DIFFERENTIAL GEOMETRY:
A First Course in
Curves and Surfaces

Preliminary Version

Spring, 2016

Theodore Shifrin
University of Georgia

Dedicated to the memory of **Shiing-Shen Chern**,
my adviser and friend

CONTENTS

1. CURVES 1

1. Examples, Arclength Parametrization 1
2. Local Theory: Frenet Frame 10
3. Some Global Results 23

2. SURFACES: LOCAL THEORY 35

1. Parametrized Surfaces and the First Fundamental Form 35
2. The Gauss Map and the Second Fundamental Form 44
3. The Codazzi and Gauss Equations and the Fundamental Theorem of Surface Theory 57
4. Covariant Differentiation, Parallel Translation, and Geodesics 66

3. SURFACES: FURTHER TOPICS 79

1. Holonomy and the Gauss-Bonnet Theorem 79
2. An Introduction to Hyperbolic Geometry 91
3. Surface Theory with Differential Forms 101
4. Calculus of Variations and Surfaces of Constant Mean Curvature 107

Appendix.
REVIEW OF LINEAR ALGEBRA AND CALCULUS . . . 114

1. Linear Algebra Review 114
2. Calculus Review 116
3. Differential Equations 118

SOLUTIONS TO SELECTED EXERCISES 121

INDEX 124

Problems to which answers or hints are given at the back of the book are marked with an asterisk (*). Fundamental exercises that are particularly important (and to which reference is made later) are marked with a sharp ($^\sharp$).

August, 2015

CHAPTER 1

Curves

1. Examples, Arclength Parametrization

We say a vector function $\mathbf{f}: (a,b) \to \mathbb{R}^3$ is $\mathcal{C}^k$ $(k = 0, 1, 2, \dots)$ if $\mathbf{f}$ and its first k derivatives, $\mathbf{f}'$, $\mathbf{f}''$, $\dots$, $\mathbf{f}^{(k)}$, exist and are all continuous. We say $\mathbf{f}$ is *smooth* if $\mathbf{f}$ is $\mathcal{C}^k$ for every positive integer k. A *parametrized curve* is a $\mathcal{C}^3$ (or smooth) map $\boldsymbol{\alpha}: I \to \mathbb{R}^3$ for some interval $I = (a,b)$ or $[a,b]$ in $\mathbb{R}$ (possibly infinite). We say $\boldsymbol{\alpha}$ is *regular* if $\boldsymbol{\alpha}'(t) \neq \mathbf{0}$ for all $t \in I$.

We can imagine a particle moving along the path $\boldsymbol{\alpha}$, with its position at time t given by $\boldsymbol{\alpha}(t)$. As we learned in vector calculus,

$$\boldsymbol{\alpha}'(t) = \frac{d\boldsymbol{\alpha}}{dt} = \lim_{h\to 0} \frac{\boldsymbol{\alpha}(t+h) - \boldsymbol{\alpha}(t)}{h}$$

is the *velocity* of the particle at time t. The velocity vector $\boldsymbol{\alpha}'(t)$ is tangent to the curve at $\boldsymbol{\alpha}(t)$ and its length, $\|\boldsymbol{\alpha}'(t)\|$, is the speed of the particle.

Example 1. We begin with some standard examples.

(a) Familiar from linear algebra and vector calculus is a parametrized line: Given points P and Q in $\mathbb{R}^3$, we let $\mathbf{v} = \overrightarrow{PQ} = Q - P$ and set $\boldsymbol{\alpha}(t) = P + t\mathbf{v}$, $t \in \mathbb{R}$. Note that $\boldsymbol{\alpha}(0) = P$, $\boldsymbol{\alpha}(1) = Q$, and for $0 \le t \le 1$, $\boldsymbol{\alpha}(t)$ is on the line segment $\overline{PQ}$. We ask the reader to check in Exercise 8 that of all paths from P to Q, the "straight line path" $\boldsymbol{\alpha}$ gives the shortest. This is typical of problems we shall consider in the future.

(b) Essentially by the very definition of the trigonometric functions cos and sin, we obtain a very natural parametrization of a circle of radius a, as pictured in Figure 1.1(a):

$$\boldsymbol{\alpha}(t) = a\big(\cos t, \sin t\big) = \big(a\cos t, a\sin t\big), \quad 0 \le t \le 2\pi.$$

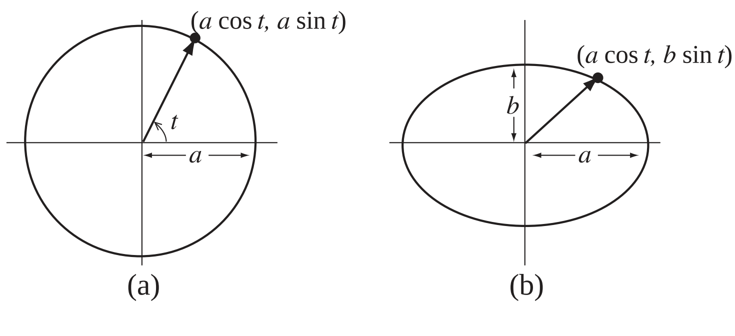


FIGURE 1.1

(c) Now, if $a, b > 0$ and we apply the linear map

$$T\colon \mathbb{R}^2 \to \mathbb{R}^2, \quad T(x, y) = (ax, by),$$

we see that the unit circle $x^2+y^2 = 1$ maps to the ellipse $x^2/a^2+y^2/b^2 = 1$. Since $T(\cos t, \sin t) = (a \cos t, b \sin t)$, the latter gives a natural parametrization of the ellipse, as shown in Figure 1.1(b).

(d) Consider the two cubic curves in $\mathbb{R}^2$ illustrated in Figure 1.2. On the left is the *cuspidal cubic*

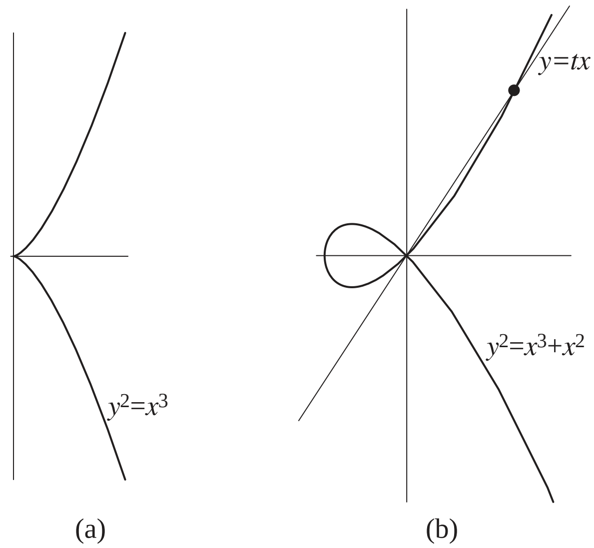


FIGURE 1.2

$y^2 = x^3$, and on the right is the *nodal cubic* $y^2 = x^3+x^2$. These can be parametrized, respectively, by the functions

$$\boldsymbol{\alpha}(t) = (t^2, t^3) \qquad \text{and} \qquad \boldsymbol{\alpha}(t) = (t^2 - 1, t(t^2 - 1)).$$

(In the latter case, as the figure suggests, we see that the line $y = tx$ intersects the curve when $(tx)^2 = x^2(x + 1)$, so $x = 0$ or $x = t^2 - 1$.)

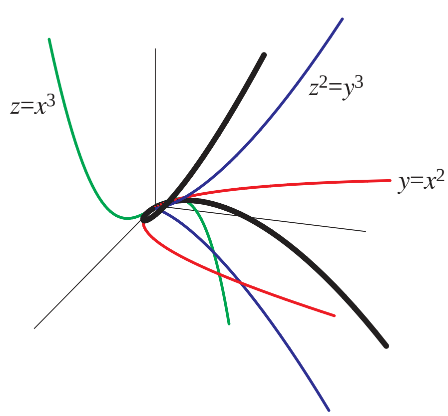


FIGURE 1.3

(e) Now consider the *twisted cubic* in $\mathbb{R}^3$, illustrated in Figure 1.3, given by

$$\boldsymbol{\alpha}(t) = (t, t^2, t^3), \quad t \in \mathbb{R}.$$

Its projections in the xy-, xz-, and yz-coordinate planes are, respectively, $y = x^2$, $z = x^3$, and $z^2 = y^3$ (the cuspidal cubic).

(f) Our next example is a classic called the *cycloid*: It is the trajectory of a dot on a rolling wheel (circle). Consider the illustration in Figure 1.4. Assuming the wheel rolls without slipping, the

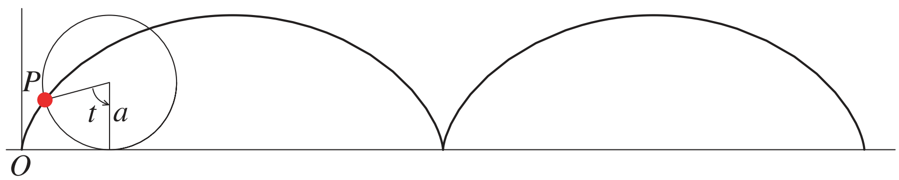


FIGURE 1.4

distance it travels along the ground is equal to the length of the circular arc subtended by the angle through which it has turned. That is, if the radius of the circle is a and it has turned through angle t, then the point of contact with the x-axis, Q, is at units to the right. The vector from the origin to

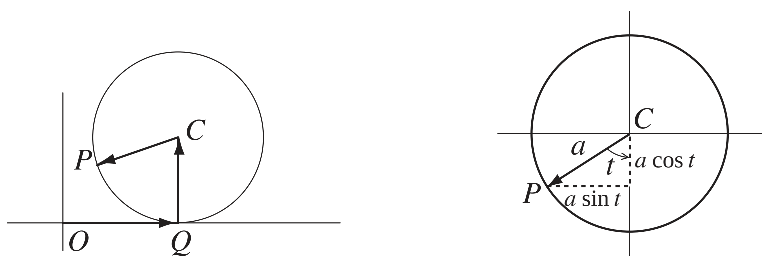


FIGURE 1.5

the point P can be expressed as the sum of the three vectors $\overrightarrow{OQ}$, $\overrightarrow{QC}$, and $\overrightarrow{CP}$ (see Figure 1.5):

$$\begin{aligned}\overrightarrow{OP} &= \overrightarrow{OQ} + \overrightarrow{QC} + \overrightarrow{CP} \\ &= (at, 0) + (0, a) + (-a\sin t, -a\cos t),\end{aligned}$$

and hence the function

$$\boldsymbol{\alpha}(t) = (at - a\sin t, a - a\cos t) = a(t - \sin t, 1 - \cos t), \quad t \in \mathbb{R}$$

gives a parametrization of the cycloid.

(g) A (circular) *helix* is the screw-like path of a bug as it walks uphill on a right circular cylinder at a constant slope or pitch. If the cylinder has radius a and the slope is b/a, we can imagine drawing a line of that slope on a piece of paper $2\pi a$ units long, and then rolling the paper up into a cylinder. The line gives one revolution of the helix, as we can see in Figure 1.6. If we take the axis of the cylinder to be vertical, the projection of the helix in the horizontal plane is a circle of radius a, and so we obtain the parametrization $\boldsymbol{\alpha}(t) = (a\cos t, a\sin t, bt)$.

