Chapter 3
Local Properties of Surfaces

In this chapter, we start discussing representations that can be associated with three-dimensional shapes, where surfaces now replace curves. We begin with some basic definitions and results on the theory of surfaces in $\mathbb{R}^3$. Although some parts are redundant with the abstract discussion of submanifolds that is provided in Appendix B, we have chosen to give a more elementary presentation here, very close to [86], to ensure that this important section can be read independently.

3.1 Curves in Three Dimensions

Before addressing surfaces, we extend our developments on plane curves to the three-dimensional case. A three-dimensional parametrized curve is a function $\gamma : [a, b] \mapsto \mathbb{R}^3$. It is regular if it is C^1 and $|\dot{\gamma}| \neq 0$ for all $t \in [a, b]$. For regular curves, the unit tangent is defined by $T = \dot{\gamma}/|\dot{\gamma}|$ and the arc length is $ds = |\dot{\gamma}(t)|dt$.

Assume that γ is C^2 and parametrized by arc length. One then defines the *curvature* of γ at s by $\kappa(s) = |\dot{T}|$. This differs from the planar case, for which a sign was attributed to the curvature: here, the curvature is always non-negative.

One says that the γ is bi-regular if $\kappa(s) \neq 0$ for all s. In this case, one uniquely defines a unit vector N by the relation $\dot{T} = \kappa N$; N is perpendicular to T because T has unit norm. Finally, the binormal is the unique unit vector B which completes (T, N) into a positive orthonormal basis of[1] $\mathbb{R}^3$: $B = T \times N$. The frame (T, N, B) is called the Frénet frame, and the plane passing through $\gamma(t)$ and generated by T and N is called the osculating plane.

[1] If $h = (a, b, c)$ and $k = (a', b', c')$ are three-dimensional vectors, their cross product $h \times k$ is defined by

$$h \times k = (bc' - cb', a'c - ac', ab' - a'b).$$

It is orthogonal to both h and k and vanishes if and only if h and k are collinear. Moreover, for any third vector l: $(h \times k)^T l = \det(h, k, l)$.

L. Younes, *Shapes and Diffeomorphisms*, Applied Mathematical Sciences 171,
https://doi.org/10.1007/978-3-662-58496-5_3

The derivative of the normal is orthogonal to N and can therefore be written $\dot{N} = aT + bB$. We have $N^T T = 0$ so that $a = \dot{N}^T T = -N^T \dot{T} = -\kappa$. The *torsion* of the curve is given by $-b$ by definition and denoted τ so that $\dot{N} = -\kappa T - \tau B$. Using the fact that $\dot{B}^T T = -B^T \dot{T} = -\kappa B^T N = 0$ and $\dot{B}^T N = -\dot{N}^T B = \tau$, we have $\dot{B} = \tau N$, which provides the third equation of Frénet's formulas for three-dimensional curves:

$$\begin{cases} \partial_s T = \kappa N, \\ \partial_s N = -\kappa T - \tau B, \\ \partial_s B = \tau N. \end{cases} \tag{3.1}$$

(These equations are valid for any parametrization if one defines $\partial_s f = \dot{f}/|\dot{m}|$.) Note that, if F is the 3 by 3 rotation matrix associated with the Frénet frame, i.e., $F = [T, N, B]$, then the Frénet formulas can be written as

$$\partial_s F = F S_m, \tag{3.2}$$

where S_m is the skew-symmetric matrix

$$S_m = \begin{pmatrix} 0 & -\kappa & 0 \\ \kappa & 0 & \tau \\ 0 & -\tau & 0 \end{pmatrix}.$$

There is a three-dimensional version of Theorem 1.13. The proof, based on Eq. (3.2), is identical to the alternative proof given in two dimensions in Sect. 1.9.

Theorem 3.1 *Two C^2 curves γ and $\tilde{\gamma}$ have the same curvature and torsion as functions of their arc length if and only if there exist a rotation R, a vector b and a change of parameter ϕ such that $\tilde{\gamma} = R\gamma \circ \phi + b$.*

3.2 Regular Surfaces

Curves being represented by one parameter, one may think of surfaces as bi-parametrized objects, i.e., functions $(u, v) = m(u, v)$ defined on some subset of $\mathbb{R}^2$.

Definition 3.2 A C^p parametrized (regular) surface is a C^p map $m : U \mapsto \mathbb{R}^3$, where U is an open subset of $\mathbb{R}^2$, such that:

1. m is one-to-one and its inverse, $m^{-1} : V = m(U) \to U$ is continuous (m is a homeomorphism between U and V), i.e., if a sequence u_n is such that $m(u_n)$ converges to $p = m(u) \in V$, then u_n converges to u.
2. For all $q \in U$, the differential $dm(q)$ is one-to-one.

The last statement is equivalent to the fact that the 3 by 2 matrix of partial derivatives $[\partial_1 m, \partial_2 m]$ has rank 2. It is a direct generalization of regularity for curves.

We did not assume that curve parametrizations were one to one, but this assumption provides an important simplification for surfaces. The second part of Condition 1 also prevents situations in which the boundary of some part of the surface intersects another part (see examples).

Finally, letting $S = m(U)$ be the *range* of m, we will often abuse the terminology by saying that S (the geometric object) is a parametrized surface. However, for many interesting surfaces, it is generally impossible (or simply not convenient) to find a parametrization which satisfies the previous requirement and covers the whole surface. This is a fundamental difference with the theory of plane curves. To be able to handle interesting cases, we need to limit our requirement for parametrizations to hold only within *patches* that together cover the surface, with additional conditions ensuring that the surface is smooth and non-intersecting, and that the patches fit well together.

Definition 3.3 A subset $S \subset \mathbb{R}^3$ is a C^k regular surface if, for each $p \in S$, there exists an open set V in $\mathbb{R}^3$, with $p \in V$, and a C^k parametrization of the surface patch $V \cap S$. The local parametrizations are also called local charts.

This definition requires more than just M being covered with parametrized patches. These patches must be obtained from intersections of S with three-dimensional open sets. In particular, this prevents non-local self-intersection, since, along such an intersection, the surface would contain two local patches and would not be locally parametrizable. Figure 3.3 provides an illustration of how local parametrized patches can be combined to cover a surface.

If $m : U \to V \cap S$ is as specified in the definition, for any p in $V \cap S$, there exist parameters $(u(p), v(p))$ in U such that $m(u(p), v(p)) = p$. The functions $p \mapsto u(p)$ and $p \mapsto v(p)$ are called the local coordinates on $V \cap S$.

3.2.1 Examples

Graphs of Functions

The simplest example of a parametrized surface is the graph of a C^1 function $f : U \subset \mathbb{R}^2 \to \mathbb{R}$. The parametrization is then $m(u, v) = (u, v, f(u, v))$. Since the inverse of (u, v, z) on the surface is (u, v), this is a homeomorphism, and the differential is

$$(u, v) \mapsto \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ \partial_1 f & \partial_2 f \end{pmatrix}$$

which has rank 2.

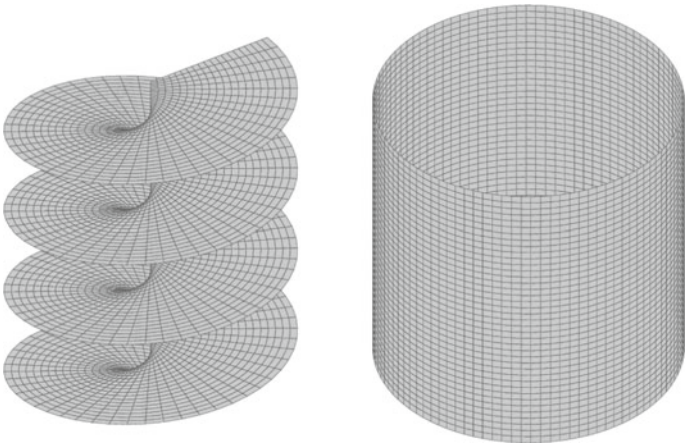

Fig. 3.1 Helicoid (left) and cylinder (right)

Helicoid

A parametrized surface does not have to be a graph. An example is the helicoid (Fig. 3.1), defined over $(0, a) \times \mathbb{R}$ by

$$m(u, v) = (u\cos(v), u\sin(v), \lambda v)$$

for some $a, \lambda > 0$.

Cylinder

The cylinder, which can be defined by the set of $m(u, v) = (\cos u, \sin u, v)$, for $u \in [0, 2\pi)$ and $v \in (-1, 1)$ (Fig. 3.1), is an example of a surface which, according to our definition, cannot be globally parametrized. This map is one-to-one and in fact a homeomorphism, and the only reason why this is not a parametrization is that we have required parametrizations to be defined on open sets ($[0, 2\pi) \times (-1, 1)$ is not open). The cylinder is a regular surface, by considering patches for the same map m, defined on $(0, 2\pi) \times (-1, 1)$ and say $(-\pi, \pi) \times (-1, 1)$.

