compute the curvature κ of $\boldsymbol{\gamma}$ at P. The **normal curvature in the direction of v** denoted by $\kappa_n(\mathbf{v})$ is taken to be

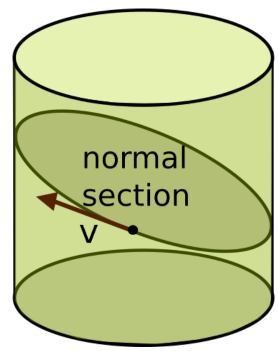
$$\kappa_n(\mathbf{v}) = \pm\kappa.$$

If the normal vector of the curve and the normal vector of the tangent plane are on the *same* side of the tangent plane, then $\kappa_n(\mathbf{v}) = \kappa$. If they are on opposite sides of the tangent plane, then $\kappa_n(\mathbf{v}) = -\kappa$. Otherwise $\kappa_n(\mathbf{v}) = 0$.

Examples. (1) Consider a **plane**. The normal section of the plane at any point in any direction is a straight line because any plane perpendicular to the given plane (equal to its own tangent plane) intersects the plane in a straight line. A line has zero curvature, so the normal curvature of a plane at any point in any direction is zero.

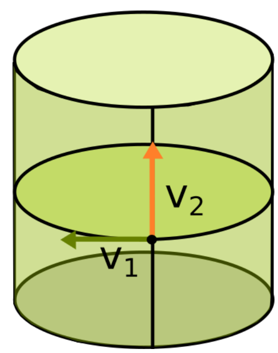
(2) Consider a **sphere** with radius a. Any plane perpendicular to the tangent plane at any point intersects the sphere in a circle containing the center of the sphere and, thus, having the radius equal to a. So, all normal sections are circles of radii a and, hence, with the curvatures of $\frac{1}{a}$. Thus, the normal curvature is $\pm\frac{1}{a}$.

(3) Consider the **cylinder** $x^2 + y^2 = a^2$.



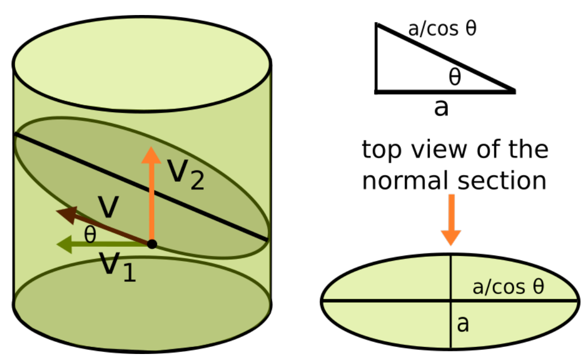
At any point, a plane perpendicular to the tangent plane and not parallel to the central axis intersects the cylinder in an *ellipse*. The curvature of an ellipse in the normal section changes when $\mathbf{v}$ changes so the normal curvature κ_n is not constant also. The curvature of the ellipse $\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1$ with semiaxis a and b at the point $(a, 0)$ can be computed to be $\frac{a}{b^2}$ (Practice Problem 4) has more details.

There are two choices of $\mathbf{v}$ which stand out:



- Let $\mathbf{v}_1$ be a vector in tangent plane perpendicular to the central axis. The normal section in direction of $\mathbf{v}_1$ is a circle of radius a and its curvature is $\frac{1}{a}$.
- Let $\mathbf{v}_2$ be a vector in the tangent plane parallel to the central axis. The normal section in direction of $\mathbf{v}_2$ is a straight line and its curvature is zero.

These two directions determine the direction of the largest and of the smallest curving.



