THE GAUSS-BONNET THEOREM

CHRISTIAN SCHNELL

1. A short biography

Johann Carl Friedrich Gauss was born on April 30, 1777 in the German city of Braunschweig. His precociousness (well-known from the old and tired anectdote about his adding the numbers from 1 to 100 in a clever way) impressed the Duke of Braunschweig, who financed Gauss's studies at the University of Göttingen. Gauss obtained his Doctor of Philosophy in 1799, and until 1807, when the funding from Braunschweig ceased, had his freest and most productive time. In 1807, he became director of the astronomical observatory at Göttingen; besides this, he also held a professorship at the university of that city. In later years, he often complained about his manifold duties that prevented him from doing his mathematical work. Gauss died in Göttingen on February 23, 1855.

Figure 1. A map of Germany, showing the area in which Gauss lived.

Gauss followed his maxim "Pauca sed matura."[1] when publishing; accordingly, the best source for historians of mathematics are his notebooks, where the development of his ideas is more obvious than in his papers.

2. Introduction

Gauss's major published work on differential geometry is contained in the "Disquisitiones generales circa superficies curvas" from 1827; this short paper of only 40 pages was presented to the *Königliche Gesellschaft der Wissenschaften* in the form of a lecture on October 8. An English translation, together with the original Latin text and various other information, can be found in Dombrowski's book [1].

[1] "Few, but mature."

The surveying work that Gauss did for the Kingdom of Hannover (then in personal union with Great Britain) from 1816 on, as well as the practical aspects of surveying, helped to make him interested in the geometry of curved surfaces. During the 19th century, people began to make precise geological surveys, both to determine the exact shape of the Earth (definition of the meter in France), and for administrative reasons; when the Kingdom of Hannover was triangulated from 1821 until 1844, Gauss was in charge of first the measurements and later the evaluation of the numerical data.

3. The Gauss-Bonnet theorem on the sphere

We begin the mathematical part of this paper by looking at the Gauss-Bonnet theorem on the simplest of curved surfaces, the sphere. We study triangles on the two-sphere $\mathbb{S}^2$, which is the set of points $(x, y, z) \in \mathbb{R}^3$ with $x^2 + y^2 + z^2 = 1$. If the sides of such a triangle come to lie on great circles, and are therefore *geodesics* or curves of shortest length, we call it a *geodesic triangle.*

Theorem 3.1. *Let Δ be a geodesic triangle on $\mathbb{S}^2$ with interior angles α, β, γ. The area of Δ equals $|\Delta| = \alpha + \beta + \gamma - \pi$.*

The clever *proof* of this formula goes as follows. Let us label the great circles bounding Δ as a, b, c, in such a manner that the angle α is enclosed between b and c and so forth. Together, b and c divide the surface of the sphere into four crescent-shaped pieces; the combined area of the two pieces with interior angle α is $2 \cdot \alpha/(2\pi) \cdot 4\pi = 4\alpha$. A similar formula holds for the the other two regions corresponding to β and γ.

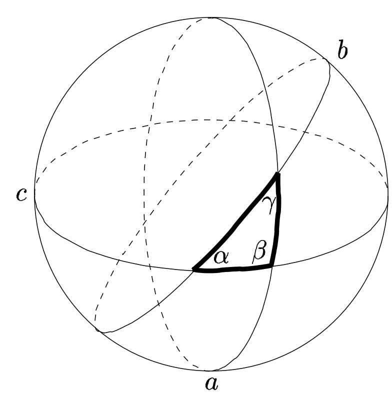

Figure 2. A geodesic triangle Δ on the two-sphere.

All three regions together cover the entire sphere, but Δ and its antipodal image (on the opposite side of the sphere) are covered exactly three times. Therefore

$$4\alpha + 4\beta + 4\gamma = 4\pi + 2 \cdot 2|\Delta| \qquad \text{or} \qquad |\Delta| = \alpha + \beta + \gamma - \pi,$$

which is the claimed formula for the area of the triangle.

Consequences. This simple result on the area of triangles has several interesting consequences. First, the area of a spherical triangle depends only on it angles; this is quite different from the case of plane triangles, which can always be rescaled. Secondly, the sum of angles $\alpha + \beta + \gamma$ in a spherical triangle exceeds π; not only that, but the discrepancy grows with the size of the triangle.

The usual laws of Euclidean geometry are thus violated on the surface of a sphere, precisely because the sphere is not flat, but curved; this fact becomes more and more apparent the bigger a portion of the surface one looks at. Interestingly, by measuring the angles inside sufficiently big triangles and finding that their sum exceeds π, a person on the sphere can discover that he is not living in the plane but on a curved surface.