Sphere

Consider now the example of the unit sphere (Fig. 3.2), which is denoted

$$S^2 = \left\{p \in \mathbb{R}^3, |p| = 1\right\}.$$

Like the cylinder, this surface cannot be globally parametrized. The simplest choice of local charts are the projections: $(u, v) \mapsto (u, v, \sqrt{1 - u^2 - v^2})$ and $(u, v) \mapsto (u, v, -\sqrt{1 - u^2 - v^2})$, both defined for $u^2 + v^2 < 1$, the open unit disc. The two maps cover the whole sphere, except the equator for which the third coordinate is 0. One can add other projections, like $(u, v) \mapsto (u, \pm\sqrt{1 - u^2 - v^2}, v)$,

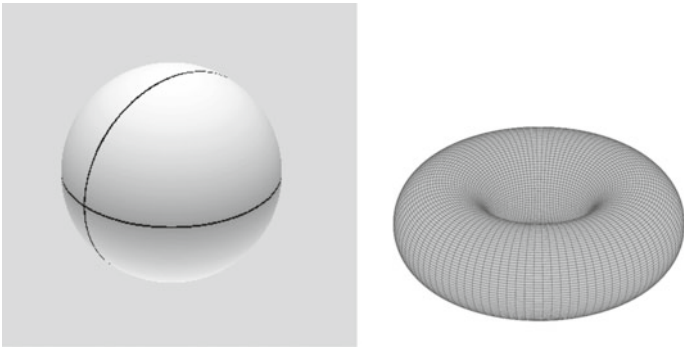

Fig. 3.2 Sphere (left) and torus (right)

$(u, v) \mapsto (\pm\sqrt{1-u^2-v^2}, u, v)$ to cover everything, or use cylindrical-like charts close to the equator.

Another useful coordinate system for the sphere is the (properly named) spherical coordinate system: $(u, v) \mapsto (\cos u \cos v, \sin u \cos v, \sin v)$. These coordinates cover the whole sphere when (u, v) varies in $[0, 2\pi) \times [-\pi/2, \pi/2]$ but they do not provide a local parametrization, since this set is not open (and the map is not one-to-one for $v = -\pi/2$ and $v = \pi/2$). Restricting to the open intervals requires using other charts to cover the meridian $u = 0$, for example the same coordinates on $(-\pi, \pi) \times (-\pi/2, \pi/2)$ which now only leave the poles uncovered. A neighborhood of the poles can be covered by the previous projection maps.

Torus

The torus (a surface with a shape like a donut, see Fig. 3.2) can be represented as the image of $[0, 2\pi) \times [0, 2\pi)$ under the map

$$m(u, v) = ((R + r\cos v)\cos u, (R + r\cos v)\sin u, r\sin v),$$

where $0 < r < R$, which is one-to-one but once again not defined on an open set. The whole torus can be covered by considering this map restricted to open subsets of $[0, 2\pi) \times [0, 2\pi)$. Let us check that the rank of the differential of m is always 2. We have

$$dm = \begin{pmatrix} -(R + r\cos v)\sin u & -r\sin v\cos u \\ (R + r\cos v)\cos u & -r\sin v\sin u \\ 0 & r\cos v \end{pmatrix}.$$

The determinant of the first two rows is $-r\sin v(R + r\cos v)$. Since $r < R$, it can only vanish when $\sin v = 0$. For the remaining two determinants, which are $r(R + r\cos v)\sin u\cos v$ and $r(R + r\cos v)\cos u\cos v$, to vanish together, one needs $\cos v = 0$. So at least one of the three two-by-two determinants does not vanish (Fig. 3.3).

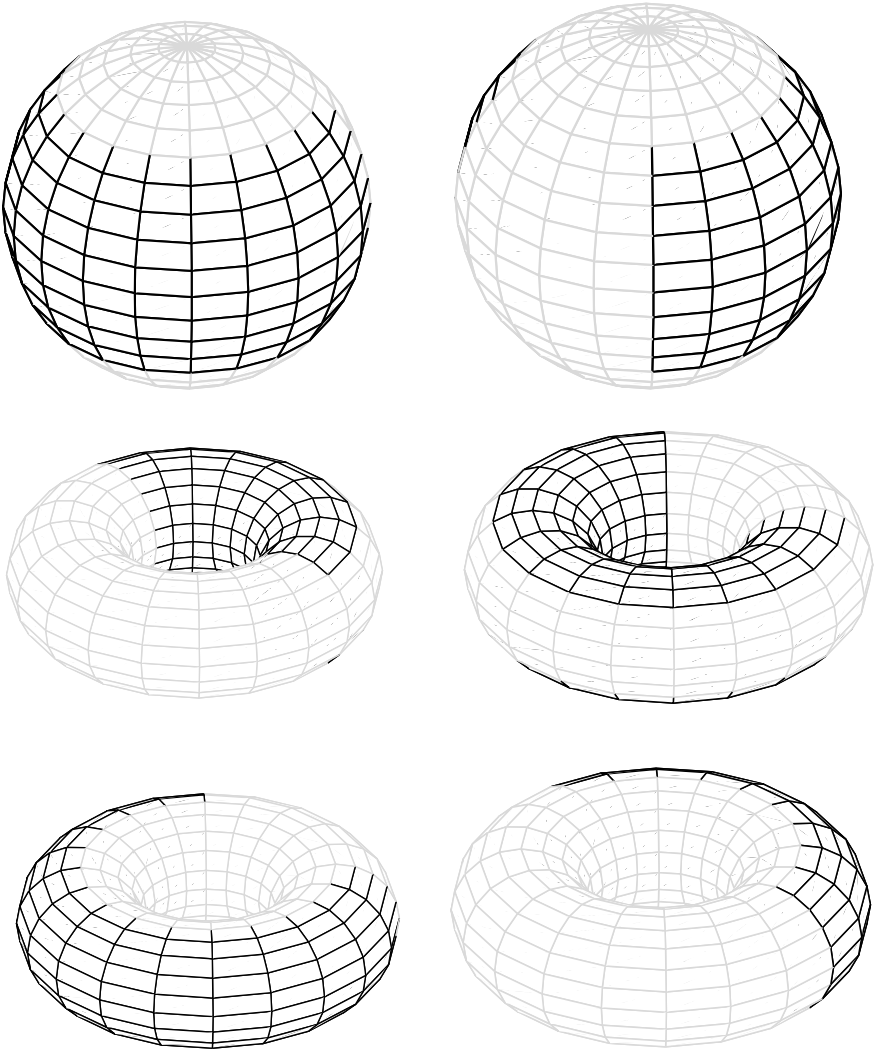

Fig. 3.3 Examples of decompositions in local charts for the sphere and the torus. Parametrizations are represented by grids over the surface, black inside local patches and gray outside

A Non-regular Surface

As a last example, consider the set S defined by

$$S = \{m(u, v) : (u, v) \in (-1, 1) \times (-\pi/2, \pi)\}$$

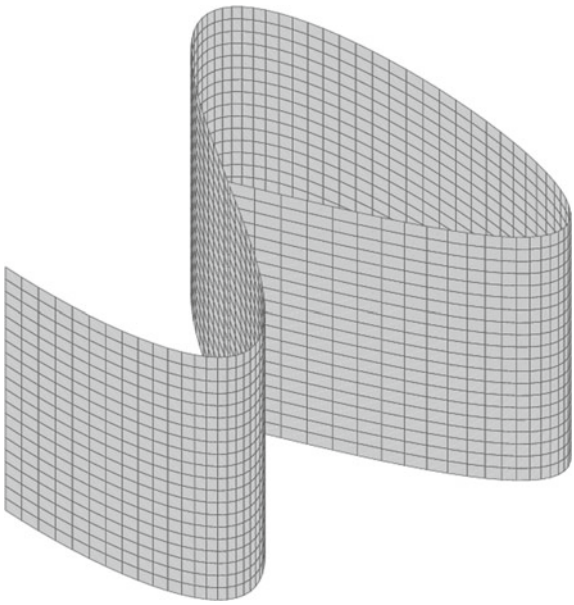

Fig. 3.4 A non-regular surface

with

$$m(u, v) = \begin{cases} \cos v \\ \sin 2v \\ u \end{cases}$$

(see Fig. 3.4). The parametrization is defined over an open set, it is one to one and

$$dm = \begin{pmatrix} 0 & -\sin v \\ 0 & 2\cos 2v \\ 1 & 0 \end{pmatrix}$$

has rank two everywhere. S is not a parametrized surface, however, because

$$\lim_{v \to -\pi/2} m(0, v) = m(0, \pi/2),$$

which contradicts the assumption that m^{-1} is continuous. The same contradiction can be obtained for $S \cap V$ where V is any open subset of $\mathbb{R}^3$ that contains 0, so that S is not a regular surface either.

3.2.2 Changing Coordinates

As we have seen, several different valid parametrizations can be defined at a single point of a surface. Like for curves, "geometric" properties or quantities should not depend on the parametrization. We will define a few of them in the following: normals, curvature, length, area, etc.

It can be deduced from the requirements in Definition 3.2 that changes of coordinates are C^1 homeomorphisms. To be more specific, assume that in a neighborhood V of a point p on S, there exist two parametrizations $m : U \to V$ and $\tilde{m} : \tilde{U} \to V$. Then, because of the invertibility of the parametrization, one can go from U to V via m, then from V to $\tilde{U}$ via the inverse of $\tilde{m}$. The resulting map, $\varphi = \tilde{m}^{-1} \circ m : U \to \tilde{U}$, is called a change of coordinates, and is a *diffeomorphism* between U and $\tilde{U}$ (it is C^1, invertible, with a C^1 inverse). This consequence of Definition 3.2 can be proved using the inverse mapping theorem.

3.2.3 Implicit Surfaces

An implicit surface is defined by an equation of the form $f(p) = 0$ where $f : \mathbb{R}^3 \to \mathbb{R}$ is a scalar function which is such that $\nabla f(p) \neq 0$ if $f(p) = 0$. In this case, the set

$$S = \left\{ p \in \mathbb{R}^3,\, f(p) = 0 \right\}$$

is a regular surface. (This is a consequence of the implicit function theorem.)