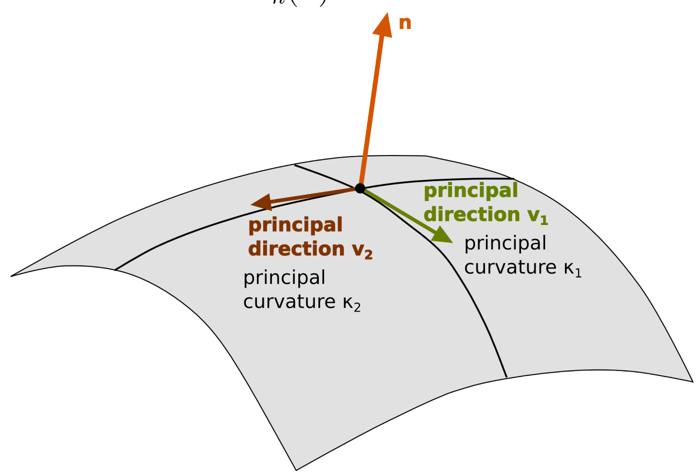
Indeed, consider an arbitrary vector $\mathbf{v}$ in the tangent plane. Let θ be the angle between $\mathbf{v}_1$ and $\mathbf{v}$. Considering the right triangle in the figure on the right, we obtain that the semi-axes of the ellipse in the normal section are a and $\frac{a}{|\cos\theta|}$.

The curvature at the relevant point is $\frac{a}{\left(\frac{a}{|\cos\theta|}\right)^2} = \frac{\cos^2\theta}{a}$. Since $\cos^2\theta$ is taking values between 0 and 1, we obtain that

$$0 \leq |\kappa_n(\mathbf{v})| = \frac{\cos^2\theta}{a} \leq \frac{1}{a}.$$

Since these bounds are reached when considering directions $\mathbf{v}_1$ and $\mathbf{v}_2$, this shows that $|\kappa_n(\mathbf{v}_1)| = \frac{1}{a}$ is the maximal value of $|\kappa_n(\mathbf{v})|$ and $\kappa_n(\mathbf{v}_2) = 0$ is the minimal value.

Principal curvatures. The above example with the cylinder turns to be more general than it may seem. Namely, for every point on any surface, one can choose orthogonal, unit-length vectors $\mathbf{v}_1$ and $\mathbf{v}_2$ called the **principal directions** and the normal curvatures determined by them correspond to the **maximal and minimal values of the normal curvature** $\kappa_n(\mathbf{v})$.



These two values are denoted by κ_1 and κ_2 and are called **principal curvatures.**

The product of the principal curvatures is

$$\boxed{\text{the } \textbf{Gaussian curvature } K = \kappa_1\kappa_2.}$$

We note a significant difference between the curvature of a curve and the normal (and Gaussian) curvature of a surface: while the curvature of a curve is defined to have just nonnegative values (because it is the magnitude of a vector), the principal and Gaussian curvatures of surfaces can have negative values.

In the example with the plane, $\kappa_1 = \kappa_2 = 0$ so $K = 0$. In the example with the sphere, either $\kappa_1 = \kappa_2 = \frac{1}{a}$ or $\kappa_1 = \kappa_2 = \frac{-1}{a}$. In either case, $K = \frac{1}{a^2}$, so the Gaussian curvature is positive.

In the example with the cylinder, $\kappa_1 = \pm\frac{1}{a}$ and $\kappa_2 = 0$ so that $K = 0$. The fact that at every point of the cylinder there is a direction in the tangent plane with the straight line as the normal section causes the relation $K = 0$.

The sign of the Gaussian curvature. In the previous example with the sphere, we have seen that the signs of κ_1 and κ_2 did not impact the sign of K. In fact, only the directions of the normal vectors $\mathbf{N}_1$ and $\mathbf{N}_2$ of the two principal sections impact the sign of the Gaussian curvature. This enables us to avoid considering the sense of the normal vector of the tangent plane and deal with the fact that this sense is often chosen completely arbitrary (think that both $3x-y=0$ and $-3x+y=0$ represent the same plane, so it is arbitrary if we want to use $(3,-1,0)$ or $(-3,1,0)$).

Example. To understand the sign of K, let us consider the following three surfaces. First, let us consider the paraboloid $z = x^2+y^2$ at the origin. The tangent plane is the xy-plane $z = 0$. Note that we can choose the normal vector of the plane to be $\mathbf{n} = (0,0,1)$ but we can also choose $\mathbf{n} = (0,0,-1)$.

Since the paraboloid curves equally in every direction of the tangent plane at the origin, we can choose the directions of the coordinate axes x and y to be the principal directions. Hence, the principal sections are in xz-plane and yz-plane respectively.