Our formula also shows that there can be no completely accurate map of any part of the Earth. For say we had a map that accurately represented all lengths and angles, of at least a small part of the Globe. Then pieces of great circles would have to become pieces of straight lines on our map, since great circles and straight lines are both curves of shortest length. A small geodesic triangle thus has to become an ordinary triangle on the map, but this is clearly not possible because the angle sums in the two triangles are not the same. (One is greater than, the other exactly equal to, π.)

General version of the Gauss-Bonnet theorem. On an arbitrary curved surface M in $\mathbb{R}^3$, a similar formula for the area of geodesic triangles exists. For the time being, we think of geodesics simply as curves of shortest length; they are the generalization of lines or great circles to the surface M. Let Δ be a small geodesic triangle on M, with interior angles α, β, γ. Then

$$\int_{\Delta} k d\sigma = \alpha + \beta + \gamma - \pi,$$

where k, the *Gaussian curvature* of M, measures how curved the surface is. We will see below how k is defined; for now, note that the sphere, which is certainly curved in the same way at each point, has $k = 1$.

4. Ideas and results from the *Disquisitiones*

In this section, we finally look at some of the results from the *Disquisitiones generales circa superficies curvas.* In a few places, we shall see a bit of Gauss's style, and of his way of proving things; it is already quite modern, but has enough of the charming quaintness of older mathematics (e.g., use of infinitesimals) to be interesting. The Gauss-Bonnet theorem is obviously not at the beginning of the paper...

Gauss' main point is to study the *intrinsic geometry* of surfaces, meaning, those properties that do not depend on how exactly the surface sits in space. We shall see an example of such a notion below, in the *Gaussian curvature* of a suface.

Object of study. In the *Disquisitiones*, Gauss studies what we now call orientable surfaces in three-space.[2] Here is how Gauss himself defines his object of study (see [1, p. 9]):

[2] An *orientable* surface in $\mathbb{R}^3$ is one on which a consistent choice of unit normal vector can be made. A sphere is orientable, a Möbius band is not.

> A curved surface is said to possess continuous curvature at one of its points A, if the directions of all the straight lines drawn from A to points of the surface at an infinitely small distance from A are deflected infinitely little from one and the same plane passing through A. This plane is said to *touch* the surface at the point A. If this condition is not satisfied for any point, the continuity of the curvature is here interrupted, as happens, for example, at the vertex of a cone. The following investigations will be restricted to such surfaces, or to such parts of surfaces, as have the continuity of their curvature nowhere interrupted.

The modern term for continuous curvature is *differentiability*; the kind of surface considered by Gauss is thus an example of a differentiable manifold. Note that no name is given the surface—most people nowadays would begin "A curved surface M is said..."

The Gauss map. The first important idea is the introduction of the so-called *Gauss map* $\zeta\colon M \to \mathbb{S}^2$ from the surface to the two-sphere; ζ associates to each point the unit normal vector at that point. Gauss himself (see [1, p. 85]) mentions astronomy in this context.

> In researches in which an infinity of directions of straight lines in space is concerned, it is advantageous to represent these directions by means of those points upon a fixed sphere, which are the end points of the radii drawn parallel to the lines. The centre and the radius of this *auxiliary sphere* are here quite arbitrary. The radius may be taken equal to unity. This procedure agrees fundamentally with that which is constantly employed in astronomy, where all directions are referred to a fictitious celestial sphere of infinite radius.

How curved the surface is can be seen from the behavior of ζ, that is, of the normal vector to the surface. On a flat surface, the normal vector varies little; on a highly curved surface, it varies rapidly. It is therefore reasonable that Gauss should use ζ to study the curvature of the surface.

Gaussian curvature. Gauss uses the following quantities to measure how curved a surface is. The *total curvature* of a subset $D \subseteq M$ is defined to be the area of its image $\zeta(D)$ under the Gauss map, with positive (resp. negative) sign if the position of $\zeta(D)$ on the sphere is similar (resp. opposite) to that of D.

The *measure of curvature* of M at a point, on the other hand, is the limit

$$k = \lim_{D} \frac{\text{area of } \zeta(D)}{\text{area of } D}$$

as D becomes infinitely small. Again, k is to be taken with a negative sign if ζ reverses the orientation of D. Defined in a more rigorous way, k (as a function on M) is nowadays called the *Gaussian curvature*. It determines the total curvature, because if $d\sigma$ is the surface element on M, then

$$\text{total curvature of } D = \int_D k\, d\sigma.$$

To illustrate these ideas, we should look at a few examples of surfaces.