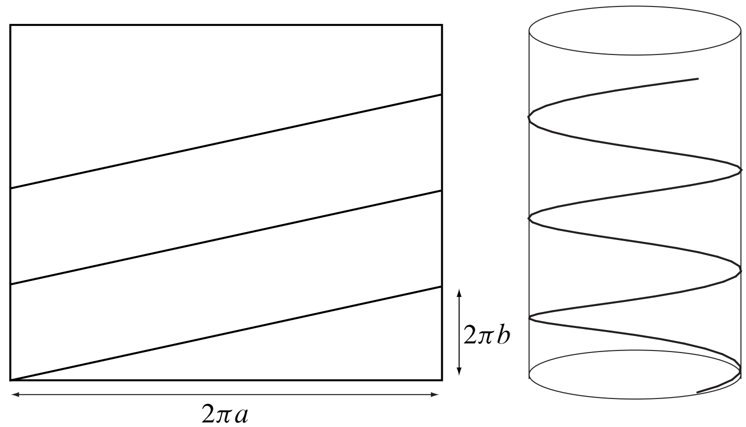


FIGURE 1.6

Brief review of hyperbolic trigonometric functions. Just as the circle $x^2+y^2=1$ is parametrized by $(\cos\theta, \sin\theta)$, the portion of the hyperbola $x^2-y^2=1$ lying to the right of the y-axis, as shown in Figure 1.7, is parametrized by $(\cosh t, \sinh t)$, where

$$\cosh t = \frac{e^t+e^{-t}}{2} \qquad \text{and} \qquad \sinh t = \frac{e^t-e^{-t}}{2}.$$

By analogy with circular trigonometry, we set $\tanh t = \dfrac{\sinh t}{\cosh t}$ and $\operatorname{sech} t = \dfrac{1}{\cosh t}$. The following

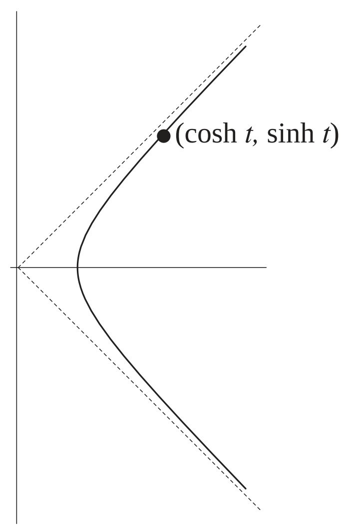


FIGURE 1.7

formulas are easy to check:

$$\cosh^2 t - \sinh^2 t = 1, \qquad \tanh^2 t + \operatorname{sech}^2 t = 1$$

$$\sinh'(t) = \cosh t, \qquad \cosh'(t) = \sinh t, \qquad \tanh'(t) = \operatorname{sech}^2 t, \qquad \operatorname{sech}'(t) = -\tanh t \ \operatorname{sech} t.$$

(h) When a uniform and flexible chain hangs from two pegs, its weight is uniformly distributed along its length. The shape it takes is called a *catenary*.[1] As we ask the reader to check in Exercise 9, the catenary is the graph of $f(x) = C\cosh(x/C)$, for any constant $C > 0$. This curve will appear

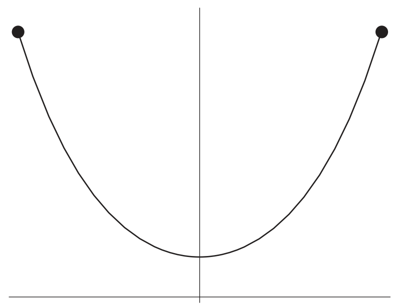


FIGURE 1.8

numerous times in this course. ∇

Example 2. One of the more interesting curves that arise "in nature" is the *tractrix*.[2] The traditional story is this: A dog is at the end of a 1-unit leash and buries a bone at $(0,1)$ as his owner begins to walk down the x-axis, starting at the origin. The dog tries to get back to the bone, so he always pulls the leash taut as he is dragged along the tractrix by his owner. His pulling the leash taut means that the leash will be tangent to the curve. When the master is at $(t,0)$, let the dog's position be $(x(t), y(t))$, and let the leash

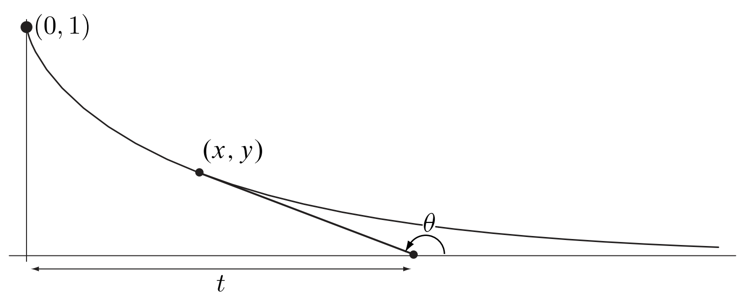


FIGURE 1.9

make angle $\theta(t)$ with the positive x-axis. Then we have $x(t) = t + \cos\theta(t)$, $y(t) = \sin\theta(t)$, so

$$\tan\theta(t) = \frac{dy}{dx} = \frac{y'(t)}{x'(t)} = \frac{\cos\theta(t)\theta'(t)}{1 - \sin\theta(t)\theta'(t)}.$$

Therefore, $\theta'(t) = \sin\theta(t)$. Separating variables and integrating, we have $\int d\theta/\sin\theta = \int dt$, and so $t = -\ln(\csc\theta + \cot\theta) + c$ for some constant c. Since $\theta = \pi/2$ when $t = 0$, we see that $c = 0$. Now, since $\csc\theta + \cot\theta = \dfrac{1+\cos\theta}{\sin\theta} = \dfrac{2\cos^2(\theta/2)}{2\sin(\theta/2)\cos(\theta/2)} = \cot(\theta/2)$, we can rewrite this as $t = \ln\tan(\theta/2)$. Thus, we can parametrize the tractrix by

$$\boldsymbol{\alpha}(\theta) = \big(\cos\theta + \ln\tan(\theta/2), \sin\theta\big), \quad \pi/2 \le \theta < \pi.$$

[1] From the Latin *catēna*, chain.

[2] From the Latin *trahere, tractus*, to pull.

Alternatively, since $\tan(\theta/2) = e^t$, we have

$$\sin\theta = 2\sin(\theta/2)\cos(\theta/2) = \frac{2e^t}{1+e^{2t}} = \frac{2}{e^t+e^{-t}} = \operatorname{sech} t$$
$$\cos\theta = \cos^2(\theta/2) - \sin^2(\theta/2) = \frac{1-e^{2t}}{1+e^{2t}} = \frac{e^{-t}-e^t}{e^t+e^{-t}} = -\tanh t,$$

and so we can parametrize the tractrix instead by

$$\boldsymbol{\beta}(t) = \big(t - \tanh t, \operatorname{sech} t\big), \quad t \geq 0. \qquad \nabla$$

The fundamental concept underlying the geometry of curves is the arclength of a parametrized curve.

Definition. If $\boldsymbol{\alpha}\colon [a,b] \to \mathbb{R}^3$ is a parametrized curve, then for any $a \leq t \leq b$, we define its *arclength* from a to t to be $s(t) = \displaystyle\int_a^t \|\boldsymbol{\alpha}'(u)\|\,du$. That is, the distance a particle travels—the arclength of its trajectory—is the integral of its speed.

An alternative approach is to start with the following

Definition. Let $\boldsymbol{\alpha}\colon [a,b] \to \mathbb{R}^3$ be a (continuous) parametrized curve. Given a partition $\mathcal{P} = \{a = t_0 < t_1 < \cdots < t_k = b\}$ of the interval $[a,b]$, let

$$\ell(\boldsymbol{\alpha}, \mathcal{P}) = \sum_{i=1}^{k} \|\boldsymbol{\alpha}(t_i) - \boldsymbol{\alpha}(t_{i-1})\|.$$

That is, $\ell(\boldsymbol{\alpha}, \mathcal{P})$ is the length of the inscribed polygon with vertices at $\boldsymbol{\alpha}(t_i)$, $i = 0, \ldots, k$, as indicated in

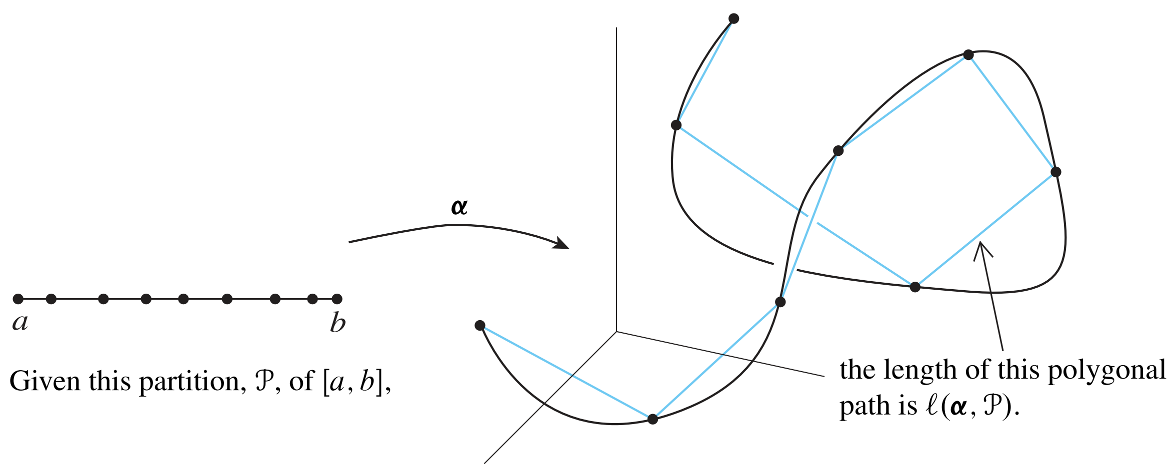


FIGURE 1.10

Figure 1.10. We define the *arclength* of $\boldsymbol{\alpha}$ to be

$$\operatorname{length}(\boldsymbol{\alpha}) = \sup\{\ell(\boldsymbol{\alpha}, \mathcal{P}) : \mathcal{P} \text{ a partition of } [a,b]\},$$

provided the set of polygonal lengths is bounded above.

Now, using this definition, we can *prove* that the distance a particle travels is the integral of its speed. We will need to use the result of Exercise A.2.4.

Proposition 1.1. *Let* $\boldsymbol{\alpha}\colon [a,b] \to \mathbb{R}^3$ *be a piecewise-*$\mathcal{C}^1$ *parametrized curve. Then*

$$\text{length}(\boldsymbol{\alpha}) = \int_a^b \|\boldsymbol{\alpha}'(t)\|\,dt.$$

Proof. For any partition $\mathcal{P}$ of $[a,b]$, we have

$$\ell(\boldsymbol{\alpha},\mathcal{P}) = \sum_{i=1}^k \|\boldsymbol{\alpha}(t_i) - \boldsymbol{\alpha}(t_{i-1})\| = \sum_{i=1}^k \left\| \int_{t_{i-1}}^{t_i} \boldsymbol{\alpha}'(t)\,dt \right\| \le \sum_{i=1}^k \int_{t_{i-1}}^{t_i} \|\boldsymbol{\alpha}'(t)\|\,dt = \int_a^b \|\boldsymbol{\alpha}'(t)\|\,dt,$$

so $\text{length}(\boldsymbol{\alpha}) \le \int_a^b \|\boldsymbol{\alpha}'(t)\|\,dt$. The corresponding inequality holds on any interval.

Now, for $a \le t \le b$, define $s(t)$ to be the arclength of the curve $\boldsymbol{\alpha}$ on the interval $[a,t]$. Then for $h > 0$ we have

$$\frac{\|\boldsymbol{\alpha}(t+h) - \boldsymbol{\alpha}(t)\|}{h} \le \frac{s(t+h) - s(t)}{h} \le \frac{1}{h}\int_t^{t+h} \|\boldsymbol{\alpha}'(u)\|\,du,$$

since $s(t+h) - s(t)$ is the arclength of the curve $\boldsymbol{\alpha}$ on the interval $[t, t+h]$. (See Exercise 8 for the first inequality and the first paragraph for the second.) Now

$$\lim_{h\to 0^+} \frac{\|\boldsymbol{\alpha}(t+h) - \boldsymbol{\alpha}(t)\|}{h} = \|\boldsymbol{\alpha}'(t)\| = \lim_{h\to 0^+} \frac{1}{h}\int_t^{t+h} \|\boldsymbol{\alpha}'(u)\|\,du.$$

Therefore, by the squeeze principle,

$$\lim_{h\to 0^+} \frac{s(t+h) - s(t)}{h} = \|\boldsymbol{\alpha}'(t)\|.$$

A similar argument works for $h < 0$, and we conclude that $s'(t) = \|\boldsymbol{\alpha}'(t)\|$. Therefore,

$$s(t) = \int_a^t \|\boldsymbol{\alpha}'(u)\|\,du, \quad a \le t \le b,$$

and, in particular, $s(b) = \text{length}(\boldsymbol{\alpha}) = \int_a^b \|\boldsymbol{\alpha}'(t)\|\,dt$, as desired. □

If $\|\boldsymbol{\alpha}'(t)\| = 1$ for all $t \in [a,b]$, i.e., $\boldsymbol{\alpha}$ always has speed 1, then $s(t) = t - a$. We say the curve $\boldsymbol{\alpha}$ is *parametrized by arclength* if $s(t) = t$ for all t. In this event, we usually use the parameter $s \in [0, L]$ and write $\boldsymbol{\alpha}(s)$.

Example 3. (a) Let $\boldsymbol{\alpha}(t) = \left(\frac{1}{3}(1+t)^{3/2}, \frac{1}{3}(1-t)^{3/2}, \frac{1}{\sqrt{2}}t\right)$, $t \in (-1,1)$. Then we have $\boldsymbol{\alpha}'(t) = \left(\frac{1}{2}(1+t)^{1/2}, -\frac{1}{2}(1-t)^{1/2}, \frac{1}{\sqrt{2}}\right)$, and $\|\boldsymbol{\alpha}'(t)\| = 1$ for all t. Thus, $\boldsymbol{\alpha}$ always has speed 1.