The xz-plane $y = 0$ intersects the paraboloid in the parabola $z = x^2$. The curve $\boldsymbol{\gamma}_1 = (x,0,x^2)$ has normal vector $\mathbf{N}_1 = (0,0,1)$ pointing upwards.

The yz-plane $x = 0$ intersects the paraboloid in the parabola $z = y^2$. The curve $\boldsymbol{\gamma}_1 = (0, y, y^2)$ has normal vector $\mathbf{N}_2 = (0, 0, 1)$ also pointing upwards.

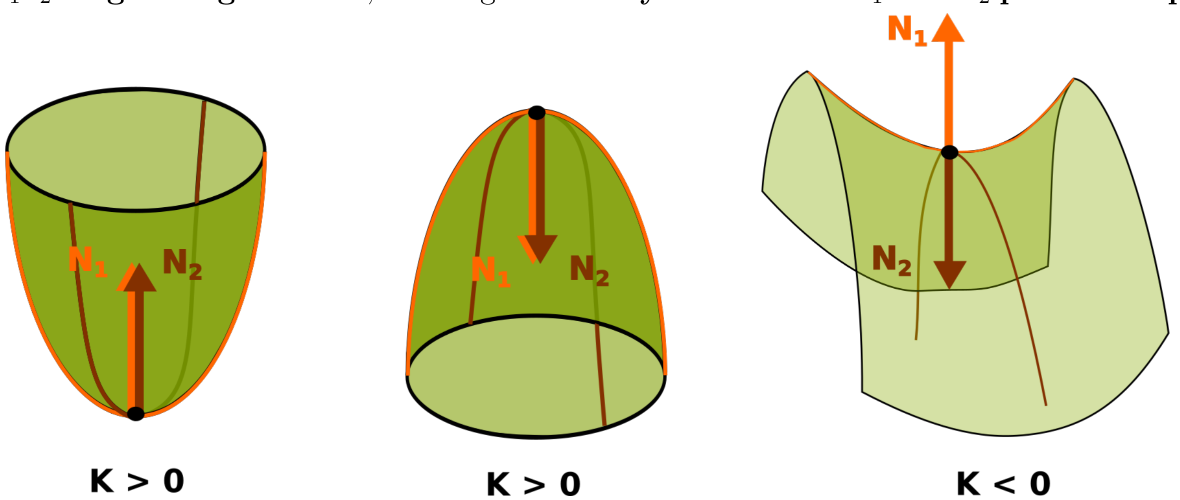
If we choose $\mathbf{n} = (0, 0, 1)$, then both κ_1 and κ_2 are positive and their product $K = \kappa_1\kappa_2$ is positive. If we choose $\mathbf{n} = (0, 0, -1)$, then both κ_1 and κ_2 are negative. However, their product $K = \kappa_1\kappa_2$ is **again** positive. So, K is positive in **any** case because $\mathbf{N}_1$ and $\mathbf{N}_2$ **point the same way.**

For the same reason, the Gaussian of the paraboloid $z = -x^2 - y^2$ at the origin is also positive (both $\mathbf{N}_1$ and $\mathbf{N}_2$ here point downwards).

Let us compare this situation with the **hyperbolic paraboloid** $z = y^2 - x^2$. At the origin, the tangent plane is also the xy-plane $z = 0$. The principal sections are in the xz-plane and the yz-plane, respectively.

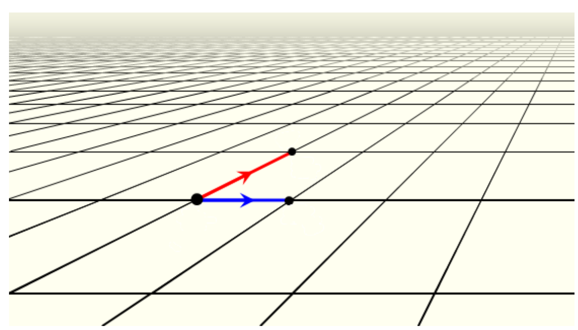
The xz-plane $y = 0$ intersects the paraboloid in the parabola $z = -x^2$. The curve $\boldsymbol{\gamma}_1 = (x, 0, -x^2)$ has normal vector $\mathbf{N}_1 = (0, 0, -1)$ pointing downwards. The yz-plane $x = 0$ intersects the paraboloid in the parabola $z = y^2$. The curve $\boldsymbol{\gamma}_1 = (0, y, y^2)$ has normal vector $\mathbf{N}_2 = (0, 0, 1)$ pointing upwards.