(1) A sphere of radius r. At a point (x, y, z) on the sphere, the unit normal vector is $N = (x/r, y/r, z/r)$; consequently, the Gauss map ζ expands area by a factor of $1/r^2$, and $k = 1/r^2$. The smaller the radius, the larger the curvature—this certainly agrees with our intuition.
(2) A plane. Since the normal vector to a plane is everywhere the same, ζ is a constant map, and $k = 0$.
(3) A cylinder. Now the image of ζ is a great circle, which has zero area, and so $k = 0$. A cylinder is therefore not curved in Gauss's sense, although it certainly looks curved. The reason, as we shall see below, is that the cylinder can be flattened out, and is, at least locally, equivalent to the plane.
(4) A saddle. A look at Figure 3 shows that the Gauss map reverses the direction of curves; this means that a saddle-shaped surface has negative curvature.

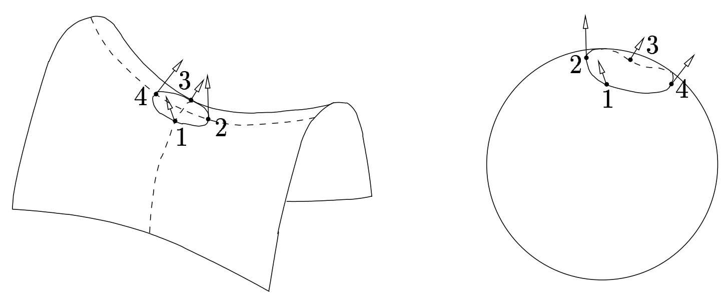

FIGURE 3. The behavior of the normal map on a saddle-shaped surface.

A computation. Gauss derives several formulas for the measure of curvature k, depending on how the surface M is presented. Out of curiosity, let us try to follow one of his computations.

Say the surface is given in the form $z = z(x, y)$. Let X, Y, Z stand for the components of the unit normal vector to the surface.

Let $d\sigma$ be the area of an infinitesimal triangle on the curved surface, and $d\Sigma$ the area of the corresponding triangle on the unit sphere. Then $Zd\sigma$ and $Zd\Sigma$ are, respectively, the areas of the projections of the two triangles to the x, y-plane; this is because the vector with components X, Y, Z is both the normal vector at a point on the surface, and the normal vector at the corresponding point on the sphere.

Suppose that when the triangle on the surface is projected to the x, y-plane, its points have coordinates

$$x, y, \qquad x + dx, y + dy, \qquad x + \delta x, y + \delta y,$$

respectively; then $Zd\sigma = (dx \cdot \delta y - dy \cdot \delta x)/2$.

In like manner, $Zd\Sigma = (dX \cdot \delta Y - dY \cdot \delta X)/2$, and therefore the measure of curvature at this point on the surface is

$$k = \frac{d\Sigma}{d\sigma} = \frac{dX \cdot \delta Y - dY \cdot \delta X}{dx \cdot \delta y - dy \cdot \delta x}.$$

But since X, Y, Z are functions of the quantities x, y, we have

$$dX = \frac{\partial X}{\partial x} dx + \frac{\partial X}{\partial y} dy, \quad \delta X = \frac{\partial X}{\partial x} \delta x + \frac{\partial X}{\partial y} \delta y, \quad \text{etc.};$$

when these values are substituted, the expression above becomes

$$k = \frac{\partial X}{\partial x} \cdot \frac{\partial Y}{\partial y} - \frac{\partial X}{\partial y} \cdot \frac{\partial Y}{\partial x}.$$

Gauss's other interpretation of k. To have better notation, let us now write the equation for the surface as $z = f(x, y)$. If we denote partial derivatives by subscripts, then the normal vector to M is the normalized cross product of $(1, 0, f_x)$ and $(0, 1, f_y)$,

$$(X, Y, Z) = \frac{1}{\sqrt{1 + f_x^2 + f_y^2}} (-f_x, -f_y, 1).$$

From this, one easily finds that

$$k = \frac{f_{xx} f_{yy} - f_{xy}^2}{(1 + f_x^2 + f_y^2)^2}.$$

Take now a fixed point P on the surface, and adjust the coordinate system (by translating and rotating) to make

$$P = (0, 0, 0), \qquad f_x(0, 0) = f_y(0, 0) = 0, \qquad f_{xy}(0, 0) = 0.$$

If we put $T = f_{xx}(0, 0)$ and $V = f_{yy}(0, 0)$, then

$$f(x, y) = \frac{1}{2} T x^2 + \frac{1}{2} V y^2 + \text{terms of higher order};$$

on the other hand, the Gaussian curvature at P is now simply

$$k = TV.$$

This admits of the following interpretation. Suppose the surface M is cut with planes containing the normal vector at P, which is $(0, 0, 1)$ in our coordinates. The sections are plane curves, whose curvature at P, as plane curves, can easily be computed. For example, if we cut with the plane $x = 0$, the resulting curve is

$$\left(0, y, \frac{1}{2} V y^2 + \text{terms of higher order}\right),$$

and the curvature at P of this curve equals V. If we cut with the plane $y = 0$, the value of the curvature is similarly T.