(b) The standard parametrization of the circle of radius a is $\boldsymbol{\alpha}(t) = (a\cos t, a\sin t)$, $t \in [0, 2\pi]$, so $\boldsymbol{\alpha}'(t) = (-a\sin t, a\cos t)$ and $\|\boldsymbol{\alpha}'(t)\| = a$. It is easy to see from the chain rule that if we reparametrize the curve by $\boldsymbol{\beta}(s) = (a\cos(s/a), a\sin(s/a))$, $s \in [0, 2\pi a]$, then $\boldsymbol{\beta}'(s) = (-\sin(s/a), \cos(s/a))$ and $\|\boldsymbol{\beta}'(s)\| = 1$ for all s. Thus, the curve $\boldsymbol{\beta}$ is parametrized by arclength. ▽

An important observation from a theoretical standpoint is that any regular parametrized curve can be *re*parametrized by arclength. For if $\boldsymbol{\alpha}$ is regular, the arclength function $s(t) = \displaystyle\int_a^t \|\boldsymbol{\alpha}'(u)\| du$ is an increasing differentiable function (since $s'(t) = \|\boldsymbol{\alpha}'(t)\| > 0$ for all t), and therefore has a differentiable inverse function $t = t(s)$. Then we can consider the parametrization

$$\boldsymbol{\beta}(s) = \boldsymbol{\alpha}(t(s)).$$

Note that the chain rule tells us that

$$\boldsymbol{\beta}'(s) = \boldsymbol{\alpha}'(t(s))t'(s) = \boldsymbol{\alpha}'(t(s))/s'(t(s)) = \boldsymbol{\alpha}'(t(s))/\|\boldsymbol{\alpha}'(t(s))\|$$

is everywhere a unit vector; in other words, $\boldsymbol{\beta}$ moves with speed 1.

EXERCISES 1.1

*1. Parametrize the unit circle (less the point $(-1, 0)$) by the length t indicated in Figure 1.11.

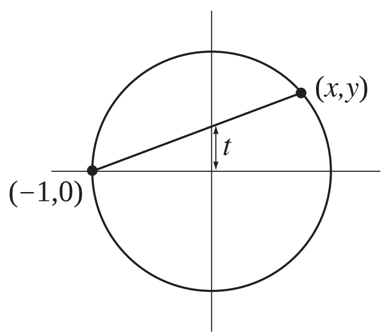


FIGURE 1.11

♯2. Consider the helix $\boldsymbol{\alpha}(t) = (a\cos t, a\sin t, bt)$. Calculate $\boldsymbol{\alpha}'(t)$, $\|\boldsymbol{\alpha}'(t)\|$, and reparametrize $\boldsymbol{\alpha}$ by arclength.

3. Let $\boldsymbol{\alpha}(t) = \left(\frac{1}{\sqrt{3}}\cos t + \frac{1}{\sqrt{2}}\sin t, \frac{1}{\sqrt{3}}\cos t, \frac{1}{\sqrt{3}}\cos t - \frac{1}{\sqrt{2}}\sin t\right)$. Calculate $\boldsymbol{\alpha}'(t)$, $\|\boldsymbol{\alpha}'(t)\|$, and reparametrize $\boldsymbol{\alpha}$ by arclength.

*4. Parametrize the graph $y = f(x)$, $a \le x \le b$, and show that its arclength is given by the traditional formula

$$\text{length} = \int_a^b \sqrt{1 + \left(f'(x)\right)^2}\, dx.$$

5. a. Show that the arclength of the catenary $\boldsymbol{\alpha}(t) = (t, \cosh t)$ for $0 \le t \le b$ is $\sinh b$.
   b. Reparametrize the catenary by arclength. (Hint: Find the inverse of sinh by using the quadratic formula.)

*6. Consider the curve $\boldsymbol{\alpha}(t) = (e^t, e^{-t}, \sqrt{2}t)$. Calculate $\boldsymbol{\alpha}'(t)$, $\|\boldsymbol{\alpha}'(t)\|$, and reparametrize $\boldsymbol{\alpha}$ by arclength, starting at $t = 0$.

7. Find the arclength of the tractrix, given in Example 2, starting at $(0, 1)$ and proceeding to an arbitrary point.

$^\sharp$8. Let $P, Q \in \mathbb{R}^3$ and let $\boldsymbol{\alpha}\colon [a, b] \to \mathbb{R}^3$ be any parametrized curve with $\boldsymbol{\alpha}(a) = P$, $\boldsymbol{\alpha}(b) = Q$. Let $\mathbf{v} = Q - P$. Prove that length$(\boldsymbol{\alpha}) \geq \|\mathbf{v}\|$, so that the line segment from P to Q gives the shortest possible path. (Hint: Consider $\int_a^b \boldsymbol{\alpha}'(t) \cdot \mathbf{v}\, dt$ and use the Cauchy-Schwarz inequality $\mathbf{u} \cdot \mathbf{v} \leq \|\mathbf{u}\|\|\mathbf{v}\|$. Of course, with the alternative definition on p. 6, it's even easier.)

9. Consider a uniform cable with density δ hanging in equilibrium. As shown in Figure 1.12, the tension forces $\mathbf{T}(x + \Delta x)$, $-\mathbf{T}(x)$, and the weight of the piece of cable lying over $[x, x + \Delta x]$ all balance. If the bottom of the cable is at $x = 0$, T_0 is the magnitude of the tension there, and the cable is

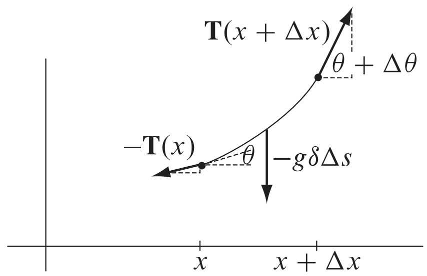


FIGURE 1.12

the graph $y = f(x)$, show that $f''(x) = \dfrac{g\delta}{T_0}\sqrt{1 + f'(x)^2}$. (Remember that $\tan\theta = f'(x)$.) Letting $C = T_0/g\delta$, show that $f(x) = C\cosh(x/C) + c$ for some constant c. (Hint: To integrate $\displaystyle\int \frac{du}{\sqrt{1 + u^2}}$, make the substitution $u = \sinh v$.)

10. As shown in Figure 1.13, Freddy Flintstone wishes to drive his car with square wheels along a strange road. How should you design the road so that his ride is perfectly smooth, i.e., so that the center of his wheel travels in a horizontal line? (Hints: Start with a square with vertices at $(\pm 1, \pm 1)$, with center

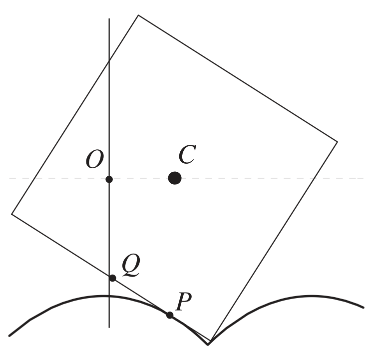


FIGURE 1.13

C at the origin. If $\boldsymbol{\alpha}(s) = (x(s), y(s))$ is an arclength parametrization of the road, starting at $(0, -1)$, consider the vector $\overrightarrow{OC} = \overrightarrow{OP} + \overrightarrow{PQ} + \overrightarrow{QC}$, where $P = \boldsymbol{\alpha}(s)$ is the point of contact and Q is the midpoint of the edge of the square. Use $\overrightarrow{QP} = s\boldsymbol{\alpha}'(s)$ and the fact that $\overrightarrow{QC}$ is a unit vector orthogonal to

$\overrightarrow{QP}$. Express the fact that C moves horizontally to show that $s = -\dfrac{y'(s)}{x'(s)}$; you will need to differentiate unexpectedly. Now use the result of Exercise 4 to find $y = f(x)$. Also see the hint for Exercise 9.)

11. Show that the curve $\boldsymbol{\alpha}(t) = \begin{cases} \big(t, t\sin(\pi/t)\big), & t \neq 0 \\ (0,0), & t = 0 \end{cases}$ has infinite length on $[0,1]$. (Hint: Consider $\ell(\boldsymbol{\alpha}, \mathcal{P}_N)$ with $\mathcal{P}_N = \{0, 1/N, 2/(2N-1), 1/(N-1), \dots, 1/2, 2/3, 1\}$.)

12. Prove that no four distinct points on the twisted cubic (see Example 1(e)) lie on a plane.

13. Consider the "spiral" $\boldsymbol{\alpha}(t) = r(t)(\cos t, \sin t)$, where r is $\mathcal{C}^1$ and $0 \le r(t) \le 1$ for all $t \ge 0$.
    a. Show that if $\boldsymbol{\alpha}$ has finite length on $[0,\infty)$ and r is decreasing, then $r(t) \to 0$ as $t \to \infty$.
    b. Show that if $r(t) = 1/(t+1)$, then $\boldsymbol{\alpha}$ has infinite length on $[0,\infty)$.
    c. If $r(t) = 1/(t+1)^2$, does $\boldsymbol{\alpha}$ have finite length on $[0,\infty)$?
    d. Characterize (in terms of the existence of improper integral(s)) the functions r for which $\boldsymbol{\alpha}$ has finite length on $[0,\infty)$.
    e. Use the result of part d to show that the result of part a holds even without the hypothesis that r be decreasing.

14. (a special case of a recent *American Mathematical Monthly* problem) Suppose $\boldsymbol{\alpha}\colon [a,b] \to \mathbb{R}^2$ is a smooth parametrized plane curve (perhaps not arclength-parametrized). Prove that if the chord length $\|\boldsymbol{\alpha}(s) - \boldsymbol{\alpha}(t)\|$ depends only on $|s-t|$, then $\boldsymbol{\alpha}$ must be a (subset of) a line or a circle. (How many derivatives of $\boldsymbol{\alpha}$ do you need to use?)

2. Local Theory: Frenet Frame

What distinguishes a circle or a helix from a line is their *curvature*, i.e., the tendency of the curve to change direction. We shall now see that we can associate to each smooth ($\mathcal{C}^3$) arclength-parametrized curve $\boldsymbol{\alpha}$ a natural "moving frame" (an orthonormal basis for $\mathbb{R}^3$ chosen at each point on the curve, adapted to the geometry of the curve as much as possible).

We begin with a fact from vector calculus that will appear throughout this course.

Lemma 2.1. *Suppose* $\mathbf{f}, \mathbf{g}\colon (a,b) \to \mathbb{R}^3$ *are differentiable and satisfy* $\mathbf{f}(t) \cdot \mathbf{g}(t) = \text{const}$ *for all* t. *Then* $\mathbf{f}'(t) \cdot \mathbf{g}(t) = -\mathbf{f}(t) \cdot \mathbf{g}'(t)$. *In particular,*

$$\boxed{\|\mathbf{f}(t)\| = \text{const} \quad \text{if and only if} \quad \mathbf{f}(t) \cdot \mathbf{f}'(t) = 0 \quad \text{for all } t.}$$

Proof. Since a function is constant on an interval if and only if its derivative is zero everywhere on that interval, we deduce from the product rule,

$$(\mathbf{f} \cdot \mathbf{g})'(t) = \mathbf{f}'(t) \cdot \mathbf{g}(t) + \mathbf{f}(t) \cdot \mathbf{g}'(t),$$

that if $\mathbf{f} \cdot \mathbf{g}$ is constant, then $\mathbf{f} \cdot \mathbf{g}' = -\mathbf{f}' \cdot \mathbf{g}$. In particular, $\|\mathbf{f}\|$ is constant if and only if $\|\mathbf{f}\|^2 = \mathbf{f} \cdot \mathbf{f}$ is constant, and this occurs if and only if $\mathbf{f} \cdot \mathbf{f}' = 0$. $\square$

Remark. This result is intuitively clear. If a particle moves on a sphere centered at the origin, then its velocity vector must be orthogonal to its position vector; any component in the direction of the position vector would move the particle off the sphere. Similarly, suppose $\mathbf{f}$ and $\mathbf{g}$ have constant length and a constant angle between them. Then in order to maintain the constant angle, as $\mathbf{f}$ turns towards $\mathbf{g}$, we see that $\mathbf{g}$ must turn *away* from $\mathbf{f}$ at the same rate.