If we choose $\mathbf{n} = (0, 0, 1)$, then κ_1 is negative and κ_2 is positive and their product $K = \kappa_1\kappa_2$ is **negative**. If we choose $\mathbf{n} = (0, 0, -1)$, then κ_1 is positive and κ_2 is negative and their product $K = \kappa_1\kappa_2$ is **again negative**. So, K is negative in **any** case because $\mathbf{N}_1$ and $\mathbf{N}_2$ **point the opposite way.**



Thus, if the principal sections curve in the same direction (both upwards or both downwards, for example), the Gaussian is positive. If the principal sections curve in the opposite directions (one upwards one downwards, for example) then the Gaussian is negative.

Theorema Egregium. The calculation of curvature involves the normal vectors "sticking out" of the surface and, thus, invisible to the locals on the surface. Because of this, it seems that the locals cannot comprehend the sign nor the values of the Gaussian. This uses the same argument we mention before: for people on the surface of the Earth, the Earth appears to be flat.



Locals view their home surface as flat

A concept involving only measurements on the surface (conducted by the locals) is said to be **intrinsic**, while a concept whose definition involves objects external to the surface (like $\mathbf{n}$, for example) is said to be **extrinsic**. Thus, using the definition we presented, the Gaussian curvature seems to be extrinsic and, thus, incomprehensible to the locals.

If one is to generalize the concept of curvature to higher dimensions, in particular the curvature of our physical space, we would have to be able to describe curvature intrinsically. In particular, to determine the curvature of our three-dimensional physical space (and the curvature of Einstein four-dimensional space-time universe), we do not want to rely on more than three (or four) dimensions.

Fortunately for locals, the crowning achievement of theory of surfaces states that **the Gaussian curvature can be calculated intrinsically**. This means that the Gaussian curvature of a surface can be determined entirely by measuring angles, distances and their rates on the surface itself, without further reference to the particular way in which the surface is embedded into the three-dimensional space. This result, proved by Carl Friedrich Gauss, is considered to be one of foundational results in differential geometry. It is usually referred to as **Theorema Egregium** (Latin for remarkable or extraordinary theorem).

Local isometry. Note also that the cylinder can be slit and unrolled into a flat sheet of paper without stretching or tearing and without affecting the length any curve. A surface with this property will also have $K = 0$. This means that the geometry of the cylinder locally is indistinguishable from the geometry of a plane. In cases like this, we say that the two surfaces are **locally isometric**. Note that globally cylinder is very different from the plane, though.

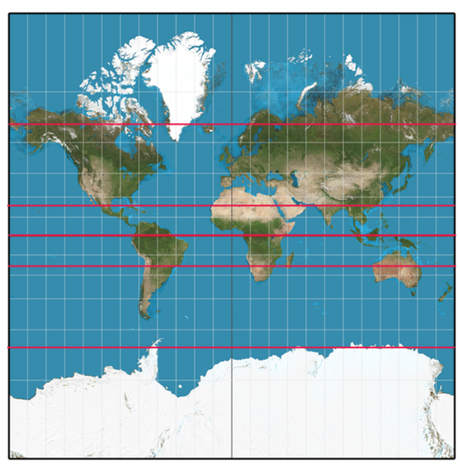
Theorema Egregium also implies that *the Gaussian curvature is invariant under a local isometry.* This means that any bending of a surface (without stretching or tearing) does not impact the Gaussian curvature. The principal curvatures do not share the property of Gaussian curvature given by Theorema Egregium – the principal curvatures do vary with bending. The fact that their product does not vary with bending makes Theorema Egregium even more remarkable.

Theorema Egregium also implies that if two surfaces have different sign of the Gaussian curvature, than one cannot be transformed into another without tearing or crumpling. To further motivate our study, we list several corollaries of this fact.