For any other plane, the value of this *sectional curvature* is between $\min(T, V)$ and $\max(T, V)$; therefore, as Gauss observes, k is the product of the largest and the smallest sectional curvature at each point; this is closer to the modern definition of Gaussian curvature.

The *Theorema egregium*. In the case when M, or a part of M, is given as the image of an open set in $\mathbb{R}^2$ under a map[3], Gauss also derives a longer, but "most productive" formula for k.

[3]Since M is a differentiable manifold, every point on M has a neighborhood for which that is the case.

Suppose $f = (f_1, f_2, f_3)$ is an isomorphism of $U \subseteq \mathbb{R}^2$ to some open set in M. Using coordinates p and q on U, we may define

$$E = \left\langle \frac{\partial f}{\partial p}, \frac{\partial f}{\partial p} \right\rangle = \left(\frac{\partial f_1}{\partial p}\right)^2 + \left(\frac{\partial f_2}{\partial p}\right)^2 + \left(\frac{\partial f_3}{\partial p}\right)^2$$

$$F = \left\langle \frac{\partial f}{\partial p}, \frac{\partial f}{\partial q} \right\rangle = \frac{\partial f_1}{\partial p} \cdot \frac{\partial f_1}{\partial q} + \frac{\partial f_2}{\partial p} \cdot \frac{\partial f_2}{\partial q} + \frac{\partial f_3}{\partial p} \cdot \frac{\partial f_3}{\partial q}$$

$$G = \left\langle \frac{\partial f}{\partial q}, \frac{\partial f}{\partial q} \right\rangle = \left(\frac{\partial f_1}{\partial q}\right)^2 + \left(\frac{\partial f_2}{\partial q}\right)^2 + \left(\frac{\partial f_3}{\partial q}\right)^2;$$

In modern usage, these three quantities are part of a *metric* on the surface. A slightly old-fashioned way of writing this metric is as

$$ds^2 = Edp^2 + 2Fdpdq + Gdq^2;$$

the length element ds is used when measuring lengths on M.

Gauss manages to express k purely in terms of E, F and G, and their first and second partial derivatives. He then writes (see [1, p. 39]):

> Thus the formula of the preceding article leads of itself to the remarkable
>
> **Theorem.** *If a curved surface is developed upon any other surface whatever, the measure of curvature in each point remains unchanged.*

The word "remarkable" is a translation of "egregium" in the original Latin text; the theorem is still known as the *Theorema egregium.* This result is important because it shows that the Gaussian curvature k is intrinsic to the surface; it depends only on the metric, that is, on how lengths are measured inside the surface, but not on how the surface sits in its ambient space $\mathbb{R}^3$.

Geodesics. Geodesics on a curved surface play a similar role as lines in the plane, at least locally. In any sufficiently small region U of the surface, any two points can be joined inside U by a unique geodesic, which is the curve of smallest length between those two points. On a sphere, for example, the great circles are geodesics; since we usually have two ways, a long one and a short one, to move between two points on a great circle, it is clear that not every geodesic is a path of shortest length. Geodesics are more like "critical points" for curve length.

In the sections of the *Disquisitiones* that treat geodesics, Gauss proves that a curve $\gamma\colon [0,1] \to M$ on a surface is a normal geodesic, meaning a geodesic with constant velocity $\|\dot\gamma\|$, if and only if the second derivative, or acceleration, $\ddot\gamma$ is always perpendicular to the surface.

Let us see how this works in the case of the sphere. For any curve γ on the sphere, we have $\|\gamma\| = 1$, and then by differentiation $\langle \gamma, \dot\gamma \rangle = 0$. Thus γ and $\dot\gamma$ are everywhere orthogonal.

As we said, any geodesic is a great circle, and is thus contained in a plane, and so $\ddot\gamma$ is always a linear combination of γ and $\dot\gamma$. If the geodesic is in addition normal, the relation $\|\dot\gamma\| = 1$ shows that $\ddot\gamma$ is also orthogonal to $\dot\gamma$, and therefore a multiple of γ and perpendicular to the sphere.

If, on the other hand, $\ddot\gamma$ is perpendicular to the sphere, and in particular to $\dot\gamma$, then a simple computation shows that both $\|\dot\gamma\|$ and the vector $\gamma \times \dot\gamma$ are constant, which means that the curve has constant velocity and lies entirely in a plane. It is then a great circle, and thus a normal geodesic.