Using Lemma 2.1 repeatedly, we now construct the *Frenet frame* of suitable regular curves. We assume throughout that the curve $\boldsymbol{\alpha}$ is parametrized by arclength. Then, for starters, $\boldsymbol{\alpha}'(s)$ is the *unit tangent vector* to the curve, which we denote by $\mathbf{T}(s)$. Since $\mathbf{T}$ has constant length, $\mathbf{T}'(s)$ will be orthogonal to $\mathbf{T}(s)$. Assuming $\mathbf{T}'(s) \neq \mathbf{0}$, define the *principal normal vector* $\mathbf{N}(s) = \mathbf{T}'(s)/\|\mathbf{T}'(s)\|$ and the *curvature* $\kappa(s) = \|\mathbf{T}'(s)\|$. So far, we have

$$\mathbf{T}'(s) = \kappa(s)\mathbf{N}(s).$$

If $\kappa(s) = 0$, the principal normal vector is not defined. Assuming $\kappa \neq 0$, we continue. Define the *binormal vector* $\mathbf{B}(s) = \mathbf{T}(s) \times \mathbf{N}(s)$. Then $\{\mathbf{T}(s), \mathbf{N}(s), \mathbf{B}(s)\}$ form a right-handed orthonormal basis for $\mathbb{R}^3$.

Now, $\mathbf{N}'(s)$ must be a linear combination of $\mathbf{T}(s)$, $\mathbf{N}(s)$, and $\mathbf{B}(s)$. But we know from Lemma 2.1 that $\mathbf{N}'(s)\cdot\mathbf{N}(s) = 0$ and $\mathbf{N}'(s)\cdot\mathbf{T}(s) = -\mathbf{T}'(s)\cdot\mathbf{N}(s) = -\kappa(s)$. We define the *torsion* $\tau(s) = \mathbf{N}'(s)\cdot\mathbf{B}(s)$. This gives us

$$\mathbf{N}'(s) = -\kappa(s)\mathbf{T}(s) + \tau(s)\mathbf{B}(s).$$

Finally, $\mathbf{B}'(s)$ must be a linear combination of $\mathbf{T}(s)$, $\mathbf{N}(s)$, and $\mathbf{B}(s)$. Lemma 2.1 tells us that $\mathbf{B}'(s)\cdot\mathbf{B}(s) = 0$, $\mathbf{B}'(s)\cdot\mathbf{T}(s) = -\mathbf{T}'(s)\cdot\mathbf{B}(s) = 0$, and $\mathbf{B}'(s)\cdot\mathbf{N}(s) = -\mathbf{N}'(s)\cdot\mathbf{B}(s) = -\tau(s)$. Thus,

$$\mathbf{B}'(s) = -\tau(s)\mathbf{N}(s).$$

In summary, we have:

Frenet formulas

$$\boxed{\begin{aligned}
\mathbf{T}'(s) &= & \kappa(s)\mathbf{N}(s) & \\
\mathbf{N}'(s) &= -\kappa(s)\mathbf{T}(s) & & + \tau(s)\mathbf{B}(s)\\
\mathbf{B}'(s) &= & -\tau(s)\mathbf{N}(s) &
\end{aligned}}$$

The skew-symmetry of these equations is made clearest when we state the Frenet formulas in matrix form:

$$\begin{bmatrix} | & | & | \\ \mathbf{T}'(s) & \mathbf{N}'(s) & \mathbf{B}'(s) \\ | & | & | \end{bmatrix} = \begin{bmatrix} | & | & | \\ \mathbf{T}(s) & \mathbf{N}(s) & \mathbf{B}(s) \\ | & | & | \end{bmatrix} \begin{bmatrix} 0 & -\kappa(s) & 0 \\ \kappa(s) & 0 & -\tau(s) \\ 0 & \tau(s) & 0 \end{bmatrix}.$$

Indeed, note that the coefficient matrix appearing on the right is skew-symmetric. This is the case whenever we differentiate an orthogonal matrix depending on a parameter (s in this case). (See Exercise A.1.4.)

Note that, by definition, the curvature, κ, is always nonnegative; the torsion, τ, however, has a sign, as we shall now see.

Example 1. Consider the helix, given by its arclength parametrization (see Exercise 1.1.2) $\boldsymbol{\alpha}(s) = \big(a\cos(s/c), a\sin(s/c), bs/c\big)$, where $c = \sqrt{a^2+b^2}$ and $a > 0$. Then we have

$$\mathbf{T}(s) = \frac{1}{c}\big(-a\sin(s/c), a\cos(s/c), b\big)$$

$$\mathbf{T}'(s) = \frac{1}{c^2}\big(-a\cos(s/c), -a\sin(s/c), 0\big) = \underbrace{\frac{a}{c^2}}_{\kappa(s)}\underbrace{\big(-\cos(s/c), -\sin(s/c), 0\big)}_{\mathbf{N}(s)}.$$

Summarizing,

$$\kappa(s) = \frac{a}{c^2} = \frac{a}{a^2+b^2} \qquad \text{and} \qquad \mathbf{N}(s) = \big(-\cos(s/c), -\sin(s/c), 0\big).$$

Now we deal with **B** and the torsion:

$$\mathbf{B}(s) = \mathbf{T}(s) \times \mathbf{N}(s) = \frac{1}{c}\big(b\sin(s/c), -b\cos(s/c), a\big)$$
$$\mathbf{B}'(s) = \frac{1}{c^2}\big(b\cos(s/c), b\sin(s/c), 0\big) = -\frac{b}{c^2}\mathbf{N}(s),$$

so we infer that $\tau(s) = \dfrac{b}{c^2} = \dfrac{b}{a^2+b^2}$.

Note that both the curvature and the torsion are constants. The torsion is positive when the helix is "right-handed" ($b > 0$) and negative when the helix is "left-handed" ($b < 0$). It is interesting to observe that, fixing $a > 0$, as $b \to 0$, the helix becomes very tightly wound and almost planar, and $\tau \to 0$; as $b \to \infty$, the helix twists extremely slowly and looks more and more like a straight line on the cylinder and, once again, $\tau \to 0$. As the reader can check, the helix has the greatest torsion when $b = a$; why does this seem plausible?

In Figure 2.1 we show the Frenet frames of the helix at some sample points. (In the latter two pictures,

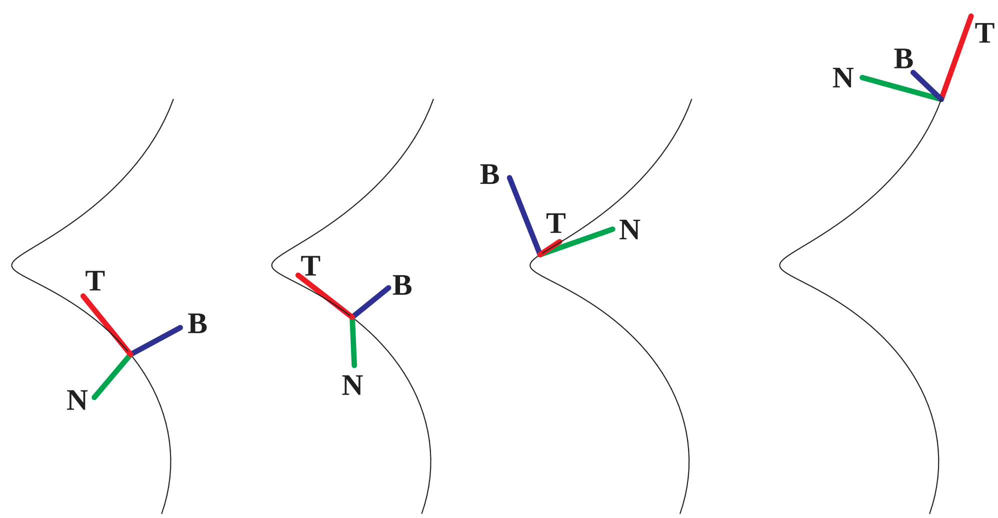


FIGURE 2.1

the perspective is misleading. $\mathbf{T}, \mathbf{N}, \mathbf{B}$ still form a right-handed frame: In the third, **T** is in front of **N**, and in the last, **B** is pointing upwards and out of the page.) ∇

We stop for a moment to contemplate what happens with the Frenet formulas when we are dealing with a *non*-arclength-parametrized, regular curve $\boldsymbol{\alpha}$. As we did in Section 1, we can (theoretically) reparametrize by arclength, obtaining $\boldsymbol{\beta}(s)$. Then we have $\boldsymbol{\alpha}(t) = \boldsymbol{\beta}(s(t))$, so, by the chain rule,

$$(*) \qquad \boldsymbol{\alpha}'(t) = \boldsymbol{\beta}'(s(t))s'(t) = \upsilon(t)\mathbf{T}(s(t)),$$

where $\upsilon(t) = s'(t)$ is the speed.[3] Similarly, by the chain rule, once we have the unit tangent vector as a function of t, differentiating with respect to t, we have

$$(\mathbf{T}\circ s)'(t) = \mathbf{T}'(s(t))s'(t) = \upsilon(t)\kappa(s(t))\mathbf{N}(s(t)).$$

Using the more casual—but convenient—Leibniz notation for derivatives,

$$\frac{d\mathbf{T}}{dt} = \frac{d\mathbf{T}}{ds}\frac{ds}{dt} = \upsilon\kappa\mathbf{N} \qquad \text{or} \qquad \kappa\mathbf{N} = \frac{d\mathbf{T}}{ds} = \frac{\frac{d\mathbf{T}}{dt}}{\frac{ds}{dt}} = \frac{1}{\upsilon}\frac{d\mathbf{T}}{dt}.$$

Example 2. Let's calculate the curvature of the tractrix (see Example 2 in Section 1). Using the first parametrization, we have $\boldsymbol{\alpha}'(\theta) = (-\sin\theta + \csc\theta, \cos\theta)$, and so

$$\upsilon(\theta) = \|\boldsymbol{\alpha}'(\theta)\| = \sqrt{(-\sin\theta + \csc\theta)^2 + \cos^2\theta} = \sqrt{\csc^2\theta - 1} = -\cot\theta.$$

(Note the negative sign because $\frac{\pi}{2} \le \theta < \pi$.) Therefore,

$$\mathbf{T}(\theta) = -\frac{1}{\cot\theta}(-\sin\theta + \csc\theta, \cos\theta) = -\tan\theta(\cot\theta\cos\theta, \cos\theta) = (-\cos\theta, -\sin\theta).$$

Of course, looking at Figure 1.9, we should expect the formula for $\mathbf{T}$. Then, to find the curvature, we calculate

$$\kappa\mathbf{N} = \frac{d\mathbf{T}}{ds} = \frac{\frac{d\mathbf{T}}{d\theta}}{\frac{ds}{d\theta}} = \frac{(\sin\theta, -\cos\theta)}{-\cot\theta} = (-\tan\theta)(\sin\theta, -\cos\theta).$$

Since $-\tan\theta > 0$ and $(\sin\theta, -\cos\theta)$ is a unit vector we conclude that

$$\kappa(\theta) = -\tan\theta \qquad \text{and} \qquad \mathbf{N}(\theta) = (\sin\theta, -\cos\theta).$$

Later on we will see an interesting geometric consequence of the equality of the curvature and the (absolute value of) the slope. ∇

Example 3. Let's calculate the "Frenet apparatus" for the parametrized curve

$$\boldsymbol{\alpha}(t) = (3t - t^3, 3t^2, 3t + t^3).$$

We begin by calculating $\boldsymbol{\alpha}'$ and determining the unit tangent vector $\mathbf{T}$ and speed υ:

$$\boldsymbol{\alpha}'(t) = 3(1 - t^2, 2t, 1 + t^2), \qquad \text{so}$$

$$\upsilon(t) = \|\boldsymbol{\alpha}'(t)\| = 3\sqrt{(1-t^2)^2 + (2t)^2 + (1+t^2)^2} = 3\sqrt{2(1+t^2)^2} = 3\sqrt{2}(1+t^2) \qquad \text{and}$$

$$\mathbf{T}(t) = \frac{1}{\sqrt{2}}\frac{1}{1+t^2}(1-t^2, 2t, 1+t^2) = \frac{1}{\sqrt{2}}\left(\frac{1-t^2}{1+t^2}, \frac{2t}{1+t^2}, 1\right).$$

Now

$$\begin{aligned} \kappa\mathbf{N} &= \frac{d\mathbf{T}}{ds} = \frac{\frac{d\mathbf{T}}{dt}}{\frac{ds}{dt}} = \frac{1}{\upsilon(t)}\frac{d\mathbf{T}}{dt} \\ &= \frac{1}{3\sqrt{2}(1+t^2)}\frac{1}{\sqrt{2}}\left(\frac{-4t}{(1+t^2)^2}, \frac{2(1-t^2)}{(1+t^2)^2}, 0\right) \end{aligned}$$

[3] υ is the Greek letter *upsilon*, not to be confused with ν, the Greek letter *nu*.

$$= \underbrace{\frac{1}{3\sqrt{2}(1+t^2)}\frac{1}{\sqrt{2}}\cdot\frac{2}{1+t^2}}_{\kappa}\underbrace{\left(-\frac{2t}{1+t^2},\frac{1-t^2}{1+t^2},0\right)}_{\mathbf{N}}.$$