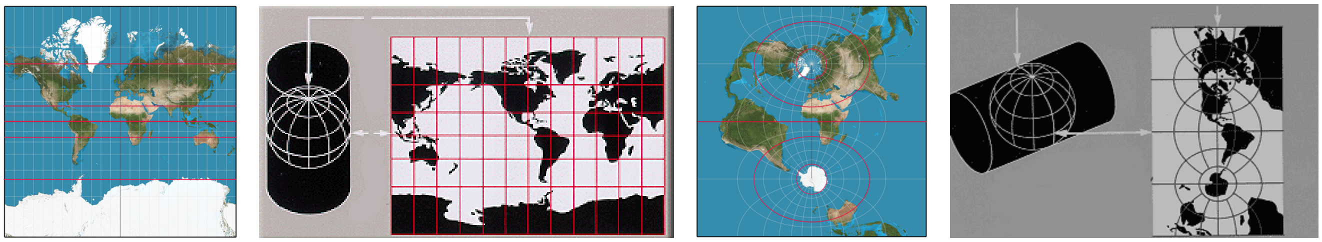
- A sphere (with $K > 0$) and a plane (with $K = 0$) cannot be morphed one into another. Thus, a piece of paper cannot be bent onto a sphere without crumpling.

- As opposed to the cylinder (with $K = 0$), the sphere (with $K > 0$) cannot be unfolded into a flat surface. Thus, if one were to step on an empty egg shell, its edges have to split in expansion before being flattened. An orange peel can be flattened just with tearing or stretching.

  As a consequence of previous observations, the Earth cannot be displayed on a map without distortion. Thus, no perfect map of the Earth can be created, even for a portion of the Earth's surface and every cartographic projection necessarily distorts at least some distances. This fact is of enormous significance for cartography. Every distinct map projection distorts in a distinct way. The study of map projections is the characterization of these distortions.

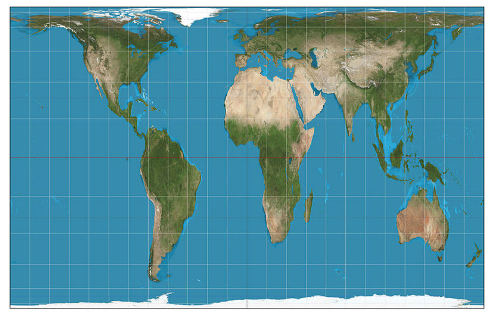


A frequently used projection, Mercator projection, preserves angles but fails to preserve area (that is why the areas around north and south pole look disproportionately large compared to the areas further away from the poles). The controversy surrounding the Mercator projections arose from political implications of map design since representing some countries larger than the others may implied that some are less significant.

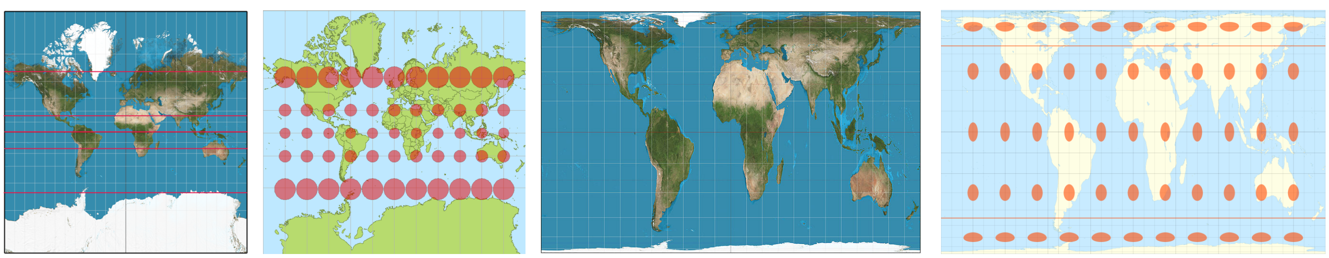


Normal and Transverse Mercator projections

Another projection used in some cases is Gall-Peters projection (you can see it in some world maps on airplanes). On this projection areas of equal size on the globe are also equally sized on the map. This has a consequence that areas around the equator looks elongated when compared to areas with larger geographical width.