3.3 Tangent Planes and Differentials

3.3.1 Tangent Planes

For a curve, we were able to define a unique unit tangent, but this is obviously no longer possible for surfaces. Still, curves provide a simple way to define tangent vectors to surfaces.

A curve $m : I \to \mathbb{R}^3$ is supported by a surface S if and only if, for all $t \in I$, one has $m(t) \in S$. We have the following definition:

Definition 3.4 Let S be a regular surface. A vector $T \in \mathbb{R}^3$ is tangent to S at a point $p \in S$ if and only if, for some $\varepsilon > 0$, there exists a C^1 curve $\gamma : (-\varepsilon, \varepsilon) \to S$ such that $\gamma(0) = p$ and $\dot{\gamma}(0) = T$.

Assume, in the previous definition, that ε is chosen small enough so that the curve γ is completely inscribed in a parametrized patch of the surface S. Let $m : (u, v) \mapsto m(u, v)$ be the parametrization. Since m is one-to-one, one can express

$\gamma(t) = m(u(t), v(t))$. The plane curve $t \mapsto (u(t), v(t))$ is the expression of γ in the local coordinates. From the chain rule, we have

$$\dot{\gamma} = \dot{u}\partial_1 m + \dot{v}\partial_2 m.$$

Thus, $\dot{\gamma}$ must be a linear combination of the two independent vectors $\partial_1 m$ and $\partial_2 m$. Conversely, if $p = m(u_0, v_0)$, then, for any $\alpha, \beta \in \mathbb{R}$, the vector $\alpha\partial_1 m + \beta\partial_2 m$ is the derivative of $t \mapsto m(u_0 + \alpha t, v_0 + \beta t)$ and is therefore tangent to S at p. This proves the following proposition:

Proposition 3.5 *Let S be a regular surface, $p \in S$ and $m : U \to S$ a parametrization of S in a neighborhood of p. The set of tangent vectors to S at p is the plane generated by $\partial_1 m$ and $\partial_2 m$.*

The tangent plane to S at p will be denoted $T_p S$. Although the generating vectors $\partial_1 m$ and $\partial_2 m$ depend on the local parametrization m, the plane itself does not, because we gave a parametrization-independent definition of tangent vectors.

If S is defined implicitly by $f(p) = 0$, the tangent plane at p is characterized by the equation $\nabla f(p)^T T = 0$ (recall that f must be such that $\nabla f(p) \neq 0$ if $f(p) = 0$). Indeed, if γ is a curve on S, then $f \circ \gamma(t) = 0$ for all t, and the chain rule implies: $\nabla f(\gamma(0))^T \dot{\gamma}(0) = 0$. This implies that $T_p S \subset (\nabla f(p))^\perp$. Because $T_p S$ and $(\nabla f(p))^\perp$ have the same dimension (two), they coincide.

3.3.2 Differentials

Differentials describe how measurements made on a surface vary locally. Consider a scalar function $f : S \to \mathbb{R}$ and take a local parametrization on S, $m : U \to V \cap S$. For $(u, v) \in U$, we can define the function $f_m(u, v) = f(m(u, v))$; this is a function from an open subset of $\mathbb{R}^2$ to $\mathbb{R}$, which provides the expression of f in the local system of coordinates: we have $f(p) = f_m(u(p), v(p))$. We have the following definition:

Definition 3.6 Let S be a regular surface. A function $f : S \to \mathbb{R}$ is C^1 at $p \in S$ if and only if, for some local parametrization m on S around p, the function f_m is C^1 at $m^{-1}(p)$.

We say that f is C^1 on S if it is C^1 at all $p \in S$.

(Because changes of coordinates are C^1, the definition does not depend on the choice of local parametrization at p.)

We now want to evaluate the effect that small variations in p have on the function f, i.e., we want to define the derivative of f. Usually, a first-order variation of $p \in \mathbb{R}^3$ in the direction h is represented by $p + \varepsilon h$, with small ε. This cannot be applied to S, since there is no reason for $p + \varepsilon h$ to belong to S if p does. It is reasonable, and rather intuitive, to define a first-order variation of p as an element of a curve on S containing p. This leads to:

Definition 3.7 Let S be a regular surface and $p \in S$. A first-order variation of p in the direction $h \in \mathbb{R}^3$ is a C^1 curve $\gamma : (-\varepsilon, \varepsilon) \to S$ such that $\gamma(0) = p$ and $\dot{\gamma}(0) = h$.

Note that, from this definition, first-order variations on S can only arise in directions which are tangent to S.

Now, we can define the differential of a scalar function f defined on S as the limit (if it exists) of the ratio $(f(\gamma(\delta)) - f(p))/\delta$ as δ tends to 0, γ being a first-order variation of p. This will be denoted $df(p)h$, with $h = \dot{\gamma}(0)$. Implicit in this notation is the fact that this limit only depends on $\dot{\gamma}(0)$, which is true if f is C^1 as stated in the next proposition.

Proposition 3.8 *Let f be a C^1 scalar function on a regular surface S. Then, for any $p \in S$, and $h \in T_pS$, the differential of f at p in the direction h exists, and is equal to the limit of the ratio $(f(\gamma(\delta)) - f(p))/\delta$ for any C^1 curve γ on S with $\gamma(0) = p$ and $\dot{\gamma}(0) = h$.*

Proof What we need to prove is that the limit of the ratio exists for any γ and only depends on h. Take a local parametrization m around p. We know that the function $f(m(u, v))$ is C^1, and letting $\gamma(t) = m(u(t), v(t))$, we have

$$\begin{aligned}\lim_{\delta \to 0} \frac{f(\gamma(\delta)) - f(p)}{\delta} &= \lim_{\delta \to 0} \frac{f_m(u(\delta), v(\delta)) - f_m(u(0), v(0))}{\delta} \\ &= \partial_1 f_m \dot{u}(0) + \partial_2 f_m \dot{v}(0).\end{aligned}$$

This proves the existence of the limit. We have $h = \dot{\gamma}(0) = \dot{u}(0)\partial_1 m + \dot{v}(0)\partial_2 m$: since $(\partial_1 m, \partial_2 m)$ has rank 2, $\dot{u}(0)$ and $\dot{v}(0)$ are uniquely specified by h and thus the limit above only depends on h. The notation $df(p)h$ is therefore valid. □

Note that the expression provided in this proof shows that $df(p)h$ is linear with respect to h. In other terms, $df(p)$ is a linear form from T_pS to $\mathbb{R}$. Most of the time, the computation of $df(p)$ is easy, because f can be expressed as the restriction to S of a differentiable function which is defined on $\mathbb{R}^3$. In this case, $df(p)h$ coincides with the usual differential of f, but restricted to the two-dimensional plane T_pS.

The proof above also provides a simple way to compute differentials in local charts: let $f : S \to \mathbb{R}$ be C^1, $p \in S$ and m be a local parametrization around p. Then, if $h = \alpha\partial_1 m + \beta\partial_2 m$, we have

$$df(p)h = \alpha\partial_1 f_m + \beta\partial_2 f_m. \tag{3.3}$$

When f is a vector-valued function ($f : S \to \mathbb{R}^d$), the differential $df(p)$ is defined in the same way, and is also vector-valued. It is a linear map from T_pS to $\mathbb{R}^d$.

The simplest examples of differentiable maps are the coordinates: if $m : U \to V \cap S$ is a local chart, the function $f = m^{-1}$ is such that $f_m(u, v) = (u, v)$, which is the identity map, and therefore differentiable. In particular, the coordinates: $p \mapsto u(p)$ and $p \mapsto v(p)$ are scalar differentiable maps. If $T = \alpha\partial_1 m + \beta\partial_2 m$, we have $du(p)T = \alpha$, $dv(p)T = \beta$ and $df(p)T = (\alpha, \beta)$.

Consider now the example of the sphere S^2. The tangent plane is easy to describe if one uses the fact that S^2 can be defined by the implicit equation $|p|^2 = 1$. If $\phi(p) = |p|^2$, we have $\nabla\phi(p)^T h = 2p^T h$ so that h is tangent to S^2 at p if and only if $p^T h = 0$ (h is perpendicular to p).

Fix a vector $p_0 \in S^2$ and consider the function $f(p) = p^T p_0$. Then, since f is well-defined on $\mathbb{R}^3$, we can use its restriction, which yields $df(p)h = h^T p_0$. This was an easy result, but for illustration purposes, let us retrieve it via local charts, which will require a little more computation.

Consider the parametrization $m(u, v) = (\cos u \cos v, \sin u \cos v, \sin v)$. Then,

$$\partial_1 m = (-\sin u \cos v, \cos u \cos v, 0) \text{ and}$$

$$\partial_2 m = (-\cos u \sin v, -\sin u \sin v, \cos v).$$

A straightforward computation shows that both $\partial_1 m$ and $\partial_2 m$ are orthogonal to $m(u, v)$. In the chart, letting $p_0 = (a, b, c)$, the function f_m is

$$f_m(u, v) = a \cos u \cos v + b \sin u \cos v + c \sin v.$$

Obviously, $\partial_1 f_m = {p_0}^T \partial_1 m$ and $\partial_2 f_m = {p_0}^T \partial_2 m$, so that, if $h = \alpha\partial_1 m + \beta\partial_2 m$, we get, by Eq. (3.3),

$$df(p)h = \alpha\partial_1 f_m + \beta\partial_2 f_m = {p_0}^T h.$$

3.4 Orientation and Normals

Let S be a surface and m a local parametrization on S. The vector $\partial_1 m \times \partial_2 m$ is non-vanishing and orthogonal to both $\partial_1 m$ and $\partial_2 m$. Since $\partial_1 m$ and $\partial_2 m$ generate $T_p S$ at $p = m(u, v)$, $\partial_1 m \times \partial_2 m$ is normal to the tangent plane at p.