Gauss also establishes the following very interesting lemma; it allows him to use geodesical polar coordinates, which are a major ingredient in his proof of the Gauss-Bonnet theorem. Fix a point P on the surface, and consider a small open neighborhood U of P. For any point $Q \in U$ sufficiently close to P, we can find a unique geodesic contained in U and connecting P and Q; if this geodesic is parametrized by arc length, we can use its length to define the *geodesic distance* from P to Q. All points Q at a fixed distance to P form a *geodesic circle.*

Gauss' Lemma. *Any sufficiently small geodesic circle around P is perpendicular to all geodesics through P.*

This is well illustrated by circles on the surface of a sphere.

5. A simple proof of the Gauss-Bonnet theorem

Here, we give a simple proof of the general Gauss-Bonnet theorem, essentially following an article by Mark Levi [3]. The argument is also very computational, but of course totally different from the one used by Gauss in the *Disquisitiones.*

The version we are going to prove is the following.

Theorem 5.1. *Let $M \subset \mathbb{R}^3$ be an orientable surface, and let R be a sufficiently small open subset of M diffeomorphic to a disc, with smooth boundary $\gamma\colon [0,1] \to M$. We assume that γ is positively oriented and let κ_g be its geodesic curvature. Then*

$$\int_R k d\sigma + \int_\gamma k_g ds = 2\pi.$$

The *geodesic curvature* of a curve γ in M is defined as follows: to compute κ_g at a point P on the curve, project the curve to the tangent plane to M at P, and take the curvature of the resulting plane curve at the point P. From this, it is not hard to get a formula for κ_g. Say $n(t)$ is the unit normal vector to the surface M at the point $\gamma(t)$. Let $w = \|\gamma(t)\|$ be the speed of the curve, and $U(t) = \gamma(t)/w(t)$ its unit tangent vector. Then

$$(\dot{U}, n \times U) = w \cdot \kappa_g.$$

There is also a version of the Gauss-Bonnet theorem allowing piecewise smooth boundaries; it is derived from the result above by rounding off the corners.

Theorem 5.2. *Let $M \subset \mathbb{R}^3$ be an orientable surface, and let R be a sufficiently small open subset of M diffeomorphic to a disc, with piecewise smooth boundary consisting of smooth curves $\gamma_i\colon [0,1] \to M$ meeting at points P_i, for $i = 1, 2, \dots, n$. Let κ_g denote the geodesic curvature along each of the curves, and let α_i be the change in direction from γ_{i-1} to γ_i at P_i (see Figure 4 below). Then*

$$\int_R k d\sigma + \sum_{i=1}^n \int_{\gamma_i} k_g ds + \sum_{i=1}^n \alpha_i = 2\pi.$$

The remainder of this section gives a full proof for Theorem 5.1—well, nearly.

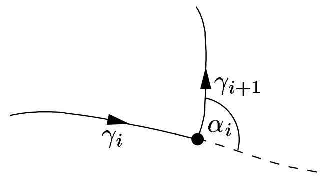

FIGURE 4. The change of direction between γ_i and γ_{i+1}.

Vector identities. We state a few useful identities for inner products and cross products of vectors in $\mathbb{R}^3$. Let $a, b, c, \ldots$ be arbitrary vectors in $\mathbb{R}^3$; the inner product of a and b is written (a, b), the cross product $a \times b$.

$$a \times (b \times c) = (a, c)b - (a, b)c$$
$$(a \times b, c) = (c \times a, b)$$
$$(a \times b, c \times d) = (a, c)(b, d) - (a, d)(b, c)$$

These will be used repeatedly throughout the proof...

Curves and dual curves. We begin by defining the notion of a *(closed) directed curve* on the sphere $\mathbb{S}^2$. Such a curve is a pair $c = (n, T)$, with both n and T smooth maps from $[0, 1]$ to $\mathbb{S}^2$ satisfying $n(0) = n(1)$ and $T(0) = T(1)$, and such that $T(t)$ is always parallel to $\dot{n}(t)$. In other words, c consists of a closed curve n on the sphere, together with a choice of unit tangent vector T. Note that n and T are always perpendicular (because n is a unit vector).

Given a curve $c = (n, T)$, we define its *dual curve* to be $c^* = (n^*, T^*) = (n \times T, -T)$. This is again a closed directed curve:

$$\dot{n}^* = \dot{n} \times T + n \times \dot{T} = n \times \dot{T},$$

which is parallel to T^* because both n and $\dot{T}$ are perpendicular to $T^* = -T$.

The reason for putting $T^* = -T$ instead of just T is to make $(c^*)^* = c$; indeed,

$$n^* \times T^* = -(n \times T) \times T = T \times (n \times T) = n.$$

At this point, the reader should check that if c is a great circle on the sphere, then c^* is a point—but with a tangent vector T^* that rotates around once as t moves from 0 to 1.