Here we have factored out the length of the derivative vector and left ourselves with a unit vector in its direction, which must be the principal normal $\mathbf{N}$; the magnitude that is left must be the curvature κ. In summary, so far we have

$$\kappa(t) = \frac{1}{3(1+t^2)^2} \qquad \text{and} \qquad \mathbf{N}(t) = \left(-\frac{2t}{1+t^2},\frac{1-t^2}{1+t^2},0\right).$$

Next we find the binormal $\mathbf{B}$ by calculating the cross product

$$\mathbf{B}(t) = \mathbf{T}(t)\times\mathbf{N}(t) = \frac{1}{\sqrt{2}}\left(-\frac{1-t^2}{1+t^2},-\frac{2t}{1+t^2},1\right).$$

And now, at long last, we calculate the torsion by differentiating $\mathbf{B}$:

$$\begin{aligned}
-\tau\mathbf{N} &= \frac{d\mathbf{B}}{ds} = \frac{\frac{d\mathbf{B}}{dt}}{\frac{ds}{dt}} = \frac{1}{\upsilon(t)}\frac{d\mathbf{B}}{dt} \\
&= \frac{1}{3\sqrt{2}(1+t^2)}\frac{1}{\sqrt{2}}\left(\frac{4t}{(1+t^2)^2},\frac{2(t^2-1)}{(1+t^2)^2},0\right) \\
&= -\underbrace{\frac{1}{3(1+t^2)^2}}_{\tau}\underbrace{\left(-\frac{2t}{1+t^2},\frac{1-t^2}{1+t^2},0\right)}_{\mathbf{N}},
\end{aligned}$$

so $\tau(t) = \kappa(t) = \dfrac{1}{3(1+t^2)^2}$. $\quad\nabla$

Now we see that curvature enters naturally when we compute the acceleration of a moving particle. Differentiating the formula $(*)$ on p. 12, we obtain

$$\begin{aligned}
\boldsymbol{\alpha}''(t) &= \upsilon'(t)\mathbf{T}(s(t)) + \upsilon(t)\mathbf{T}'(s(t))s'(t) \\
&= \upsilon'(t)\mathbf{T}(s(t)) + \upsilon(t)^2\big(\kappa(s(t))\mathbf{N}(s(t))\big).
\end{aligned}$$

Suppressing the variables for a moment, we can rewrite this equation as

$$(**) \qquad \boldsymbol{\alpha}'' = \upsilon'\mathbf{T} + \kappa\upsilon^2\mathbf{N}.$$

The tangential component of acceleration is the derivative of speed; the normal component (the "centripetal acceleration" in the case of circular motion) is the product of the curvature of the path *and* the square of the speed. Thus, from the physics of the motion we can recover the curvature of the path:

Proposition 2.2. *For any regular parametrized curve* $\boldsymbol{\alpha}$*, we have* $\kappa = \dfrac{\|\boldsymbol{\alpha}'\times\boldsymbol{\alpha}''\|}{\|\boldsymbol{\alpha}'\|^3}$.

Proof. Since $\boldsymbol{\alpha}'\times\boldsymbol{\alpha}'' = (\upsilon\mathbf{T})\times(\upsilon'\mathbf{T}+\kappa\upsilon^2\mathbf{N}) = \kappa\upsilon^3\mathbf{T}\times\mathbf{N}$ and $\kappa\upsilon^3 > 0$, we obtain $\kappa\upsilon^3 = \|\boldsymbol{\alpha}'\times\boldsymbol{\alpha}''\|$, and so $\kappa = \|\boldsymbol{\alpha}'\times\boldsymbol{\alpha}''\|/\upsilon^3$, as desired. $\square$

We next proceed to study various theoretical consequences of the Frenet formulas.

Proposition 2.3. *A space curve is a line if and only if its curvature is everywhere* 0.

Proof. The general line is given by $\boldsymbol{\alpha}(s) = s\mathbf{v} + \mathbf{c}$ for some unit vector $\mathbf{v}$ and constant vector $\mathbf{c}$. Then $\boldsymbol{\alpha}'(s) = \mathbf{T}(s) = \mathbf{v}$ is constant, so $\kappa = 0$. Conversely, if $\kappa = 0$, then $\mathbf{T}(s) = \mathbf{T_0}$ is a constant vector, and, integrating, we obtain $\boldsymbol{\alpha}(s) = \displaystyle\int_0^s \mathbf{T}(u)du + \boldsymbol{\alpha}(0) = s\mathbf{T_0} + \boldsymbol{\alpha}(0)$. This is, once again, the parametric equation of a line. $\square$

Example 4. Suppose all the tangent lines of a space curve pass through a fixed point. What can we say about the curve? Without loss of generality, we take the fixed point to be the origin and the curve to be arclength-parametrized by $\boldsymbol{\alpha}$. Then there is a scalar function λ so that for every s we have $\boldsymbol{\alpha}(s) = \lambda(s)\mathbf{T}(s)$. Differentiating, we have

$$\mathbf{T}(s) = \boldsymbol{\alpha}'(s) = \lambda'(s)\mathbf{T}(s) + \lambda(s)\mathbf{T}'(s) = \lambda'(s)\mathbf{T}(s) + \lambda(s)\kappa(s)\mathbf{N}(s).$$

Then $(\lambda'(s) - 1)\mathbf{T}(s) + \lambda(s)\kappa(s)\mathbf{N}(s) = \mathbf{0}$, so, since $\mathbf{T}(s)$ and $\mathbf{N}(s)$ are linearly independent, we infer that $\lambda(s) = s + c$ for some constant c and $\kappa(s) = 0$. Therefore, the curve must be a line through the fixed point. ∇

Somewhat more challenging is the following

Proposition 2.4. *A space curve is planar if and only if its torsion is everywhere* 0. *The only planar curves with nonzero constant curvature are (portions of) circles.*

Proof. If a curve lies in a plane $\mathcal{P}$, then $\mathbf{T}(s)$ and $\mathbf{N}(s)$ span the plane $\mathcal{P}_0$ parallel to $\mathcal{P}$ and passing through the origin. Therefore, $\mathbf{B} = \mathbf{T} \times \mathbf{N}$ is a constant vector (the normal to $\mathcal{P}_0$), and so $\mathbf{B}' = -\tau\mathbf{N} = \mathbf{0}$, from which we conclude that $\tau = 0$. Conversely, if $\tau = 0$, the binormal vector $\mathbf{B}$ is a constant vector $\mathbf{B_0}$. Now, consider the function $f(s) = \boldsymbol{\alpha}(s) \cdot \mathbf{B_0}$; we have $f'(s) = \boldsymbol{\alpha}'(s) \cdot \mathbf{B_0} = \mathbf{T}(s) \cdot \mathbf{B}(s) = 0$, and so $f(s) = c$ for some constant c. This means that $\boldsymbol{\alpha}$ lies in the plane $\mathbf{x} \cdot \mathbf{B_0} = c$.

We leave it to the reader to check in Exercise 2a. that a circle of radius a has constant curvature $1/a$. (This can also be deduced as a special case of the calculation in Example 1.) Now suppose a planar curve $\boldsymbol{\alpha}$ has constant curvature κ_0. Consider the auxiliary function $\boldsymbol{\beta}(s) = \boldsymbol{\alpha}(s) + \dfrac{1}{\kappa_0}\mathbf{N}(s)$. Then we have $\boldsymbol{\beta}'(s) = \boldsymbol{\alpha}'(s) + \dfrac{1}{\kappa_0}(-\kappa_0(s)\mathbf{T}(s)) = \mathbf{T}(s) - \mathbf{T}(s) = \mathbf{0}$. Therefore $\boldsymbol{\beta}$ is a constant function, say $\boldsymbol{\beta}(s) = P$ for all s. Now we claim that $\boldsymbol{\alpha}$ is a (subset of a) circle centered at P, for $\|\boldsymbol{\alpha}(s) - P\| = \|\boldsymbol{\alpha}(s) - \boldsymbol{\beta}(s)\| = 1/\kappa_0$. $\square$

We have already seen that a circular helix has constant curvature and torsion. We leave it to the reader to check in Exercise 10 that these are the only curves with constant curvature and torsion. Somewhat more interesting are the curves for which τ/κ is a constant.

A *generalized helix* is a space curve with $\kappa \neq 0$ all of whose tangent vectors make a constant angle with a fixed direction. As shown in Figure 2.2, this curve lies on a generalized cylinder, formed by taking the union of the lines (rulings) in that fixed direction through each point of the curve. We can now characterize generalized helices by the following

Proposition 2.5. *A curve is a generalized helix if and only if* τ/κ *is constant.*

Proof. Suppose $\boldsymbol{\alpha}$ is an arclength-parametrized generalized helix. Then there is a (constant) unit vector $\mathbf{A}$ with the property that $\mathbf{T} \cdot \mathbf{A} = \cos\theta$ for some constant θ. Differentiating, we obtain $\kappa\mathbf{N} \cdot \mathbf{A} = 0$, whence $\mathbf{N} \cdot \mathbf{A} = 0$. Differentiating yet again, we have

$$(-\kappa\mathbf{T} + \tau\mathbf{B}) \cdot \mathbf{A} = 0. \tag{†}$$

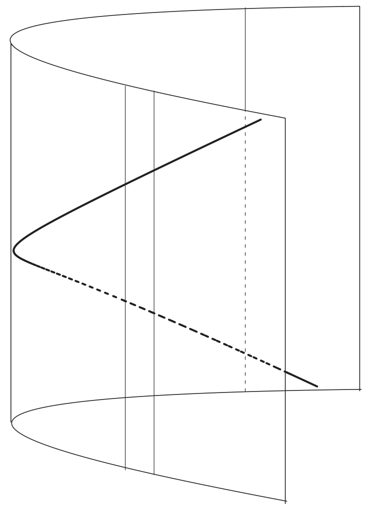


FIGURE 2.2

Now, note that $\mathbf{A}$ lies in the plane spanned by $\mathbf{T}$ and $\mathbf{B}$, and thus $\mathbf{B} \cdot \mathbf{A} = \pm \sin\theta$. Thus, we infer from equation (†) that $\tau/\kappa = \pm \cot\theta$, which is indeed constant.

Conversely, if τ/κ is constant, set $\tau/\kappa = \cot\theta$ for some angle $\theta \in (0, \pi)$. Set $\mathbf{A}(s) = \cos\theta \mathbf{T}(s) + \sin\theta \mathbf{B}(s)$. Then $\mathbf{A}'(s) = (\kappa\cos\theta - \tau\sin\theta)\mathbf{N}(s) = \mathbf{0}$, so $\mathbf{A}(s)$ is a constant unit vector $\mathbf{A}$, and $\mathbf{T}(s) \cdot \mathbf{A} = \cos\theta$ is constant, as desired. $\square$

Example 5. In Example 3 we saw a curve $\boldsymbol{\alpha}$ with $\kappa = \tau$, so from the proof of Proposition 2.5 we see that the curve should make a constant angle $\theta = \pi/4$ with the vector $\mathbf{A} = \dfrac{1}{\sqrt{2}}(\mathbf{T} + \mathbf{B}) = (0, 0, 1)$ (as should have been obvious from the formula for $\mathbf{T}$ alone). We verify this in Figure 2.3 by drawing $\boldsymbol{\alpha}$ along with the vertical cylinder built on the projection of $\boldsymbol{\alpha}$ onto the xy-plane. ∇

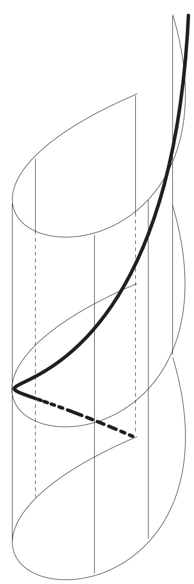


FIGURE 2.3

The Frenet formulas actually characterize the local picture of a space curve.