In particular, the vector $N = \partial_1 m \times \partial_2 m / |\partial_1 m \times \partial_2 m|$ is a unit normal to the tangent plane. One also says that N is *normal to the surface* S. Since unit normals to a plane are defined up to a sign change, the one obtained from another parametrization must be either N or $-N$. This leads to the following definition:

Definition 3.9 Two local parametrizations, m and $\tilde{m}$, on a regular surface S have the same orientation at a given point at which they are both defined if

$$\frac{\partial_1 m \times \partial_2 m}{|\partial_1 m \times \partial_2 m|} = \frac{\partial_1 \tilde{m} \times \partial_2 \tilde{m}}{|\partial_1 \tilde{m} \times \partial_2 \tilde{m}|}$$

and have opposite orientation otherwise.

The surface S is said to be orientable if it can be covered by local parametrizations that have the same orientation wherever they intersect.

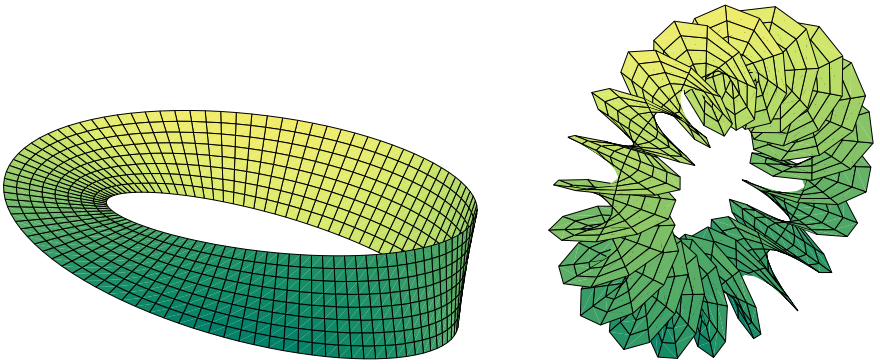

Fig. 3.5 Two examples of non-orientable surfaces. On the left is the Möbius band; the surface on the right is similar, with an odd number of twists

A surface is therefore orientable if there is a consistent (continuous) definition of a normal all over it. Not all surfaces are orientable (Fig. 3.5). A typical example is a twisted ring (the Möbius band).

3.5 Integration on an Orientable Surface

Let S be an orientable surface and $f : S \to \mathbb{R}$ be a continuous function. We want to compute the integral of f over S. We first define it within a local chart. Let $m : U \to V \cap S$ be a parametrized patch of the surface S. To motivate the definition, let U be divided into small rectangular cells (neglecting boundary issues). Consider a cell of the form $(u_0 - \varepsilon/2, u_0 + \varepsilon/2) \times (v_0 - \varepsilon/2, v_0 + \varepsilon/2)$. In this cell, we can make a first-order expansion of m in the form

$$m(u, v) = m(u_0, v_0) + (u - u_0)\partial_1 m(u_0, v_0) + (v - v_0)\partial_2 m(u_0, v_0) + o(\varepsilon)$$

so that, at first order, the image of the rectangular cell by m is a parallelogram in space, centered at $p_0 = m(u_0, v_0)$, namely

$$\sigma_0 = \{p_0 + \alpha\partial_1 m + \beta\partial_2 m, \alpha \in (-\varepsilon/2, \varepsilon/2), \beta \in (-\varepsilon/2, \varepsilon/2)\}.$$

Its area is given by $\varepsilon^2|\partial_1 m \times \partial_2 m|$, and the integral of a function f over this parallelogram can legitimately be estimated by $\varepsilon^2 f(p_0)|\partial_1 m \times \partial_2 m|$. Summing over cells and letting ε tend to 0 leads to the following definition:

Definition 3.10 Let f be a function defined on a regular surface S, and $m : U \to V \cap S$ a regular patch on S. The integral of f on $V \cap S$ is defined and denoted by

$$\int_{V\cap S} f(p)d\sigma_S(p) = \int_U f_m(u,v)\,|\partial_1 m \times \partial_2 m|\,du\,dv\,. \tag{3.4}$$

The integral of f over the whole surface S is defined as the sum of such integrals over non-overlapping local patches that cover S (maybe leaving out a finite number of curves or points on S). It is denoted

$$\int_S f(p)\,d\sigma_S(p) \text{ or } \int_S f\,d\sigma_S.$$

This can be shown to be independent of the chosen family of patches. The notation $d\sigma_S$ refers to the area form on S, defined on a local chart by $d\sigma_S = |\partial_1 m \times \partial_2 m|\,du\,dv$.

Note that the area form that we have defined here is a special case of a volume form in an arbitrary finite-dimensional manifold. For this reason, it is also often called the volume form of S (even though it measures areas).

Another (equivalent) way to globally define the integral is to use partitions of unity. Given a family $((U_i, m_i), i = 1, \ldots, n)$ of local parametrizations which cover the surface (so that $\bigcup_i m_i(U_i) = S$), but may overlap, one defines a partition of unity as a family of continuous functions $(\omega_i, i = 1, \ldots, n)$ where each ω_i is defined on S and takes values in $[0, 1]$, with $\omega_i(p) = 0$ if $p \notin m_i(U_i)$, and for all $p \in S$,

$$\sum_{i=1}^{n} \omega_i(p) = 1.$$

Such partitions of unity always exist, and one can define

$$\int_S f(p)\,d\sigma_S(p) = \sum_{i=1}^{N} \int_{U_i} \omega_i(m_i(u,v))\,f_{m_i}(u,v)|\partial_1 m_i \times \partial_2 m_i|\,du\,dv\,.$$

Here also, the result does not depend on the local parametrizations, or on which partition of unity is chosen.

That the right-hand side of (3.4) does not depend on the chosen parametrization should be clear from the approximation process which led to its definition (which was purely geometric), and can be checked directly as follows. Let $\tilde{m} : \tilde{U} \to V \cap S$ be another parametrization of the same patch. For $p \in V \cap S$, the equation $p = m(u,v) = \tilde{m}(\tilde{u}, \tilde{v})$ provides a relation between homologous coordinates given by

$$\begin{cases} \partial_1 m = \partial_1 \tilde{u}\partial_1 \tilde{m} + \partial_1 \tilde{v}\partial_2 \tilde{m} \\ \\ \partial_2 m = \partial_2 \tilde{u}\partial_1 \tilde{m} + \partial_2 \tilde{v}\partial_2 \tilde{m}. \end{cases}$$

The left-hand sides are computed at (u, v) and the right-hand sides at $(\tilde{u}, \tilde{v})$. This implies

$$\partial_1 m \times \partial_2 m = (\partial_1 \tilde{u} \partial_2 \tilde{v} - \partial_1 \tilde{v} \partial_2 \tilde{u})\, \partial_1 \tilde{m} \times \partial_2 \tilde{m}\,.$$

Letting ϕ be the change of variables ($\phi(u, v) = (\tilde{u}, \tilde{v})$), this is $\partial_1 m \times \partial_2 m = (\det \phi)(\partial_1 \tilde{m} \times \partial_2 \tilde{m}) \circ \phi$. Therefore

$$\begin{aligned}\int_U f(m(u, v))\, |\partial_1 m \times \partial_2 m|\, du\, dv \\ = \int_U f(\tilde{m} \circ \phi(u, v))\, |\partial_1 m \times \partial_2 m| \circ \phi(u, v)\, |\det \phi(u, v)|\, du\, dv \\ = \int_{\tilde{U}} f(\tilde{m}(\tilde{u}, \tilde{v}))\, |\partial_1 \tilde{m} \times \partial_2 \tilde{m}|\, d\tilde{u}\, d\tilde{v}.\end{aligned}$$

As an example, we compute the area of the unit sphere, which can be parametrized (poles excepted) by $m(u, v) = (\cos u \cos v, \sin u \cos v, \sin v)$. Then

$$\begin{aligned}\partial_1 m &= (-\sin u \cos v, \cos u \cos v, 0), \\ \partial_2 m &= (-\cos u \sin v, -\sin u \sin v, \cos v)\end{aligned}$$

and $|\partial_1 m \times \partial_2 m|^2 = \cos^2 v$, so that

$$\int_{S^2} d\sigma = \int_0^{2\pi} \int_{-\pi/2}^{\pi/2} \cos v\, du\, dv = 2\pi [\sin v]_{-\pi/2}^{\pi/2} = 4\pi.$$

3.6 Regular Surfaces with Boundary

Consider the surface S defined by $x^2 + y^2 < 1$, $z = 0$, which is the unit disc in the horizontal plane. It is natural to define the boundary of S to be the circle $x^2 + y^2 < 1$, $z = 0$. Such a definition cannot coincide with the topological boundary in $\mathbb{R}^3$, $\partial S = \bar{S} \setminus \mathring{S}$, which would be the unit disc $\bar{S}$ defined by $x^2 + y^2 \leq 1$, $z = 0$ (because S has an empty interior in $\mathbb{R}^3$). Because of this, one defines the boundary of a surface S by $\partial S = \bar{S} \setminus S$ (and never use the topological boundary).

For a regular surface to be a "regular surface with boundary", some additional requirements are made to ensure that the boundary is locally a smooth curve.