Length and area of directed curves. The *length* of a directed curve $c = (n, T)$ will be defined as

$$l(c) = \int_0^1 (\dot{n}, T)dt.$$

In case $\dot{n}$ is never zero, $l(c) = \int \|\dot{n}\| dt$ if T and $\dot{n}$ always point in the same direction; then $l(c)$ is the usual length of the curve defined by n. If $\dot{n}$ and T point in opposite directions, $l(c)$ acquires a negative sign.

In case the region bounded by the curve c is diffeomorphic to a disc, we can also define the *area* $A(c)$ of c. Since c will be the boundary of two separate regions, we have to be clear about which of the two we mean. The region bounded by c shall always be the one to the *left* of c; in other words, the one that lies in the direction of $n \times T$.

Now say c bounds a region diffeomorphic to a disc; we parametrize the region by a map $n = n(t, \tau)\colon [0,1]^2 \to \mathbb{S}^2$ with $n(t,0) = n(t)$ and $n(t,1)$ constant (corresponding to the center of the disc), and define its area as

$$A(c) = \int_0^1 \int_0^1 (n_t \times n_\tau, n) dt d\tau,$$

where n_t and n_τ denote partial derivatives. It is an easy exercise in calculus to see that this gives the correct area.

Another good exercise is to compute the area of a circle c on the surface of the unit sphere, as well as the length of its dual curve c^*, and to verify the conclusion of the following theorem in this special case.

The Dual Cones Theorem. *If a closed directed curve $c = (n, T)$ bounds a region diffeomorphic to a disc, then $l(c^*) + A(c) = 2\pi$.*

The idea of the *proof* is simple: deform c into a point and show that $l(c^*) + A(c)$ remains constant in the process. When c is nearly a point, c^* is nearly a great circle, which gives the formula.

Let us now make this precise. We consider only the case when c is a smooth curve, that is, when $\dot{n}$ is nowhere zero (the general case can be handled by perturbing the curve a little bit). With this assumption, $\dot{n} = \varepsilon \|\dot{n}\| T$ for a fixed $\varepsilon = \pm 1$.

As in the definition of $A(c)$, we parametrize the region bounded by c by a map $n = n(t,\tau)\colon [0,1]^2 \to \mathbb{S}^2$ with $n(t,0) = n(t)$, and $n(t,1)$ constant. We can also view this map n as giving us a family of curves $t \mapsto n(t,\tau)$, starting at the original curve and ending at a point P. To make the argument work, we require three conditions.

- $n(t,\tau)$ should be a diffeomorphism as long as $\tau \neq 1$.
- The various curves $t \mapsto n(t,\tau)$ should have non-zero tangent vector, at least for $\tau \neq 1$.
- The behavior near the point P should be nice, say all curves with τ close to 1 should be small circles centered at P.

In this case, we can extend T smoothly to the unit square by setting $T(t,\tau) = \varepsilon n_t / \|n_t\|$ for $\tau \neq 1$; since all curves for τ close to 1 are small circles, the map can be extended to $\tau = 1$ by continuity.

For each $\tau \in [0,1]$, we then have a directed curve

$$c_\tau = \big(n(t,\tau), T(t,\tau)\big);$$

$c_0 = c$, while c_1 is a point. We also have the dual curves

$$c_\tau^* = \big(n^*(t,\tau), -T(t,\tau)\big) = \big(n(t,\tau) \times T(t,\tau), -T(t,\tau)\big),$$

and the third item above guarantees that c_1^* is a great circle.

Consider first the area $A(c_\tau)$, as a function of τ. Clearly,

$$A(c_\tau) = \int_\tau^1 \int_0^1 (n_t \times n_\tau, n) dt d\tau,$$

and so

$$\begin{aligned} \frac{d}{d\tau} A(c_\tau) &= -\int_0^1 (n_t \times n_\tau, n) dt = -\varepsilon \int_0^1 \|n_t\| (T \times n_\tau, n) dt \\ &= -\varepsilon \int_0^1 \|n_t\| (n \times T, n_\tau) dt = -\varepsilon \int_0^1 \|n_t\| (n^*, n_\tau) dt. \end{aligned} \tag{5.1}$$

Next, let us compute the length of c^*_τ as a function of τ.

$$l(c^*_\tau) = \int_0^1 (n^*_t, -T)dt = -\int_0^1 (n^* \times n^*_t, n^* \times T)dt = \int_0^1 (n^* \times n^*_t, n)dt.$$

Again, we compute the derivative and simplify.