Gall-Peters projection



Mercator and Gall-Peters with their deformations

- When trying to preserve precious toppings on a slice of pizza, you are using Theorema Egregium too: you bend a slice horizontally along a radius so that non-zero principal curvatures are created along the bend, dictating that the other principal curvature at these points must be zero. This creates rigidity in the direction perpendicular to the fold and it prevents the toppings from falling off.

- Theorema Egregium also implies that we can measure the curvature of the Earth without leaving the surface (for example in an airplane to observe the curving) just measuring the distances and angles on the surface of the Earth.

The game plan. We devote the remainder of our study of differential geometry to accomplishing the following three goals.

Goal 1 Develop an apparatus that completely describes a surface. This is analogous to the Serret-Frenet apparatus of a curve and leads us to the first and the second fundamental forms.

Goal 2 Understand the statement of Theorema Egregium in mathematical terms and proof of the theorem. This requires consideration of geodesics and the curvature tensor.

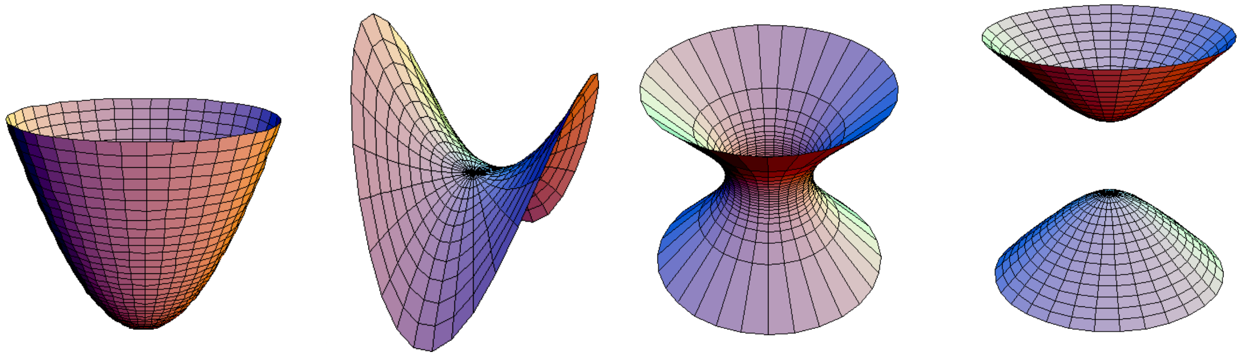
Goal 3 Theorema Egregium allows the concept of curvature to be generalized to **higher dimensions.** Two-dimensional surfaces generalize to n-dimensional **manifolds**, defined for any n and the concept of the curvature of a surfaces generalizes to the curvature of a manifold. This enables you to understand the language used in special and general relativity. It also enables you to generalize the content of this course to higher dimensions.

Practice Problems.

1. The **mean curvature** is defined as the mean of the principal curvatures $H = \frac{\kappa_1+\kappa_2}{2}$. Determine the absolute value of the mean curvature of the surfaces discussed in this section: plane, sphere of radius a and cylinder $x^2 + y^2 = a^2$.

2. Calculate the curvature of the parabolas $y = \pm ax^2$ at the origin for $a > 0$. Use this formula to find the Gaussian curvature of the hyperbolic paraboloid $z = ax^2 - by^2$ at the origin for $a, b > 0$.

3. A **quadratic surface** is any surface given by equation $ax^2 + by^2 + cz^2 + dxy + exz + fyz + gx + hy + iz + j = 0$. This class includes the following surfaces: ellipsoid ($\frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} = 1$), elliptical paraboloid ($\frac{x^2}{a^2} + \frac{y^2}{b^2} = z$), hyperbolic paraboloid ($\frac{x^2}{a^2} - \frac{y^2}{b^2} = z$), hyperboloid of one sheet ($\frac{x^2}{a^2} + \frac{y^2}{b^2} - \frac{z^2}{c^2} = 1$) and hyperboloid of two sheets ($\frac{x^2}{a^2} + \frac{y^2}{b^2} - \frac{z^2}{c^2} = -1$).