Proposition 2.6 (Local canonical form)**.** *Let* $\boldsymbol{\alpha}$ *be a smooth* ($\mathcal{C}^3$ *or better) arclength-parametrized curve. If* $\boldsymbol{\alpha}(0) = \mathbf{0}$*, then for* s *near* 0*, we have*

$$\boldsymbol{\alpha}(s) = \left(s - \frac{\kappa_0^2}{6}s^3 + \dots\right)\mathbf{T}(0) + \left(\frac{\kappa_0}{2}s^2 + \frac{\kappa_0'}{6}s^3 + \dots\right)\mathbf{N}(0) + \left(\frac{\kappa_0\tau_0}{6}s^3 + \dots\right)\mathbf{B}(0).$$

(Here κ_0*,* τ_0*, and* κ_0' *denote, respectively, the values of* κ*,* τ*, and* κ' *at* 0*, and* $\lim\limits_{s\to 0} \dots/s^3 = 0$*.)*

Proof. Using Taylor's Theorem, we write

$$\boldsymbol{\alpha}(s) = \boldsymbol{\alpha}(0) + s\boldsymbol{\alpha}'(0) + \frac{1}{2}s^2\boldsymbol{\alpha}''(0) + \frac{1}{6}s^3\boldsymbol{\alpha}'''(0) + \dots,$$

where $\lim\limits_{s\to 0} \dots/s^3 = 0$. Now, $\boldsymbol{\alpha}(0) = \mathbf{0}$, $\boldsymbol{\alpha}'(0) = \mathbf{T}(0)$, and $\boldsymbol{\alpha}''(0) = \mathbf{T}'(0) = \kappa_0\mathbf{N}(0)$. Differentiating again, we have $\boldsymbol{\alpha}'''(0) = (\kappa\mathbf{N})'(0) = \kappa_0'\mathbf{N}(0) + \kappa_0(-\kappa_0\mathbf{T}(0) + \tau_0\mathbf{B}(0))$. Substituting, we obtain

$$\begin{aligned}\boldsymbol{\alpha}(s) &= s\mathbf{T}(0) + \frac{1}{2}s^2\kappa_0\mathbf{N}(0) + \frac{1}{6}s^3\left(-\kappa_0^2\mathbf{T}(0) + \kappa_0'\mathbf{N}(0) + \kappa_0\tau_0\mathbf{B}(0)\right) + \dots \\ &= \left(s - \frac{\kappa_0^2}{6}s^3 + \dots\right)\mathbf{T}(0) + \left(\frac{\kappa_0}{2}s^2 + \frac{\kappa_0'}{6}s^3 + \dots\right)\mathbf{N}(0) + \left(\frac{\kappa_0\tau_0}{6}s^3 + \dots\right)\mathbf{B}(0),\end{aligned}$$

as required. □

We now introduce three fundamental planes at $P = \boldsymbol{\alpha}(0)$:

(i) the *osculating plane*, spanned by $\mathbf{T}(0)$ and $\mathbf{N}(0)$,
(ii) the *rectifying plane*, spanned by $\mathbf{T}(0)$ and $\mathbf{B}(0)$, and
(iii) the *normal plane*, spanned by $\mathbf{N}(0)$ and $\mathbf{B}(0)$.

We see that, locally, the projections of $\boldsymbol{\alpha}$ into these respective planes look like

(i) $\left(u - (\kappa_0^2/6)u^3 + \dots, (\kappa_0/2)u^2 + (\kappa_0'/6)u^3 + \dots\right)$
(ii) $\left(u - (\kappa_0^2/6)u^3 + \dots, (\kappa_0\tau_0/6)u^3 + \dots\right)$, and
(iii) $\left((\kappa_0/2)u^2 + (\kappa_0'/6)u^3 + \dots, (\kappa_0\tau_0/6)u^3 + \dots\right)$,

where $\lim\limits_{u\to 0} \dots/u^3 = 0$. Thus, the projections of $\boldsymbol{\alpha}$ into these planes look locally as shown in Figure 2.4. The osculating ("kissing") plane is the plane that comes closest to containing $\boldsymbol{\alpha}$ near P (see also Exercise

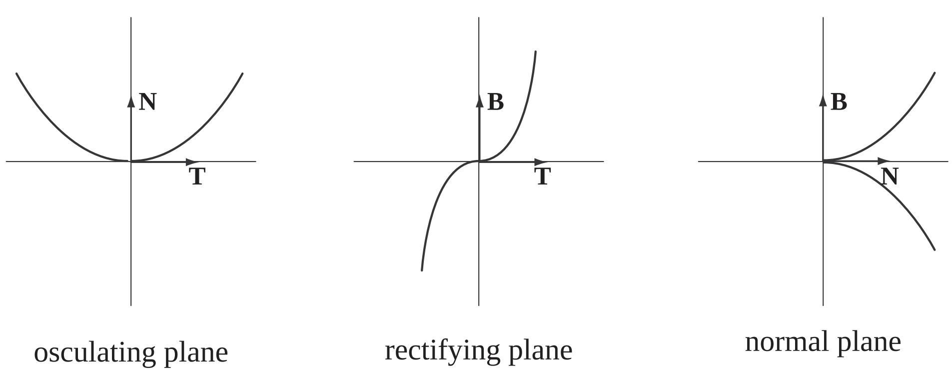


FIGURE 2.4

25); the rectifying ("straightening") plane is the one that comes closest to flattening the curve near P; the normal plane is normal (perpendicular) to the curve at P. (Cf. Figure 1.3.)

EXERCISES 1.2

1. Compute the curvature of the following arclength-parametrized curves:
   a. $\boldsymbol{\alpha}(s) = \left(\frac{1}{\sqrt{2}}\cos s, \frac{1}{\sqrt{2}}\cos s, \sin s\right)$
   b. $\boldsymbol{\alpha}(s) = \left(\sqrt{1+s^2}, \ln(s+\sqrt{1+s^2})\right)$
   *c. $\boldsymbol{\alpha}(s) = \left(\frac{1}{3}(1+s)^{3/2}, \frac{1}{3}(1-s)^{3/2}, \frac{1}{\sqrt{2}}s\right)$, $s \in (-1,1)$
2. Calculate the unit tangent vector, principal normal, and curvature of the following curves:
   a. a circle of radius a: $\boldsymbol{\alpha}(t) = (a\cos t, a\sin t)$
   b. $\boldsymbol{\alpha}(t) = (t, \cosh t)$
   c. $\boldsymbol{\alpha}(t) = (\cos^3 t, \sin^3 t)$, $t \in (0, \pi/2)$
3. Calculate the Frenet apparatus ($\mathbf{T}$, κ, $\mathbf{N}$, $\mathbf{B}$, and τ) of the following curves:
   *a. $\boldsymbol{\alpha}(s) = \left(\frac{1}{3}(1+s)^{3/2}, \frac{1}{3}(1-s)^{3/2}, \frac{1}{\sqrt{2}}s\right)$, $s \in (-1,1)$
   b. $\boldsymbol{\alpha}(t) = \left(\frac{1}{2}e^t(\sin t + \cos t), \frac{1}{2}e^t(\sin t - \cos t), e^t\right)$
   *c. $\boldsymbol{\alpha}(t) = \left(\sqrt{1+t^2}, t, \ln(t+\sqrt{1+t^2})\right)$
   d. $\boldsymbol{\alpha}(t) = (e^t\cos t, e^t\sin t, e^t)$
   e. $\boldsymbol{\alpha}(t) = (\cosh t, \sinh t, t)$
   f. $\boldsymbol{\alpha}(t) = \left(t, t^2/2, t\sqrt{1+t^2} + \ln(t+\sqrt{1+t^2})\right)$
   g. $\boldsymbol{\alpha}(t) = (t - \sin t\cos t, \sin^2 t, \cos t)$, $t \in (0,\pi)$

♯4. Prove that the curvature of the plane curve $y = f(x)$ is given by $\kappa = \dfrac{|f''|}{(1+f'^2)^{3/2}}$.

♯*5. Use Proposition 2.2 and the second parametrization of the tractrix given in Example 2 of Section 1 to recompute the curvature.

*6. By differentiating the equation $\mathbf{B} = \mathbf{T}\times\mathbf{N}$, derive the equation $\mathbf{B}' = -\tau\mathbf{N}$.

♯7. Suppose $\boldsymbol{\alpha}$ is an arclength-parametrized space curve with the property that $\|\boldsymbol{\alpha}(s)\| \le \|\boldsymbol{\alpha}(s_0)\| = R$ for all s sufficiently close to s_0. Prove that $\kappa(s_0) \ge 1/R$. (Hint: Consider the function $f(s) = \|\boldsymbol{\alpha}(s)\|^2$. What do you know about $f''(s_0)$?)

8. Let $\boldsymbol{\alpha}$ be a regular (arclength-parametrized) curve with nonzero curvature. The normal line to $\boldsymbol{\alpha}$ at $\boldsymbol{\alpha}(s)$ is the line through $\boldsymbol{\alpha}(s)$ with direction vector $\mathbf{N}(s)$. Suppose all the normal lines to $\boldsymbol{\alpha}$ pass through a fixed point. What can you say about the curve?

9. a. Prove that if all the normal planes of a curve pass through a particular point, then the curve lies on a sphere. (Hint: Apply Lemma 2.1.)
   *b. Prove that if all the osculating planes of a curve pass through a particular point, then the curve is planar.

10. Prove that if $\kappa = \kappa_0$ and $\tau = \tau_0$ are nonzero constants, then the curve is a (right) circular helix.

(Hint: Start by solving for $\mathbf{N}$. The only solutions of the differential equation $y'' + k^2 y = 0$ are $y = c_1 \cos(kt) + c_2 \sin(kt)$.)

Remark. It is an amusing exercise to give a and b (in our formula for the circular helix) in terms of κ_0 and τ_0.

*11. Proceed as in the derivation of Proposition 2.2 to show that

$$\tau = \frac{\boldsymbol{\alpha}' \cdot (\boldsymbol{\alpha}'' \times \boldsymbol{\alpha}''')}{\|\boldsymbol{\alpha}' \times \boldsymbol{\alpha}''\|^2}.$$

12. Let $\boldsymbol{\alpha}$ be a $\mathcal{C}^4$ arclength-parametrized curve with $\kappa \neq 0$. Prove that $\boldsymbol{\alpha}$ is a generalized helix if and only if $\boldsymbol{\alpha}'' \cdot (\boldsymbol{\alpha}''' \times \boldsymbol{\alpha}^{(\mathrm{iv})}) = 0$. (Here $\boldsymbol{\alpha}^{(\mathrm{iv})}$ denotes the fourth derivative of $\boldsymbol{\alpha}$.)

13. Suppose $\kappa\tau \neq 0$ at P. Of all the planes containing the tangent line to $\boldsymbol{\alpha}$ at P, show that $\boldsymbol{\alpha}$ lies *locally* on both sides only of the osculating plane.

14. Let $\boldsymbol{\alpha}$ be a regular curve with $\kappa \neq 0$ at P. Prove that the planar curve obtained by projecting $\boldsymbol{\alpha}$ into its osculating plane at P has the same curvature at P as $\boldsymbol{\alpha}$.

15. A closed, planar curve C is said to have *constant breadth* μ if the distance between parallel tangent lines to C is always μ. (No, C needn't be a circle. See Figure 2.5.) Assume for the rest of this problem that the curve is arclength parametrized by a $\mathcal{C}^2$ function $\boldsymbol{\alpha}\colon [0, L] \to \mathbb{R}^2$ with $\kappa \neq 0$. To say C is closed means $\boldsymbol{\alpha}(0) = \boldsymbol{\alpha}(L)$ and the derivatives match as well.

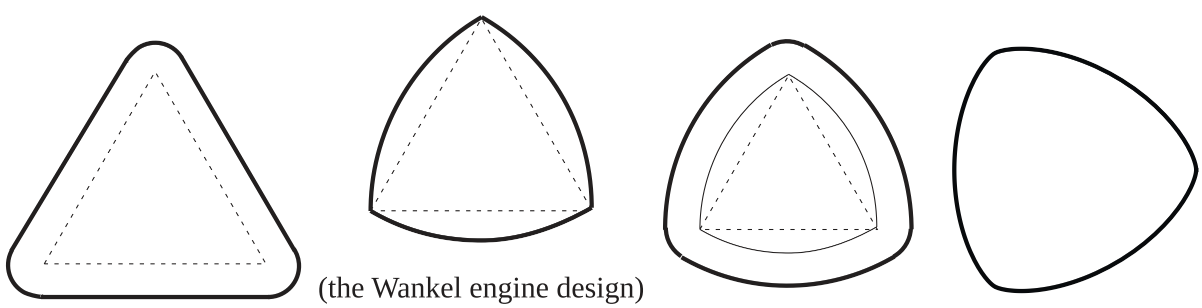


FIGURE 2.5

a. Let's call two points with parallel tangent lines opposite. Prove that if C has constant breadth μ, then the chord joining opposite points is normal to the curve at both points. (Hint: If $\boldsymbol{\beta}(s)$ is opposite $\boldsymbol{\alpha}(s)$, then $\boldsymbol{\beta}(s) = \boldsymbol{\alpha}(s) + \lambda(s)\mathbf{T}(s) + \mu\mathbf{N}(s)$. First explain why the coefficient of $\mathbf{N}$ is μ; then show that $\lambda = 0$.)

b. Prove that the sum of the reciprocals of the curvature at opposite points is equal to μ. (Warning: If $\boldsymbol{\alpha}$ is arclength-parametrized, $\boldsymbol{\beta}$ is quite unlikely to be. It might be helpful to introduce the notation $\mathbf{T}_{\boldsymbol{\beta}}$ and $\mathbf{N}_{\boldsymbol{\beta}}$ for the unit tangent vector and principal normal of $\boldsymbol{\beta}$. How are they related to $\mathbf{T}$ and $\mathbf{N}$?)