$$\frac{d}{d\tau} l(c^*_\tau) = \int_0^1 (n^*_\tau \times n^*_t, n)dt + \int_0^1 (n^* \times n^*_{t,\tau}, n)dt + \int_0^1 (n^* \times n^*_t, n_\tau)dt;$$

in the first integral, all three vectors are orthogonal to n^* and so the inner product is zero; the same is true in the third integral, but with n. Thus we may continue

$$= \int_0^1 (n^* \times n^*_{t,\tau}, n)dt = -\int_0^1 (n^*_t \times n^*_\tau, n)dt - \int_0^1 (n^* \times n^*_\tau, n_t)dt,$$

using integration by parts. This may be further simplified as

$$= -\int_0^1 (n^* \times n^*_\tau, n_t)dt = -\int_0^1 (n_t \times n^*, n^*_\tau)dt = -\varepsilon \int_0^1 (n, n^*_\tau)\|n_t\|dt,$$

because $n_t = \varepsilon \|n_t\| T$.

But now we get from $n^* = n \times T$ that $n^*_\tau = n_\tau \times T + n \times T_\tau$, and so $(n, n^*_\tau) = (n, n_\tau \times T) = (T \times n, n_\tau) = -(n^*, n_\tau)$. Substituting into the result of our long computation above gives

$$\frac{d}{d\tau} l(c^*_\tau) = \varepsilon \int_0^1 (n^*, n_\tau)\|n_t\|dt. \tag{5.2}$$

From (5.1) and (5.2), we see that

$$\frac{d}{d\tau}\Big(l(c^*_\tau) + A(c_\tau)\Big) = 0,$$

in other words, the value of the sum does not change while the curve is being deformed into a point. The formula in the theorem now follows easily: c_1 is a point, with zero area, while c^*_1 is a great circle, of length 2π; therefore

$$l(c^*) + A(c) = l(c^*_1) + A(c_1) = 2\pi.$$

The proof of Theorem 5.1. As in the statement of the theorem, let γ be the given curve on the surface M. At each point $\gamma(t)$, let $n(t)$ be the unit normal vector to the surface; n thus gives a curve in the two-sphere $\mathbb{S}^2$. Finally, let $w(t) = \|\gamma(t)\|$ be the speed of γ and $U(t) = \gamma(t)/w(t)$ the unit tangent vector to γ.

At this point, we shall have to make the following technical assumption: The Gauss map ζ should be submersive at a certain point of the region R (which means that the normal vector n moves in all directions); this will imply that ζ maps a small neighborhood of that point diffeomorphically to a small open set in the two-sphere. Also, R itself should lie well inside that neighborhood, which will certainly be the case if R is "sufficiently small". This condition can always be achieved by moving the surface a little bit, but we shall not worry about this point here.

The image of R under the Gauss map ζ is then also diffeomorphic to a disc, and we will be able to apply the Dual Curves Theorem below.

The condition above implies that $\dot{n}$ is never zero on R or on its boundary, and we can therefore work with the directed curve $c = (n, T) = (n, \dot{n}/\|\dot{n}\|)$ and its dual $c^* = (n^*, T^*) = (n^*, -T)$. For the sake of brevity, let us write $m = n^* = n \times T$.

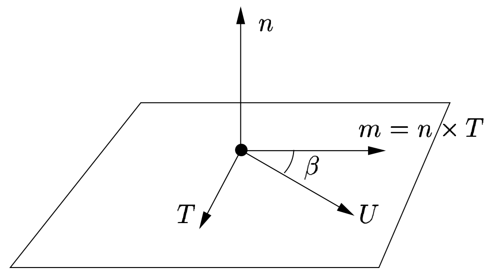

FIGURE 5. The tangent plane to M at $\gamma(t)$, which is also the tangent plane to $\mathbb{S}^2$ at $n(t)$.

The main idea is to introduce the angle $\beta(t)$ between the two unit vectors $U(t)$ and $m(t)$. We have $\cos\beta = (U, m)$, and from there by differentiation

$$-\dot{\beta} \cdot \sin\beta = (\dot{U}, m) + (U, \dot{m}). \tag{5.3}$$

Each of the two terms can be expressed using other quantities we know.

First, U is a unit vector, and so $\dot{U}$ is perpendicular to U. It can thus be written as a linear combination of n and $n \times U$, and our formula for the geodesic curvature tells us the component $\dot{U}_{n\times U} = w\kappa_g n \times U$ in the direction of $n \times U$. On the other hand, m is perpendicular to n, and so

$$(\dot{U}, m) = (\dot{U}_{n\times U}, m) = w\kappa_g(n \times U, m) = w\kappa_g(n \times U, n \times T) = w\kappa_g(U, T);$$

refering to Figure 5 above, we see that this equals

$$= w\kappa_g \cdot \sin\beta$$

since $\sin\beta = (U, T)$.