Definition 3.11 Let S be a regular surface. One says that $p \in \partial S$ is a regular boundary point if there exists a parametrized surface $m : U \to \mathbb{R}^3$, where U is open in $\mathbb{R}^2$ with $0 \in U$ such that $m(0, 0) = p$, and, if $U^+ = \{(u, v) \in U : v > 0\}$ and $U^0 = \{(u, v) \in U : v = 0\}$, one has $m(U^+) = m(U) \cap S$ and $m(U^0) = m(U) \cap \partial S$.

One says that S is a regular surface with boundary if every point $p \in \partial S$ is regular.

Equivalently, $p \in \partial S$ is regular if there exists a regular surface $\tilde{S}$ such that $S \subset \tilde{S}$, $p \in \tilde{S}$ and $\partial S \cap \tilde{S}$ is a C^1 regular curve on $\tilde{S}$.

With this definition, the horizontal unit disc considered above is a regular surface with boundary, and its boundary is the horizontal unit circle. If one removes the origin from this unit disc, one still has a regular surface, but not "with boundary" anymore, because 0, which is now in the boundary, is not a regular point. Similarly, a triangular region, such as $\{x > 0, y > 0, x + y < 1, z = 0\}$, is not a regular surface with boundary, because its boundary has angles.

Let S be a regular surface with boundary and assume that S is oriented. Let $p \in \partial S$ and m be a local parametrization such as the one defined in Definition 3.11. Assume that $m : U^+ \to S$ is positively oriented (otherwise, take its composition with the transformation $(u, v) \mapsto (-u, v)$). Then one defines the unit tangent and normal to ∂S at p by $T_{\partial S}(p) = \partial_1 m(0, 0)/|\partial_1 m(0, 0)|$ and $N_{\partial S}(p) = N_S(p) \times T_{\partial S}(p)$ where $N_S(p)$, defined by

$$N_S(p) = \frac{\partial_1 m(0, 0) \times \partial_2 m(0, 0)}{|\partial_1 m(0, 0) \times \partial_2 m(0, 0)|},$$

extends the normal to S to its boundary. We let the reader check that this definition does not depend on the chosen parametrization m (or refer to the general argument made in Sect. B.7.3). With this definition, $N_{\partial S}$ is the inward pointing normal to ∂S in the tangent plane to S.

Note that the term "boundary" is not the only difference between the terminology used for surfaces and the one for standard topology. Here is another example.

Definition 3.12 One says that a regular surface S is a "closed surface", or a "surface without boundary", if and only if S is a compact subset of $\mathbb{R}^3$.

With this definition, a sphere and a torus are closed surfaces. However, the horizontal plane $z = 0$ is a closed subset of $\mathbb{R}^3$ and a regular surface, but not a closed surface according to this definition.

3.7 The First Fundamental Form

3.7.1 Definition and Properties

Let S be a regular surface. When h and k are two tangent vectors at $p \in S$, their dot product in $\mathbb{R}^3$ will be denoted $\langle h\ ,\ k\rangle_p$. It is simply the usual dot product, the sum of products of coordinates, but gets a specific notation because it is restricted to T_pS. The associated quadratic form is called *the first fundamental form*, and denoted

$$I_p(h) := |h|_p^2 . \tag{3.5}$$

This form is the key instrument for metric measurements on surfaces. Although its definition is straightforward, one must remember that surfaces are mostly described by local charts, and the expression of the form in such charts is not the standard

norm anymore. Indeed, let m be a local parametrization around p, and $h = \alpha\partial_1 m + \beta\partial_2 m \in T_p S$. Then

$$\begin{aligned} I_p(h) &= \alpha^2 \langle \partial_1 m\,,\, \partial_1 m \rangle_p + 2\alpha\beta \langle \partial_1 m\,,\, \partial_2 m \rangle_p + \beta^2 \langle \partial_2 m\,,\, \partial_2 m \rangle_p \\ &= \alpha^2 E + 2\alpha\beta F + \beta^2 G \end{aligned}$$

with the notation

$$E = \langle \partial_1 m\,,\, \partial_1 m \rangle_p,\ F = \langle \partial_1 m\,,\, \partial_2 m \rangle_p,\ G = \langle \partial_2 m\,,\, \partial_2 m \rangle_p. \tag{3.6}$$

E, F and G are the coefficients of the first fundamental form in the chart. They depend on the parameters u, v.

The following proposition allows one to use convenient local charts around a given point.

Proposition 3.13 *If S is a regular surface and $p \in S$, there exists a local parametrization $m : U \mapsto S$ around p such that $(\partial_1 m, \partial_2 m)$ is orthogonal on U.*

(Note that this proposition does not hold if 'orthogonal' is replaced with 'orthonormal'.)

3.7.2 Geodesics

The first fundamental form provides all the information required to compute lengths of curves on S: let γ be such a curve; assuming that γ is contained in a parametrized patch and letting $\gamma(t) = m(u(t), v(t))$, we have

$$|\dot\gamma|^2 = |\dot u \partial_1 m + \dot v \partial_2 m|^2 = \dot u^2 E + 2\dot u \dot v F + \dot v^2 G$$

so that the length of the curve from its expression in local coordinates is provided by

$$\text{length}(\gamma) = \int_a^b \sqrt{\dot u^2 E(u,v) + 2\dot u \dot v F(u,v) + \dot v^2 G(u,v)}\, dt.$$

Similarly, one defines the energy of a curve γ by

$$\text{energy}(\gamma) = \frac{1}{2}\int_a^b |\dot\gamma|^2 dt = \frac{1}{2}\int_a^b \left(\dot u^2 E(u,v) + 2\dot u \dot v F(u,v) + \dot v^2 G(u,v)\right) dt.$$

Curves of minimal energy on a surface are called *minimizing geodesics*, as formalized by the following definition.

Definition 3.14 Given two points p and p' on a surface M, a curve γ on M achieving the minimum energy among all piecewise C^1 curves on M linking p and p' is called a (minimizing) geodesic.

In addition to minimizing the energy, it can be shown that geodesics are curves of minimal length between two points [86, 87]. Moreover, if we define

$$d_M(p, p') = \inf \left\{ \text{length}(\gamma) : \gamma \text{ piecewise } C^1 \text{ on } M \right\}$$

then d_M is a distance on M, called the geodesic distance. Therefore if γ is a minimizing geodesic between p and p' then $\text{length}(\gamma) = d_M(p, p')$.

Minimizing geodesics between two given points do not always exist, however. Let M be, for example, the plane $z = 0$ with the point $(0, 0, 0)$ removed, which is a regular surface. Then the geodesic distance between p and $-p$ in M is $2|p|$, but this distance cannot be achieved because the optimal curve must be a straight line containing 0. We however have the following theorem, which is an application of the standard Hopf–Rinow theorem (see [86], for example).

Theorem 3.15 *If M is a (topologically) closed surface, then there exists a minimizing geodesic connecting any pair of its points.*

If γ is a minimizing geodesic between p and p', and $h(t)$ is for all t a vector tangent to the surface at $\gamma(t)$, one can define, for small ε, a one-parameter family of curves $\tilde{\gamma}(t, \varepsilon)$ such that $\tilde{\gamma}(t, 0) = \gamma(t)$ and $\partial_\varepsilon \tilde{\gamma}(t, 0) = h(t)$. Since γ is minimizing, the function $\varepsilon \mapsto \text{energy}(\tilde{\gamma}(\cdot, \varepsilon))$ has a vanishing derivative at $\varepsilon = 0$. This derivative is given by

$$\int_a^b \dot{\gamma}^T \dot{h} dt = -\int_a^b \ddot{\gamma}^T h dt$$

by integration by parts. The fact that this expression vanishes for any h tangent to the surface along γ implies that the "acceleration" $\ddot{\gamma}$ is normal to the surface. By extension, curves satisfying this property are also called geodesics. They generalize the notion of straight lines in a plane.

Definition 3.16 A C^2 regular curve γ on M is called a geodesic if its second derivative $\ddot{\gamma}$ is always normal to M.

Note that, using $\partial |\dot{\gamma}|^2 = 2\dot{\gamma}^T \ddot{\gamma} = 0$ for geodesics, one finds immediately that such curves have "constant speed": $|\dot{\gamma}| = \text{const}$.

Let us compute the geodesics of the unit sphere. Such geodesics must satisfy $|\gamma(t)| = 1$ for all t and, in order to be normal,

$$\ddot{\gamma}(t) = \lambda(t)\gamma(t)$$

for some real-valued function λ. On the sphere, we can write, since $\gamma^T \dot{\gamma} = 0$,

$$0 = \partial_t \gamma^T \dot{\gamma} = |\dot{\gamma}|^2 + \lambda(t)|\gamma|^2,$$

which implies (because $\dot{\gamma}$ is constant along geodesics and $|\gamma| = 1$) that λ is constant. So geodesics must satisfy the equation $\ddot{\gamma} = \lambda\gamma$. By making a constant time change, we can assume that $|\dot{\gamma}| = -\lambda(t) = 1$, and that γ is parametrized by arc length. Since $\partial\ddot{\gamma} = \dot{\gamma}$, we see that the curve has unit curvature and zero torsion and therefore coincides with a portion of unit circle. The only unit circles included in the sphere must be centered at 0, and constitute the great circles on the sphere. So we find that geodesics on the sphere are great circles parametrized at constant speed.

Finally, we note that the first fundamental form also determines the area form used in the computation of integrals over the surface. Indeed, one can easily check that $|\partial_1 m \times \partial_2 m| = \sqrt{EG - F^2}$ (both terms are equal to $|\partial_1 m|\,|\partial_2 m|\,|\sin\theta|$ where θ is the angle between the two tangent vectors) so that

$$d\sigma_S = \sqrt{EG - F^2}\, du\, dv. \tag{3.7}$$

3.7.3 The Divergence Theorem on Surfaces

A vector field on S is a function $h : S \to \mathbb{R}^3$ such that, for all p, $h(p) \in T_pS$. We start with a simple definition of the divergence of a C^1 vector field.