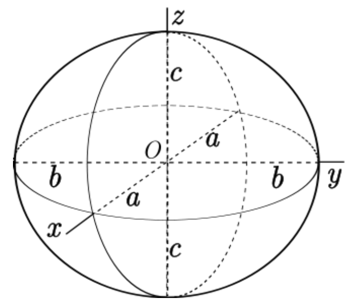


Elliptical and hyperbolic paraboloids and hyperboloids of one and two sheets

By making a suitable change of variables to eliminate some terms, any quadratic surface can be put into a certain normal form. It turns out that there are 16 such normal forms. Of these 16 forms, the above five surfaces are non-degenerate and remaining eleven are degenerate: cones ($\frac{x^2}{a^2} + \frac{y^2}{b^2} - \frac{z^2}{c^2} = 0$), cylindrical surfaces (elliptic, hyperbolic and parabolic cylinder), planes, lines, points or even no points at all.

Using argument similar to those used to show that the Gaussian curvature of a cylinder is 0, deduce that K of all the degenerate quadratic surfaces is 0. Then determine the sign of Gaussian curvature for five non-degenerate quadratic surfaces.

4. Find the Gaussian curvature of ellipsoid $\frac{x^2}{a^2}+\frac{y^2}{b^2}+\frac{z^2}{c^2}=1$ at the end points of the three semi-axes $(\pm a,0,0)$, $(0,\pm b,0)$ and $(0,0,\pm c)$.

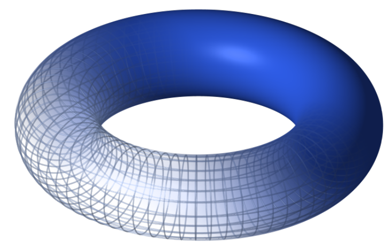


Ellipsoid

5. A **torus** is a surface obtained by revolving one circle along the other circle creating a doughnut-like shape. Consider revolving a circle $(x-a)^2+z^2=b^2$ in xz-plane along the circle $x^2+y^2=a^2$ in xy-plane.

   Assume that $a>b$ so that "the doughnut" that you obtain has a hole in the middle.

   Calculate the Gaussian curvature at any point on the "outer" circle (obtained by revolving the point $(a+b,\ 0,\ 0)$ about z-axis) and at any point on the "inner" circle (obtained by revolving the point $(a-b,\ 0,\ 0)$ about z-axis.


Torus

Solutions. (1) $H=\frac{0+0}{2}=0$ for any plane, $|H|=\frac{\frac{1}{a}+\frac{1}{a}}{2}=\frac{1}{a}$ for the sphere of radius a, and $|H|=\frac{0+\frac{1}{a}}{2}=\frac{1}{2a}$ for the cylinder $x^2+y^2=a^2$.

(2) To calculate the curvature of $y=\pm ax^2$, let $\boldsymbol{\gamma}=(x,\pm ax^2,0)$ and compute $\boldsymbol{\gamma}',\gamma''$ and then $\kappa=\frac{|\boldsymbol{\gamma}'\times\boldsymbol{\gamma}''|}{|\boldsymbol{\gamma}'|^3}=\frac{2a}{\sqrt{(1+4a^2x^2)^3}}$. At the origin, $x=0$ so $\kappa=\frac{2a}{\sqrt{1^3}}=2a$. Note that κ is positive regardless of whether we consider $y=ax^2$ or $y=-ax^2$.

Then consider the hyperbolic paraboloid $z=ax^2-by^2$. The normal sections of $z=ax^2-by^2$ are parabolas in xz and yz planes. In xz plane, $y=0$ and so $z=ax^2-b0^2=ax^2$ and hence $|\kappa_1|=2a$. In yz plane, $x=0$ and so $z=a0^2-by^2=-by^2$ and hence $|\kappa_2|=2b$. The two normal vectors have the opposite direction (because one parabola faces upwards and the other downwards) and so the two principal curvatures have the opposite signs: if $\kappa_1=\pm 2a$ then $\kappa_2=\mp 2b$. Thus, $K=-(2a)(2b)=-4ab$.