Secondly, we note that the dual curve c^* has $\dot{n}^* = \dot{m}$ parallel to T^*; this allows us to write

$$\dot{m} = (\dot{m}, T^*)T^* = -(\dot{m}, T^*)T.$$

From this we conclude that

$$(U, \dot{m}) = -(\dot{m}, T^*)(U, T) = -(\dot{m}, T^*)\sin\beta.$$

Thus (5.3) above becomes

$$\dot{\beta} = (\dot{m}, T^*) - w\kappa_g = (\dot{m}, T^*) - \|\dot{\gamma}\|\kappa_g.$$

Now integrate this over $[0, 1]$ and use the fact that $\beta(0) = \beta(1)$ to get

$$0 = \int_0^1 \dot{\beta} dt = \int_0^1 (\dot{m}, T^*) dt - \int_0^1 \|\dot{\gamma}\|\kappa_g dt = l(c^*) - \int_\gamma \kappa_g ds.$$

By the Dual Cones Theorem, $l(c^*) = 2\pi - A(c)$; but the region bounded by c (which was defined using the normal vector $n(t)$ to the surface) is exactly the image of R under the Gauss map, hence $A(c) = \int_R k d\sigma$ is the total curvature of R. So altogether, we have

$$0 = 2\pi - A(c) - \int_\gamma \kappa_g ds = 2\pi - \int_R k d\sigma - \int_\gamma \kappa_g ds,$$

which is the result we were after.

6. A topological version

Lastly, we shall look at yet another version of the Gauss-Bonnet theorem. As before, we let M be a compact orientable surface in $\mathbb{R}^3$; instead of the integral of the Gaussian curvature k over small triangles, we shall now consider

$$\text{total curvature of } M = \int_M k d\sigma.$$

To compute this integral, triangulate the surface M by using small triangles (not necessarily geodesic, but small enough for Theorem 5.2 to apply on each triangle). Say this triangulation has V vertices, E edges, and F faces. We label the vertices $P_1, P_2, \ldots, P_V$, and assume that the degree of P_j, meaning the number of edges coming into P_j, is d_j.

We now take the sum of all the identities provided by Theorem 5.2, over all triangles. Then:

(1) All integrals of the form $\int_R k d\sigma$ can be combined into $\int_M k d\sigma$.
(2) The integrals $\int_{\gamma_i} k_g ds$ cancel in pairs, since each edge occurs twice, but with opposite orientations.
(3) We sum all the changes in direction α_i by considering one vertex P_j at a time. If the change in direction is α_i, then the angle enclosed between the edges is $\pi - \alpha_i$, and all these sum up to 2π at each P_j since the triangles cover M. As there are V vertices in the triangulation, we get $\sum(\pi - \alpha_i) = 2\pi \cdot V$, or equivalently, $\sum \alpha_i = \pi \sum d_j - 2\pi V$.

Altogether, we have to sum over F triangles, and thus obtain the following formula.

$$\int_M k d\sigma + \pi \sum_{j=1}^{V} d_j - 2\pi V = 2\pi F.$$

But each edge connects exactly two vertices, and so we have

$$d_1 + d_2 + \cdots + d_V = 2E$$

for the sum of all the degrees. Substitute this into the other equation to get

$$\int_M k d\sigma = 2\pi(V - E + F) = 2\pi \cdot \chi(M).$$

We have written $\chi(M) = V - E + F$; this quantity, usually called the *Euler characteristic* of M, is independent of how we triangulate the surface—in fact, we have shown that it is equal to a quantity that depends only on M. What is more surprising is that $\chi(M)$ is actually a *topological* invariant of M, meaning it does not even depend on the Riemannian metric on the surface. It can be shown that $\chi(M) = 2 - 2g$, where g is the *genus* (or number of handles) of the surface M.

The reader should ponder the following interesting fact: while we defined the Gaussian curvature via the embedding of the surface in $\mathbb{R}^3$, it turned out to depend only on the surface itself. Now its integral over M depends only on the topology of M.

References

[1] P. Dombrowski, *150 years after Gauss' "Disquisitiones generales circa superficies curvas"*, Astérisque, vol. 62, Société Mathématique de France, Paris, 1979, With the original text of Gauss.

[2] M. Spivak, *A comprehensive introduction to differential geometry. Vol. II*, 2nd ed., Publish or Perish Inc., Wilmington, Delaware, 1979, ISBN 0-914098-83-7.
[3] M. Levi, *A "bicycle wheel" proof of the Gauss-Bonnet theorem*, Expositiones Mathematicae **12** (1994), 145–164.