Definition 3.17 Let h be a C^1 vector field on a regular surface S. The divergence of h on S is defined by

$$\mathrm{div}_S h(p) = {e_1}^T dh(p) e_1 + {e_2}^T dh(p) e_2 \tag{3.8}$$

whenever e_1, e_2 is a positively oriented orthonormal basis of T_pM (the result being independent of the choice made for e_1, e_2).

In this definition, $dh(p)$ is a linear transformation between T_pS and $\mathbb{R}^3$. If h is defined on S and takes values in $\mathbb{R}^3$ (not necessarily in TS), the definition remains meaningful. We will use the notation $\mathrm{div}'_S(h)$ for the left-hand side of (3.8) in that case. In fact, if h decomposes as $h = h_T + \mu N$ where h_T is a vector field on S, we have

$$\mathrm{div}'_S(h) = \mathrm{div}_S(h_T) + \mu \mathrm{div}'_S(N). \tag{3.9}$$

Another way of understanding the definition is by introducing the orthogonal projection on T_pS (denoted π_{T_pS}) and the operator

$$\nabla_S h(p) = \pi_{T_pS} \circ dh(p) : T_pS \to T_pS. \tag{3.10}$$

This operator is the covariant derivative on S, as described in Appendix B, and Definition 3.17 simply says that

$$\mathrm{div}_S h(p) = \mathrm{trace}(\nabla_S h(p)). \tag{3.11}$$

Note that we have, for $\xi \in T_pS$

$$\nabla_S h(p)\xi = dh(p)\xi - ((dh(p)\xi)^T N)N.$$

This definition can be made explicit in a chart. This yields the following proposition (the proof, which is just a computation, is left to the reader):

Proposition 3.18 *If m is a local chart on S and the C^1 vector field h decomposes as $h = \alpha\partial_1 m + \beta\partial_2 m$ in this chart, we have*

$$\operatorname{div}_S h = \partial_1\alpha + \partial_2\beta + (\alpha\partial_1\rho + \beta\partial_2\rho)/\rho, \tag{3.12}$$

where $\rho = |\partial_1 m \times \partial_2 m| = \sqrt{EG - F^2}$.

We also have the nice formula, still valid in a chart, that says that

$$\partial_1 h \times \partial_2 m + \partial_1 m \times \partial_2 h = \rho(\operatorname{div}_S h)N. \tag{3.13}$$

This result is a direct consequence of the following simple computation in linear algebra, the proof of which is left to the reader.

Lemma 3.19 *Let A be a linear operator from M, an oriented two-dimensional linear subspace of $\mathbb{R}^3$, to $\mathbb{R}^3$. Let n be the unit normal to M. Define, for $e_1, e_2 \in M$,*

$$\phi_A(e_1, e_2) = (Ae_1)^T(e_2 \times n) + (Ae_2)^T(n \times e_1).$$

Then, there exists a real number $\rho(A)$ such that

$$\phi_A(e_1, e_2) = \rho(A)\det(e_1, e_2, n),$$

which is also equal to $\rho(A)|e_1 \times e_2|$ if e_1, e_2 are positively oriented. Moreover, we have

$$\rho(A) = \operatorname{trace}((\mathrm{Id} - nn^T)A), \tag{3.14}$$

where $(\mathrm{Id} - nn^T)A$ (which is A followed by the projection on M) is considered as an operator from M to itself.

Equation (3.13) just comes by applying Lemma 3.19 with $M = T_pM$, $A = dh(p)$, $e_1 = \partial_1 m$ and $e_2 = \partial_2 m$.

We now give the divergence theorem on a surface, which is a direct generalization of the one we saw on $\mathbb{R}^2$ (Theorem 1.16):

Theorem 3.20 *Let S be an oriented regular surface, and h a smooth vector field on S. Then, if $\Sigma \subset S$ is a bounded subdomain of S with a regular boundary, we have*

$$\int_{\partial\Sigma} h^T N_{\partial\Sigma}\, d\sigma_{\partial\Sigma} = -\int_{\Sigma} \operatorname{div}_S(h)\, d\sigma_S,$$

where the first integral is a line integral over the curve $\partial\Sigma$, and $N_{\partial\Sigma}$ is the inward normal to Σ (normal to $\partial\Sigma$ and tangent to S).

The proof (which we skip) is an application of Green's formula in $\mathbb{R}^2$ combined with a decomposition in local coordinates.

In addition to the divergence, one can define the gradient operator on a surface S, which applies to scalar-valued functions.

Definition 3.21 Let $f : S \to \mathbb{R}$ be C^1. The gradient of f at $p \in S$ is denoted $\nabla_S f(p)$ and defined by $\nabla_S f(p) \in T_p S$ and

$$\forall \xi \in T_p S, \quad \left\langle \nabla_S f(p)\,,\, \xi \right\rangle_p = df(p)\,\xi. \tag{3.15}$$

Note that, even if they are using the same symbol ∇_S, the covariant derivative introduced in (3.10) and the gradient in (3.15) are similar, but different notions, since the former applies to vector fields on S and the latter to scalar functions. Their similarity (and some justification for the notation conflict) is supported by the following observation: if f is the restriction to S of a differentiable function $\hat{f}$ defined on $\mathbb{R}^3$, then (3.15) implies that $\nabla_S f$ is the orthogonal projection of $\nabla \hat{f}$ (the usual gradient in $\mathbb{R}^3$) on the tangent plane to S, namely

$$\nabla_S f(p) = \pi_{T_p S} \nabla \hat{f}(p) = \nabla \hat{f}(p) - (N(p)^T \nabla \hat{f}(p)) N(p). \tag{3.16}$$

In a chart $(u, v) \mapsto m(u, v)$, we have

$$\nabla_S f = \frac{G\partial_1 f - F\partial_2 f}{EG - F^2}\partial_1 m + \frac{E\partial_2 f - F\partial_1 f}{EG - F^2}\partial_2 m. \tag{3.17}$$

The usual formula, $\mathrm{div}(fh) = \nabla f^T h + f\,\mathrm{div} h$, extends to surfaces with

$$\mathrm{div}_S(fh) = \nabla_S f^T h + f\,\mathrm{div}_S h \tag{3.18}$$

for a scalar function f and a vector field h on S.

The generalization of the Laplacian on $\mathbb{R}^2$ is the Laplace–Beltrami operator on S. It is defined as follows:

Definition 3.22 The Laplace–Beltrami operator on a regular surface S associates to a scalar function f on S the scalar function $\Delta_S f$ defined by

$$\Delta_S f = \mathrm{div}_S \nabla_S f. \tag{3.19}$$

The Laplace–Beltrami operator in a chart is therefore given by the combination of (3.17) and (3.12), which yields a formula notably more complex than the ordinary Laplacian.

Theorem 3.20 relates surface integrals to linear integrals over the surface. Surface integrals can also be related to three-dimensional integrals, if the surface is closed, via the three-dimensional divergence theorem.

Theorem 3.23 *Let Ω be a bounded domain in $\mathbb{R}^3$ and assume that $S = \partial\Omega$ is a regular surface. If v is a C^1 vector field on $\mathbb{R}^3$, we have*

$$\int_\Omega \operatorname{div} v \, dx = -\int_S v^T(m) N(m) \, d\sigma_S(m), \tag{3.20}$$

where $N(m)$ is the inward normal to S at m.

From this theorem, we can derive an expression of the volume of Ω as an integral over its boundary, namely (taking $v(x, y, z) = (x, y, z)$ in (3.20))

$$\text{volume}(\Omega) = -\frac{1}{3}\int_S Om^T N(m) \, d\sigma_S(m). \tag{3.21}$$

3.8 Curvature and the Second Fundamental Form

Let S be a C^2 orientable regular surface, and N be its unit normal. The function N can be seen as a map defined on S with values in $\mathbb{R}^3$ (in fact in the unit sphere S^2 since $|N| = 1$), which is called the Gauss map. It therefore has a differential, dN. For any $p \in S$, $dN(p)$ is a linear map from T_pS to $\mathbb{R}^3$. The fact that $|N|^2 = 1$ implies that $(dN(p)h)^T N(p) = 0$ for all $h \in T_pS$ so that the range of $dN(p)$ is orthogonal to $N(p)$ and therefore coincides with T_pS. We can therefore consider $dN(p)$ as an endomorphism (a linear map from a vector space into itself)

$$dN(p) : T_pS \to T_pS.$$

This endomorphism (also called the *shape operator*) is essential for describing the curvature of the surface, which measures how the surface bends in a neighborhood of a point p. It has the interesting property of being symmetric:

Proposition 3.24 *Let S be a regular surface and $p \in S$: for any $h, k \in T_pS$, we have*

$$\langle dN(p)h \, , \, k \rangle_p = \langle h \, , \, dN(p)k \rangle_p.$$

Proof It suffices to show this for a basis of T_pS. Let us take the one provided by a local parametrization around p: $h = \partial_1 m$ and $k = \partial_2 m$. Let $N_m = N \circ m$ be the expression of N as a function of the parameters, so that

$$dN(p)(\alpha\partial_1 m + \beta\partial_2 m) = \alpha\partial_1 N_m + \beta\partial_2 N_m.$$

In particular, $dN(p)\partial_1 m = \partial_1 N_m$ and $dN(p)\partial_2 m = \partial_2 N_m$, and what we need to show is

$$\langle \partial_1 N_m \, , \, \partial_2 m \rangle_p = \langle \partial_1 m \, , \, \partial_2 N_m \rangle_p.$$

But, from $(\partial_1 m)^T N_m = 0$, we get $\langle \partial_1 m\ ,\ \partial_2 N_m \rangle_p = (\partial_1 m)^T \partial_2 N_m = -(\partial_2 \partial_1 m)^T N_m$. Similarly, $\langle \partial_2 m\ ,\ \partial_1 N_m \rangle_p = -(\partial_1 \partial_2 m)^T N_m$. Since partial derivatives commute, the two quantities are equal, yielding the required identity. □

Let γ be a curve on S, and assume that γ is parametrized by arc length. Let $T^{(\gamma)}$ be the unit tangent of γ, $\kappa^{(\gamma)}$ its curvature and $N^{(\gamma)}$ its unit normal, such that $\dot{T}^{(\gamma)} = \kappa^{(\gamma)} N^{(\gamma)}$. The normal $N^{(\gamma)}$ does not coincide with N in general, and we define the normal curvature of γ by the (algebraic) normal part of $\dot{T}^{(\gamma)}$ to the surface S. The interesting point is that it only depends on γ via $T^{(\gamma)}$.