16. Let $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ be two regular curves defined on $[a, b]$. We say $\boldsymbol{\beta}$ is an *involute* of $\boldsymbol{\alpha}$ if, for each $t \in [a, b]$,

(i) $\boldsymbol{\beta}(t)$ lies on the tangent line to $\boldsymbol{\alpha}$ at $\boldsymbol{\alpha}(t)$, and

(ii) the tangent vectors to $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ at $\boldsymbol{\alpha}(t)$ and $\boldsymbol{\beta}(t)$, respectively, are perpendicular.

Reciprocally, we also refer to $\boldsymbol{\alpha}$ as an *evolute* of $\boldsymbol{\beta}$.

a. Suppose $\boldsymbol{\alpha}$ is arclength-parametrized. Show that $\boldsymbol{\beta}$ is an involute of $\boldsymbol{\alpha}$ if and only if $\boldsymbol{\beta}(s) = \boldsymbol{\alpha}(s) + (c - s)\mathbf{T}(s)$ for some constant c (here $\mathbf{T}(s) = \boldsymbol{\alpha}'(s)$). We will normally refer to the curve $\boldsymbol{\beta}$ obtained with $c = 0$ as *the* involute of $\boldsymbol{\alpha}$. If you were to wrap a string around the curve $\boldsymbol{\alpha}$, starting at $s = 0$, the involute is the path the end of the string follows as you unwrap it, always pulling the string taut, as illustrated in the case of a circle in Figure 2.6.

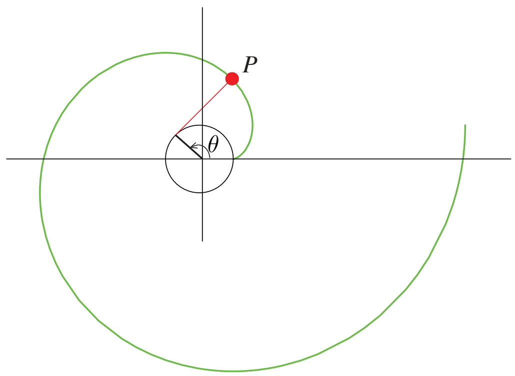


FIGURE 2.6

b. Show that the involute of a helix is a plane curve.

c. Show that the involute of a catenary is a tractrix. (Hint: You do not *need* an arclength parametrization!)

d. If $\boldsymbol{\alpha}$ is an arclength-parametrized plane curve, prove that the curve $\boldsymbol{\beta}$ given by

$$\boldsymbol{\beta}(s) = \boldsymbol{\alpha}(s) + \frac{1}{\kappa(s)}\mathbf{N}(s)$$

is the unique evolute of $\boldsymbol{\alpha}$ lying in the plane of $\boldsymbol{\alpha}$. Prove, moreover, that this curve is regular if $\kappa' \neq 0$. (Hint: Go back to the original definition.)

17. Find the involute of the cycloid $\boldsymbol{\alpha}(t) = (t + \sin t, 1 - \cos t)$, $t \in [-\pi, \pi]$, using $t = 0$ as your starting point. Give a geometric description of your answer.

18. Suppose $\boldsymbol{\alpha}$ is a generalized helix with axis in direction $\mathbf{A}$. Let $\boldsymbol{\beta}$ be the curve obtained by projecting $\boldsymbol{\alpha}$ onto a plane orthogonal to $\mathbf{A}$. Prove that the principal normals of $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are parallel at corresponding points and calculate the curvature of $\boldsymbol{\beta}$ in terms of the curvature of $\boldsymbol{\alpha}$.

19. Let $\boldsymbol{\alpha}$ be a curve parametrized by arclength with $\kappa, \tau \neq 0$.

a. Suppose $\boldsymbol{\alpha}$ lies on the surface of a sphere centered at the origin (i.e., $\|\boldsymbol{\alpha}(s)\| = \text{const}$ for all s). Prove that

$$(\star) \qquad \frac{\tau}{\kappa} + \left(\frac{1}{\tau}\left(\frac{1}{\kappa}\right)'\right)' = 0.$$

(Hint: Write $\boldsymbol{\alpha} = \lambda \mathbf{T} + \mu \mathbf{N} + \nu \mathbf{B}$ for some functions λ, μ, and ν, differentiate, and use the fact that $\{\mathbf{T}, \mathbf{N}, \mathbf{B}\}$ is a basis for $\mathbb{R}^3$.)

b. Prove the converse: If $\boldsymbol{\alpha}$ satisfies the differential equation ($\star$), then $\boldsymbol{\alpha}$ lies on the surface of some sphere. (Hint: Using the values of λ, μ, and ν you obtained in part a, show that $\boldsymbol{\alpha} - (\lambda \mathbf{T} + \mu \mathbf{N} + \nu \mathbf{B})$ is a constant vector, the candidate for the center of the sphere. If the nature of this argument puzzles you, review the latter part of the proof of Proposition 2.4.)

20. Two distinct parametrized curves $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are called *Bertrand mates* if for each t, the normal line to $\boldsymbol{\alpha}$ at $\boldsymbol{\alpha}(t)$ equals the normal line to $\boldsymbol{\beta}$ at $\boldsymbol{\beta}(t)$. An example is pictured in Figure 2.7. Suppose $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are

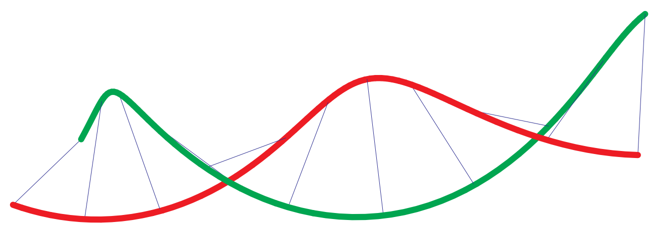


FIGURE 2.7

Bertrand mates.

a. If $\boldsymbol{\alpha}$ is arclength-parametrized, show that $\boldsymbol{\beta}(s) = \boldsymbol{\alpha}(s) + r(s)\mathbf{N}(s)$ and $r(s) = \text{const}$. Thus, corresponding points of $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are a constant distance apart.

b. Show that, moreover, the angle between the tangent vectors to $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ at corresponding points is constant. (Hint: If $\mathbf{T}_{\boldsymbol{\alpha}}$ and $\mathbf{T}_{\boldsymbol{\beta}}$ are the unit tangent vectors to $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ respectively, consider $\mathbf{T}_{\boldsymbol{\alpha}} \cdot \mathbf{T}_{\boldsymbol{\beta}}$.)

c. Suppose $\boldsymbol{\alpha}$ is arclength-parametrized and $\kappa\tau \neq 0$. Show that $\boldsymbol{\alpha}$ has a Bertrand mate $\boldsymbol{\beta}$ if and only if there are constants r and c so that $r\kappa + c\tau = 1$. (Hint for $\Longrightarrow$: Interpret the result of part b using your formula for $\boldsymbol{\beta}'$ from part a.)

d. Given $\boldsymbol{\alpha}$, prove that if there is more than one curve $\boldsymbol{\beta}$ so that $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are Bertrand mates, then there are infinitely many such curves $\boldsymbol{\beta}$ and this occurs if and only if $\boldsymbol{\alpha}$ is a circular helix.

21. (See Exercise 20.) Suppose $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are Bertrand mates. Prove that the torsion of $\boldsymbol{\alpha}$ and the torsion of $\boldsymbol{\beta}$ at corresponding points have constant product.

22. Suppose $\mathbf{Y}$ is a $\mathcal{C}^2$ vector function on $[a, b]$ with $\|\mathbf{Y}\| = 1$ and $\mathbf{Y}$, $\mathbf{Y}'$, and $\mathbf{Y}''$ everywhere linearly independent. For any nonzero constant c, define $\boldsymbol{\alpha}(t) = c \int_a^t \big(\mathbf{Y}(u) \times \mathbf{Y}'(u)\big)du$, $t \in [a, b]$. Prove that the curve $\boldsymbol{\alpha}$ has constant torsion $1/c$. (Hint: Show that $\mathbf{B} = \pm \mathbf{Y}$.)

23. (See Exercise 20.) Suppose $\mathbf{Y}$ is a $\mathcal{C}^2$ arclength-parametrized curve on the unit sphere. For any nonzero constant a and $0 < \theta \leq \pi/2$, define

$$\boldsymbol{\alpha}(t) = a\left(\int_0^t \mathbf{Y}(s)ds + \cot\theta \int_0^t \big(\mathbf{Y}(s) \times \mathbf{Y}'(s)\big)ds\right).$$

Show that the curve $\boldsymbol{\alpha}$ has a Bertrand mate. (Hint: Show that $\mathbf{N} = \pm \mathbf{Y}'$.)

24. a. Let $\boldsymbol{\alpha}$ be an arclength-parametrized *plane* curve. We create a "parallel" curve $\boldsymbol{\beta}$ by taking $\boldsymbol{\beta} = \boldsymbol{\alpha} + \varepsilon\mathbf{N}$ (for a fixed small positive value of ε). Explain the terminology and express the curvature of $\boldsymbol{\beta}$ in terms of ε and the curvature of $\boldsymbol{\alpha}$.
    b. Now let $\boldsymbol{\alpha}$ be an arclength-parametrized *space* curve. Show that we can obtain a "parallel" curve $\boldsymbol{\beta}$ by taking $\boldsymbol{\beta} = \boldsymbol{\alpha} + \varepsilon\big((\cos\theta)\mathbf{N} + (\sin\theta)\mathbf{B}\big)$ for an appropriate function θ. How many such parallel curves are there?
    c. Sketch such a parallel curve for a circular helix $\boldsymbol{\alpha}$.

25. Suppose $\boldsymbol{\alpha}$ is an arclength-parametrized curve, $P = \boldsymbol{\alpha}(0)$, and $\kappa(0) \neq 0$. Use Proposition 2.6 to establish the following:
    *a. Let $Q = \boldsymbol{\alpha}(s)$ and $R = \boldsymbol{\alpha}(t)$. Show that the plane spanned by P, Q, and R approaches the osculating plane of $\boldsymbol{\alpha}$ at P as $s, t \to 0$.
    b. The *osculating circle* at P is the limiting position of the circle passing through P, Q, and R as $s, t \to 0$. Prove that the osculating circle has center $Z = P + \big(1/\kappa(0)\big)\mathbf{N}(0)$ and radius $1/\kappa(0)$.
    c. The *osculating sphere* at P is the limiting position of the sphere through P and three neighboring points on the curve, as the latter points tend to P independently. Prove that the osculating sphere has center

$$Z = P + \big(1/\kappa(0)\big)\mathbf{N}(0) + \big(1/\tau(0)(1/\kappa)'(0)\big)\mathbf{B}(0)$$

    and radius

$$\sqrt{(1/\kappa(0))^2 + (1/\tau(0)(1/\kappa)'(0))^2}.$$

    d. How is the result of part c related to Exercise 19?

26. a. Suppose $\boldsymbol{\beta}$ is a plane curve and C_s is the circle centered at $\boldsymbol{\beta}(s)$ with radius $r(s)$. Assuming $\boldsymbol{\beta}$ and r are differentiable functions, show that the circle C_s is contained inside the circle C_t whenever $t > s$ if and only if $\|\boldsymbol{\beta}'(s)\| \le r'(s)$ for all s.
    b. Let $\boldsymbol{\alpha}$ be arclength-parametrized plane curve and suppose κ is a decreasing function. Prove that the osculating circle at $\boldsymbol{\alpha}(s)$ lies inside the osculating circle at $\boldsymbol{\alpha}(t)$ whenever $t > s$. (See Exercise 25 for the definition of the osculating circle.)

27. Suppose the front wheel of a bicycle follows the arclength-parametrized plane curve $\boldsymbol{\alpha}$. Determine the path $\boldsymbol{\beta}$ of the rear wheel, 1 unit away, as shown in Figure 2.8. (Hint: If the front wheel is turned an

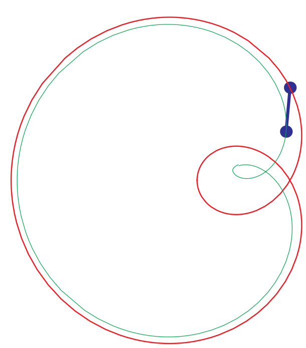


FIGURE 2.8

angle θ from the axle of the bike, start by writing $\boldsymbol{\alpha} - \boldsymbol{\beta}$ in terms of θ, $\mathbf{T}$, and $\mathbf{N}$. Your goal should be a differential equation that θ must satisfy, involving only κ. Note that the path of the rear wheel will obviously depend on the initial condition $\theta(0)$. In all but the simplest of cases, it may be impossible to solve the differential equation explicitly.)