(3) $K>0$ for ellipsoid. $K>0$ for elliptical paraboloid. $K<0$ for hyperbolic paraboloid. $K<0$ for hyperboloid of one sheet. $K>0$ for hyperboloid of two sheets.


Hyperbolical Paraboloids

(4) Let us calculate the curvature of ellipse $\frac{x^2}{a^2}+\frac{y^2}{b^2}=1$ at $(a,0)$ and $(0,b)$ first. Here the curve can be parametrized as $\boldsymbol{\gamma}=(a\cos t,b\sin t)$. Then $\boldsymbol{\gamma}'=(-a\sin t,b\cos t,0)$, $\boldsymbol{\gamma}''=(-a\cos t,-b\sin t,0)$ $\boldsymbol{\gamma}'\times\boldsymbol{\gamma}''=(0,0,ab)$. Thus $|\gamma'|=\sqrt{a^2\sin^2 t+b^2\cos^2 t}$ and $|\boldsymbol{\gamma}'\times\boldsymbol{\gamma}''|=ab$ and so $\kappa=\frac{ab}{(a^2\sin^2 t+b^2\cos^2 t)^{3/2}}$.

At $(a, 0)$ the value of parameter t is 0 and at $(0, b)$ the value of parameter t is $\frac{\pi}{2}$. Thus $\kappa(0) = \frac{ab}{b^3} = \frac{a}{b^2}$ and $\kappa(\frac{\pi}{2}) = \frac{ab}{a^3} = \frac{b}{a^2}$.

At $(\pm a, 0, 0)$, the normal sections are in xy and xz planes. In xy plane the normal section is the ellipse $\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1$ with curvature $\frac{a}{b^2}$ at $(\pm a, 0)$. In xz plane the normal section is the ellipse $\frac{x^2}{a^2} + \frac{z^2}{c^2} = 1$ with curvature $\frac{a}{c^2}$ at $(\pm a, 0)$. The normal vectors have the same directions so the two principal curvatures have the same signs. Thus the Gaussian is $K = \frac{a^2}{b^2c^2}$.

At $(0, \pm b, 0)$, the normal sections are in xy and yz planes. In xy plane the normal section is the ellipse $\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1$ with curvature $\frac{b}{a^2}$ at $(0, \pm b)$. In yz plane the normal section is the ellipse $\frac{y^2}{b^2} + \frac{z^2}{c^2} = 1$ with curvature $\frac{b}{c^2}$.

The normal vectors have the same directions so the two principal curvatures have the same signs. Thus the Gaussian is $K = \frac{b^2}{a^2c^2}$. On similar manner, we obtain that the Gaussian curvature at $(0, 0, \pm c)$ is $K = \frac{c^2}{a^2b^2}$.

(5) At point $(a + b, 0, 0)$, the normal sections are in xy plane and xz plane. In xy plane, the normal section is the circle of radius $a + b$ so its curvature is $\frac{1}{a+b}$. In xz plane the normal section is the circle of radius b so its curvature is $\frac{1}{b}$. The normal vectors have the same direction. Hence, the Gaussian curvature is positive and equal to $\frac{1}{b(a+b)}$.

At point $(a - b, 0, 0)$, the normal sections are in xy plane and xz plane as well. In xy plane, the normal section is the circle of radius $a - b$ with curvature $\frac{1}{a-b}$. In xz plane the normal section is the circle of radius b with curvature $\frac{1}{b}$. The normal vectors have the opposite direction. Hence, the Gaussian curvature is negative and equal to $\frac{-1}{b(a-b)}$.

Using this example, we can deduce that on the outer part of the torus (obtained by revolving the right half of circle $(x - a)^2 + z^2 = b^2$ about z-axis) the Gaussian curvature is positive and on the inner part of the torus (obtained by revolving the left half of circle $(x - a)^2 + z^2 = b^2$ about z-axis) the Gaussian curvature is negative. This implies the not so obvious fact that the Gaussian curvature on the "top" and "bottom" circles (obtained by revolving points $(a, 0, b)$ and $(a, 0, -b)$ about z-axis) is zero.