Definition 3.25 The normal curvature at p of an arc length parametrized curve γ on a regular surface S is $\kappa_N^{(\gamma)}(s) = (\dot{T}^{(\gamma)}(s))^T N(\gamma(s))$, where $T^{(\gamma)} = \dot{\gamma}$.

The fact that the normal curvature only depends on $T^{(\gamma)}$ can be proved as follows: let γ be a curve on S such that $\dot{\gamma}(0) = T^{(\gamma)}$. For all s, we have $(T^{(\gamma)})^T N = 0$ since $T^{(\gamma)}$ is tangent to S. Computing the derivative with respect to arc length and applying the chain rule yields

$$(\dot{T}^{(\gamma)})^T N \circ \gamma + (T^{(\gamma)})^T dN(\gamma) T^{(\gamma)} = 0$$

so that

$$\kappa_N^{(\gamma)} = -(T^{(\gamma)})^T dN(\gamma) T^{(\gamma)}. \tag{3.22}$$

One also defines the *geodesic curvature* of γ at s_0 by the curvature (at s_0) of the projection of γ on the tangent plane to S at $\gamma(s_0)$, which is

$$\bar{\gamma}(s) = \gamma(s) - (\gamma(s) - \gamma(s_0))^T N(s_0)\, N(s_0).$$

Computing first and second derivatives in s and computing them at $s = s_0$ yields $\dot{\bar{\gamma}}(s_0) = \dot{\gamma}(s_0)$ and

$$\ddot{\bar{\gamma}}(s_0) = \ddot{\gamma}(s_0) - \kappa_N^{(\gamma)}(s_0) N(s_0).$$

Denoting the geodesic curvature by $\kappa_g^{(\gamma)}(s_0)$, we find (using the definition of the (signed) curvature for plane curves in the oriented tangent plane) that

$$\kappa_g^{(\gamma)} = \det(\dot{\gamma}, \ddot{\gamma}, N) = \ddot{\gamma}^T (N \times \dot{\gamma}),$$

where $N \times \dot{\gamma}$ is the unit normal to γ that belongs to $T_\gamma M$ and complements $\dot{\gamma}$ in a positively oriented basis of the tangent plane. Writing $\ddot{\gamma} = (\ddot{\gamma}^T (N \times \dot{\gamma}))(N \times \dot{\gamma}) + (\ddot{\gamma}^T N)$, one also gets the identity

$$(\kappa_g^{(\gamma)})^2 + (\kappa_N^{(\gamma)})^2 = (\kappa^{(\gamma)})^2,$$

the squared curvature of γ.

This expression in Eq. (3.22) involves another important quantity on S, its second fundamental form.

Definition 3.26 Let S be a regular surface and $p \in S$. The second fundamental form at p is the quadratic form defined on T_pS by

$$II_p(h) = -\langle h\,,\, dN(p)h\rangle_p.$$

In particular, we have the expression of the normal curvature of an arc length parametrized curve γ:

$$\kappa_N^{(\gamma)} = II_\gamma(\dot{\gamma}).$$

Because $dN(p)$ is symmetric, it can be diagonalized in an orthonormal basis of T_pS: let (e_1, e_2) be such a basis, with corresponding eigenvalues $-\kappa_1$ and $-\kappa_2$ such that $\kappa_1 \geq \kappa_2$. The numbers κ_1 and κ_2 are called the *principal curvatures* of the surface at p. The reason for this terminology is that any unit vector in T_pS can be written, for some θ, in the form $h = \cos\theta e_1 + \sin\theta e_2$ and

$$II_p(h) = -\langle h\,,\, dN(p)h\rangle = \kappa_1\cos^2\theta + \kappa_2\sin^2\theta.$$

This implies that $\kappa_2 \leq II_p(h) \leq \kappa_1$, the lower bound being attained for $h = e_2$ and the upper bound for $h = e_1$: κ_1 and κ_2, respectively, are the maximum and minimum normal curvatures of curves passing through p.

Definition 3.27 If κ_1 and κ_2 are the principal curvatures of a surface S at $p \in S$, one defines the mean curvature at p by $H(p) = (\kappa_1 + \kappa_2)/2$, and the Gauss curvature by $K(p) = \kappa_1\kappa_2$. They respectively coincide with the trace of $-dN(p)/2$ and the determinant of $dN(p)$.

From this definition, we can also write

$$2H = -\mathrm{div}'_S(N) \tag{3.23}$$

and rewrite (3.9) as (for $h = h_T + \mu N$)

$$\mathrm{div}'_S(h) = \mathrm{div}_S(h_T) - 2\mu H. \tag{3.24}$$

3.9 Curvature in Local Coordinates

In this section, we give the expression of the curvature in local coordinates, as functions of the coefficients of the first and second fundamental forms. Recall the notation (3.6) for the first fundamental form and a local parametrization m. We introduce a similar notation for the second form, letting

$$II_p(\alpha\partial_1 m + \beta\partial_2 m) = \alpha^2 e + 2\alpha\beta f + \beta^2 g$$

and

$$e = -\partial_1 m^T \partial_1 N = \partial_1^2 m^T N,\ f = -\partial_1 m^T \partial_2 N = \partial_1 \partial_2 m^T N,$$
$$g = -\partial_2 m^T \partial_2 N = \partial_2^2 m^T N. \tag{3.25}$$

Let $dN = \begin{pmatrix} a & c \\ b & d \end{pmatrix}$ in the basis $(\partial_1 m, \partial_2 m)$ (the matrix is not necessarily symmetric since the basis is not assumed to be orthonormal). We find:

$$\begin{aligned} -e &= \partial_1 m^T dN \partial_1 m = aE + bF \\ -f &= \partial_2 m^T dN \partial_1 m = aF + bG \\ -f &= \partial_1 m^T dN \partial_2 m = cE + dF \\ -g &= \partial_2 m^T dN \partial_2 m = cF + dG \end{aligned}$$

which yields, in matrix form: $-\begin{pmatrix} e & f \\ f & g \end{pmatrix} = \begin{pmatrix} E & F \\ F & G \end{pmatrix} \begin{pmatrix} a & c \\ b & d \end{pmatrix}$. This implies that, in the basis $(\partial_1 m, \partial_2 m)$, dN is given by the matrix

$$-\begin{pmatrix} E & F \\ F & G \end{pmatrix}^{-1} \begin{pmatrix} e & f \\ f & g \end{pmatrix}.$$

From this, it can be deduced that

$$K = \frac{eg - f^2}{EG - F^2}$$

because it is just the ratio of the determinants. Also, after computation, one finds

$$H = \frac{eG - 2fF + gE}{2(EG - F^2)}.$$

The principal curvatures are then given by $\kappa = H \pm \sqrt{H^2 - K}$.

3.10 Implicit Surfaces

Assume that a surface is defined implicitly by

$$S = \left\{ p \in \mathbb{R}^3,\ f(p) = 0 \right\},$$

where f is a C^2 function from $\mathbb{R}^3$ to $\mathbb{R}$ with $\nabla f \neq 0$ on S. We have already noticed that the tangent plane to S is orthogonal to ∇f, and therefore $N(p) = -\nabla f(p)/|\nabla f(p)|$ is a smooth unit normal to S which therefore is orientable (and we take the orientation provided by this choice of N).

The interesting feature in this representation is that, since f is defined on $\mathbb{R}^3$, the function N can be extended to $\mathbb{R}^3$ (denote the extension by $\hat{N}$) so that $dN(p)$

is simply the restriction to T_pS of $d\hat{N}(p)$, In particular, the trace of $dN(p)$ is, by definition, $\langle e_1\,,\, dN(p)e_1\rangle_p + \langle e_2\,,\, dN(p)e_2\rangle_p$ for an arbitrary orthonormal basis of T_pS. It therefore suffices to add $(d\hat{N}N)^T N$ to obtain the trace of $d\hat{N}$, but this added quantity vanishes because $|\hat{N}|^2 = 1$ implies that $d\hat{N}N$ is perpendicular to N. Thus, we have, for the mean curvature:

$$H = -\text{trace}(d\hat{N})/2 = \frac{1}{2}\text{div}\frac{\nabla f}{|\nabla f|}. \tag{3.26}$$

(This is the usual divergence on $\mathbb{R}^3$, not to be confused with the S-divergence in Definition 3.17.)

Let P_N be the projection on $\hat{N}^\perp$: $P_N = \text{Id}_{\mathbb{R}^3} - \hat{N}\hat{N}^T$. The Gauss curvature can be computed after diagonalizing the matrix $P_N d\hat{N} P_N = d\hat{N} P_N$, which is symmetric and coincides with dN on T_pS. Using $\hat{N} = -\nabla f/|\nabla f|$, we get

$$\begin{aligned}(d\hat{N}P_N h)^T P_N k &= -\frac{1}{|\nabla f|}(d^2 f P_N h)^T P_N k + \frac{1}{|\nabla f|}\left((d^2 f P_N h)^T \hat{N}\right)\left((P_N k)^T N\right)\\ &= -\frac{1}{|\nabla f|}(P_N h)^T d^2 f P_N k,\end{aligned}$$

which is symmetric in h and k. The matrix $P_N d^2 f P_N/|\nabla f|$ has one vanishing eigenvalue since $P_N N = 0$, and the other two are the principal curvatures of S. Their product provides the Gauss curvature.

The Delta-Function Trick

When a surface or a curve is defined implicitly, integrals over its interior can be described in a straightforward way using the Heaviside function. Assume that S is the set $f(p) = 0$ for some smooth function f, and let Ω be its interior, defined by $f < 0$. Introduce the Heaviside function H_0 defined on $\mathbb{R}$ by $H_0(x) = 1$ if $x \geq 0$ and $H_0(x) = 0$ otherwise. Then, clearly, for any function V on $\mathbb{R}^3$, we have

$$\int_\Omega V(x)dx = \int_{\mathbb{R}^3}(1 - H_0(f(x)))V(x)dx. \tag{3.27}$$

Contour or surface integrals can be defined via a level-set representation, albeit requiring passing to a limit. For this, we need to replace H_0 by a smooth approximation denoted H_ε, which must be an increasing function that tends to H_0 as ε tends to 0. A possible example is (cf. [227, 316]) $H_\varepsilon(x) = 0$ for $x < -\varepsilon$, $H_\varepsilon(x) = 1$ for $x > \varepsilon$ and, on $[-\varepsilon, \varepsilon]$:

$$H_\varepsilon(x) = \frac{1}{2}\left(1 + \frac{x}{\varepsilon} + \frac{1}{\pi}\sin\left(\frac{\pi x}{\varepsilon}\right)\right). \tag{3.28}$$

Alternatively [58], one can take, for all $x \in \mathbb{R}$:

$$H_\varepsilon(x) = \frac{1}{2}\left(1 + \frac{2}{\pi}\arctan\left(\frac{x}{\varepsilon}\right)\right). \tag{3.29}$$

This choice being made, let δ_ε denote the derivative of H_ε. The function δ_ε can be considered as a smooth approximation of the Dirac function δ_0, in the sense that, for any bounded function u on $\mathbb{R}$ which is continuous at $t = 0$, one has

$$\lim_{\varepsilon\to 0}\int_{\mathbb{R}} \delta_\varepsilon(t)u(t)dt = u(0). \tag{3.30}$$

We leave the easy proof to the reader (simply divide the integral over domains around 0 or away from 0).

We now describe how surface integrals over implicitly defined surfaces can be approximated using δ_ε.

Proposition 3.28 *Let $f : \mathbb{R}^3 \to \mathbb{R}$ be a C^2 function with $\nabla f \neq 0$ if $f = 0$, and such that the implicit surface $S = f^{-1}(\varepsilon)$ is bounded in a neighborhood of 0. Then, if $V : \mathbb{R}^3 \to \mathbb{R}$ is continuous, we have*

$$\lim_{\varepsilon\to 0}\int_{\mathbb{R}^3} \delta_\varepsilon \circ f(x)\, V(x)\, |\nabla f(x)|\, dx = \int_S V(m)\, d\sigma_S(m). \tag{3.31}$$

The same proposition holds for curves, with $f : \mathbb{R}^2 \to \mathbb{R}$ and the surface integral replaced by the integral along the curve.

Proof Let's consider the surface case (the case of curves is similar and simpler). We also assume that δ_ε is supported in $[-\varepsilon, \varepsilon]$, like for (3.28) (the general case requiring only minor modifications). Consider a local chart $(u, v) \mapsto m(u, v)$ on $S = f^{-1}(0)$. Consider the equation

$$f(m(u, v) + tN(u, v)) = y,$$

which we want to solve for t as a function of (u, v, y) in a neighborhood of some $u = u_0$, $v = v_0$ and $y = 0$. From the implicit function theorem, this is possible, because

$$\partial_t f(m + tN) = \nabla f^T N = -|\nabla f|,$$

which is not zero by assumption. Using the compactness of S, we can find a finite number of points $p_0 = m(u_0, v_0)$ and domains around $(u_0, v_0, 0) \in \mathbb{R}^3$ over which a function $t(m(u, v), y)$ such that $f(m + tN) = y$ is well-defined and such that the union of these domains forms an open set in $\mathbb{R}^3$ that contains S, and more generally contains the set $|f(p)| < y_0$ for y_0 small enough.

Taking $\varepsilon < y_0$, we can write

$$\int_{\mathbb{R}^d} \delta_\varepsilon \circ f(x)\, V(x)\, |\nabla f(x)|\, dx = \int_{|f|<y_0} \delta_\varepsilon \circ f(x)\, V(x)\, |\nabla f(x)|\, dx.$$

(Not assuming δ_ε to be compactly supported would add a small error to this identity, which is easily shown to be negligible when $\varepsilon \to 0$.)

We can decompose the integral over a partition of unity, which reduces the problem to the situation in which V is supported by one of the domains above. Working under this assumption, we make the change of variables $x(u, v, y) = m(u, v) + t(m(u, v), y)N(u, v)$ in this domain and let $J(u, v, t)$ be the associated Jacobian determinant, so that

$$\int_{|f|<y_0} \delta_\varepsilon \circ f(x) V(x) \, |\nabla f(x)| \, dx = \int_{|y|<y_0} \delta_\varepsilon(y) \, V(x(u, v, y)) \, |\nabla f(x(u, v, y))| \, J(u, v, y) \, du \, dv \, dy.$$

Our assumptions ensure that the integral

$$u(y) = \int V(x(u, v, y)) \, |\nabla f(x(u, v, y))| \, J(u, v, y) \, du \, dv$$

is continuous in y so that,

$$\lim_{\varepsilon \to 0} \int_{\mathbb{R}^d} \delta_\varepsilon \circ f(x) \, V(x) \, |\nabla f(x)| \, dx = u(0).$$

Now,

$$J(u, v, 0) = |\det(\partial_1 m, \partial_2 m, \partial_3 t N)| = |\partial_1 m \times \partial_2 m| / |\nabla f(m)|$$

because $y = f(m + tN)$ implies $1 = \partial_3 t \, \nabla f^T N = -\partial_3 t |\nabla f|$. This implies that the $|\nabla f|$ terms cancel in the expression of $u(0)$, which is equal to

$$u(0) = \int V(m(u, v)) \, |\partial_1 m \times \partial_2 m| \, du \, dv = \int_S V \, d\sigma_S,$$

which concludes the proof. □

The theorem is particularly important for numerical computations, because it replaces computations over a surface with computations over a grid that contains the surface.

The left-hand side of (3.31) is often written using the symbolic notation

$$\int_{\mathbb{R}^2} \delta_0 \circ f(x) \, V(x) \, |\nabla f(x)| \, dx.$$

The assumption that V is continuous is important (of course, we only need continuity near $f^{-1}(0)$). Take the following simple example with curves; let $f(u, v) = u^2 + v^2 - 1$, so that $f^{-1}(0) = S^1$, the unit circle and let $V(u, v) = 1$ if $u^2 + v^2 \leq 1$ and 0 otherwise. Then

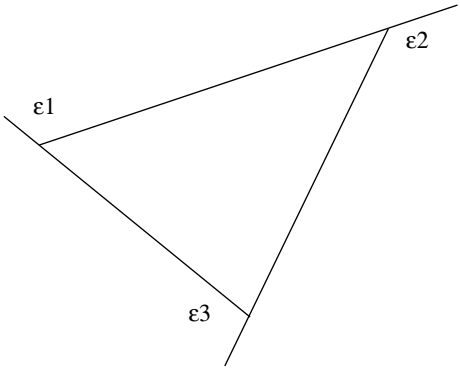

Fig. 3.6 The Gauss–Bonnet theorem in $\mathbb{R}^2$ reduces to the well-known property that the sum of consecutive angles in a polygon is 2π

$$\int_{S^1} V dl = 2\pi$$

but

$$\lim_{\varepsilon \to 0} \int_{\mathbb{R}^2} \delta_\varepsilon \circ f(x) V(x) |\nabla f(x)| dx = \pi$$

(both integrals being easily computed in radial coordinates).

3.11 The Gauss–Bonnet Theorem

The average of the Gauss curvature over a domain with piecewise geodesic boundary is provided by the Gauss–Bonnet formula [86]:

Theorem 3.29 *Let S be a regular surface and A be a domain on M such that ∂A is the union of N geodesics $\gamma^{(1)}, \ldots, \gamma^{(N)}$. Let $\varepsilon_i, i = 1, \ldots N$ be the sequence of consecutive angles between the curves at their intersection. Then*

$$\int_A K d\sigma = 2\pi - \sum_{i=1}^{N} \varepsilon_i. \tag{3.32}$$

For example, when $N = 3$ (∂A is a "geodesic triangle"), we obtain the fact that the sum of the angles of a triangle is 2π minus the integral of the Gauss curvature over its interior. This is consistent with the sum being 2π in the plane, which has zero Gauss curvature (Fig. 3.6).