3. Some Global Results

3.1. Space Curves. The fundamental notion in geometry (see Section 1 of the Appendix) is that of congruence: When do two figures differ merely by a rigid motion? If the curve $\boldsymbol{\alpha}^*$ is obtained from the curve $\boldsymbol{\alpha}$ by performing a rigid motion (composition of a translation and a rotation), then the Frenet frames at corresponding points differ by that same rigid motion, and the twisting of the frames (which is what gives curvature and torsion) should be the same. (Note that a reflection will not affect the curvature, but will change the sign of the torsion.)

Theorem 3.1 (Fundamental Theorem of Curve Theory)**.** *Two space curves C and C^* with nonzero curvature are congruent (i.e., differ by a rigid motion) if and only if the corresponding arclength parametrizations $\boldsymbol{\alpha}, \boldsymbol{\alpha}^*\colon [0, L] \to \mathbb{R}^3$ have the property that $\kappa(s) = \kappa^*(s)$ and $\tau(s) = \tau^*(s)$ for all $s \in [0, L]$.*

Proof. Suppose $\boldsymbol{\alpha}^* = \Psi \circ \boldsymbol{\alpha}$ for some rigid motion $\Psi\colon \mathbb{R}^3 \to \mathbb{R}^3$, so $\Psi(\mathbf{x}) = A\mathbf{x} + \mathbf{b}$ for some $\mathbf{b} \in \mathbb{R}^3$ and some 3×3 orthogonal matrix A with $\det A > 0$. Then $\boldsymbol{\alpha}^*(s) = A\boldsymbol{\alpha}(s) + \mathbf{b}$, so $\|\boldsymbol{\alpha}^{*\prime}(s)\| = \|A\boldsymbol{\alpha}'(s)\| = 1$, since A is orthogonal. Therefore, $\boldsymbol{\alpha}^*$ is likewise arclength-parametrized, and $\mathbf{T}^*(s) = A\mathbf{T}(s)$. Differentiating again, $\kappa^*(s)\mathbf{N}^*(s) = \kappa(s)A\mathbf{N}(s)$. Since A is orthogonal, $A\mathbf{N}(s)$ is a unit vector, and so $\mathbf{N}^*(s) = A\mathbf{N}(s)$ and $\kappa^*(s) = \kappa(s)$. But then $\mathbf{B}^*(s) = \mathbf{T}^*(s) \times \mathbf{N}^*(s) = A\mathbf{T}(s) \times A\mathbf{N}(s) = A(\mathbf{T}(s) \times \mathbf{N}(s)) = A\mathbf{B}(s)$, inasmuch as orthogonal matrices map orthonormal bases to orthonormal bases and $\det A > 0$ insures that orientation is preserved as well (i.e., right-handed bases map to right-handed bases). Last, $\mathbf{B}^{*\prime}(s) = -\tau^*(s)\mathbf{N}^*(s)$ and $\mathbf{B}^{*\prime}(s) = A\mathbf{B}'(s) = -\tau(s)A\mathbf{N}(s) = -\tau(s)\mathbf{N}^*(s)$, so $\tau^*(s) = \tau(s)$, as required.

Conversely, suppose $\kappa = \kappa^*$ and $\tau = \tau^*$. We now define a rigid motion Ψ as follows. Let A be the unique orthogonal matrix so that $A\mathbf{T}(0) = \mathbf{T}^*(0)$, $A\mathbf{N}(0) = \mathbf{N}^*(0)$, and $A\mathbf{B}(0) = \mathbf{B}^*(0)$, and let $\mathbf{b} = \boldsymbol{\alpha}^*(0) - A\boldsymbol{\alpha}(0)$. A also has positive determinant, since both orthonormal bases are right-handed. Set $\tilde{\boldsymbol{\alpha}} = \Psi \circ \boldsymbol{\alpha}$. We now claim that $\boldsymbol{\alpha}^*(s) = \tilde{\boldsymbol{\alpha}}(s)$ for all $s \in [0, L]$. Note, by our argument in the first part of the proof, that $\tilde{\kappa} = \kappa = \kappa^*$ and $\tilde{\tau} = \tau = \tau^*$. Consider

$$f(s) = \tilde{\mathbf{T}}(s) \cdot \mathbf{T}^*(s) + \tilde{\mathbf{N}}(s) \cdot \mathbf{N}^*(s) + \tilde{\mathbf{B}}(s) \cdot \mathbf{B}^*(s).$$

We now differentiate f, using the Frenet formulas.

$$\begin{aligned}
f'(s) &= \big(\tilde{\mathbf{T}}'(s) \cdot \mathbf{T}^*(s) + \tilde{\mathbf{T}}(s) \cdot \mathbf{T}^{*\prime}(s)\big) + \big(\tilde{\mathbf{N}}'(s) \cdot \mathbf{N}^*(s) + \tilde{\mathbf{N}}(s) \cdot \mathbf{N}^{*\prime}(s)\big) \\
&\qquad\qquad + \big(\tilde{\mathbf{B}}'(s) \cdot \mathbf{B}^*(s) + \tilde{\mathbf{B}}(s) \cdot \mathbf{B}^{*\prime}(s)\big) \\
&= \kappa(s)\big(\tilde{\mathbf{N}}(s) \cdot \mathbf{T}^*(s) + \tilde{\mathbf{T}}(s) \cdot \mathbf{N}^*(s)\big) - \kappa(s)\big(\tilde{\mathbf{T}}(s) \cdot \mathbf{N}^*(s) + \tilde{\mathbf{N}}(s) \cdot \mathbf{T}^*(s)\big) \\
&\qquad + \tau(s)\big(\tilde{\mathbf{B}}(s) \cdot \mathbf{N}^*(s) + \tilde{\mathbf{N}}(s) \cdot \mathbf{B}^*(s)\big) - \tau(s)\big(\tilde{\mathbf{N}}(s) \cdot \mathbf{B}^*(s) + \tilde{\mathbf{B}}(s) \cdot \mathbf{N}^*(s)\big) \\
&= 0,
\end{aligned}$$

since the first two terms cancel and the last two terms cancel. By construction, $f(0) = 3$, so $f(s) = 3$ for all $s \in [0, L]$. Since each of the individual dot products can be at most 1, the only way the sum can be 3 for all s is for each to be 1 for all s, and this in turn can happen only when $\tilde{\mathbf{T}}(s) = \mathbf{T}^*(s)$, $\tilde{\mathbf{N}}(s) = \mathbf{N}^*(s)$, and $\tilde{\mathbf{B}}(s) = \mathbf{B}^*(s)$ for all $s \in [0, L]$. In particular, since $\tilde{\boldsymbol{\alpha}}'(s) = \tilde{\mathbf{T}}(s) = \mathbf{T}^*(s) = \boldsymbol{\alpha}^{*\prime}(s)$ and $\tilde{\boldsymbol{\alpha}}(0) = \boldsymbol{\alpha}^*(0)$, it follows that $\tilde{\boldsymbol{\alpha}}(s) = \boldsymbol{\alpha}^*(s)$ for all $s \in [0, L]$, as we wished to show. $\square$

Remark. The latter half of this proof can be replaced by asserting the uniqueness of solutions of a system of differential equations, as we will see in a moment. Also see Exercise A.3.1 for a matrix-computational version of the proof we just did.

Example 1. We now see that the only curves with constant κ and τ are circular helices. ∇

Perhaps more interesting is the existence question: Given continuous functions $\kappa, \tau\colon [0, L] \to \mathbb{R}$ (with κ everywhere positive), is there a space curve with those as its curvature and torsion? The answer is yes, and this is an immediate consequence of the fundamental existence theorem for differential equations, Theorem 3.1 of the Appendix. That is, we let

$$F(s) = \begin{bmatrix} | & | & | \\ \mathbf{T}(s) & \mathbf{N}(s) & \mathbf{B}(s) \\ | & | & | \end{bmatrix} \quad \text{and} \quad K(s) = \begin{bmatrix} 0 & -\kappa(s) & 0 \\ \kappa(s) & 0 & -\tau(s) \\ 0 & \tau(s) & 0 \end{bmatrix}.$$

Then integrating the linear system of ordinary differential equations $F'(s) = F(s)K(s)$, $F(0) = F_0$, gives us the Frenet frame everywhere along the curve, and we recover $\boldsymbol{\alpha}$ by integrating $\mathbf{T}(s)$.

We turn now to the concept of *total curvature* of a *closed* space curve, which is the integral of its curvature. That is, if $\boldsymbol{\alpha}\colon [0, L] \to \mathbb{R}^3$ is an arclength-parametrized curve with $\boldsymbol{\alpha}(0) = \boldsymbol{\alpha}(L)$, $\boldsymbol{\alpha}'(0) = \boldsymbol{\alpha}'(L)$, and $\boldsymbol{\alpha}''(0) = \boldsymbol{\alpha}''(L)$, then its total curvature is $\displaystyle\int_0^L \kappa(s)\,ds$. This quantity can be interpreted geometrically as follows: The *Gauss map* of $\boldsymbol{\alpha}$ is the map to the unit sphere, Σ, given by the unit tangent vector $\mathbf{T}\colon [0, L] \to \Sigma$; its image, Γ, is classically called the *tangent indicatrix* of $\boldsymbol{\alpha}$. Observe that—provided the Gauss map is one-

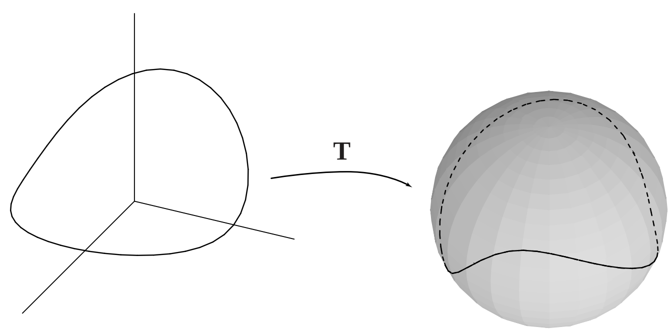


FIGURE 3.1

to-one—the length of Γ is the total curvature of $\boldsymbol{\alpha}$, since $\text{length}(\Gamma) = \displaystyle\int_0^L \|\mathbf{T}'(s)\|\,ds = \int_0^L \kappa(s)\,ds$. More generally, this integral is the length of Γ "counting multiplicities."

A preliminary question to ask is this: What curves Γ in the unit sphere can be the Gauss map of some *closed* space curve $\boldsymbol{\alpha}$? Since $\boldsymbol{\alpha}(s) = \boldsymbol{\alpha}(0) + \displaystyle\int_0^s \mathbf{T}(u)\,du$, we see that a necessary and sufficient condition is that $\displaystyle\int_0^L \mathbf{T}(s)\,ds = \mathbf{0}$. (Note, however, that this depends on the arclength parametrization of the original curve and is not a parametrization-independent condition on the image curve $\Gamma \subset \Sigma$.) We do, nevertheless, have the following geometric consequence of this condition. For any (unit) vector $\mathbf{A}$, we have

$$0 = \mathbf{A} \cdot \int_0^L \mathbf{T}(s)ds = \int_0^L (\mathbf{T}(s) \cdot \mathbf{A})ds,$$

and so the average value of $\mathbf{T} \cdot \mathbf{A}$ must be 0. In particular, the tangent indicatrix must cross the great circle with normal vector $\mathbf{A}$. That is, if the curve Γ is to be a tangent indicatrix, it must be "balanced" with respect to every direction $\mathbf{A}$. It is natural to ask for the shortest curve(s) with this property.

If $\boldsymbol{\xi} \in \Sigma$, let $\boldsymbol{\xi}^\perp$ denote the oriented great circle with normal vector $\boldsymbol{\xi}$. (By this we mean that we go around the circle $\boldsymbol{\xi}^\perp$ so that at $\mathbf{x}$, the tangent vector $\mathbf{T}$ points so that $\mathbf{x}, \mathbf{T}, \boldsymbol{\xi}$ form a right-handed basis for $\mathbb{R}^3$.)

Proposition 3.2 (Crofton's formula)**.** *Let Γ be a piecewise-$\mathcal{C}^1$ curve on the sphere. Then*

$$\begin{aligned} \text{length}(\Gamma) &= \frac{1}{4}\int_\Sigma \#(\Gamma \cap \boldsymbol{\xi}^\perp)d\boldsymbol{\xi} \\ &= \pi \times (\textit{the average number of intersections of } \Gamma \textit{ with all great circles}). \end{aligned}$$

(Here $d\boldsymbol{\xi}$ represents the usual element of surface area on Σ.)

Proof. We leave this to the reader in Exercise 11. □

Remark. Although we don't stop to justify it here, the set of $\boldsymbol{\xi}$ for which $\#(\Gamma \cap \boldsymbol{\xi}^\perp)$ is infinite is a set of measure zero, and so the integral makes sense.

Applying this to the case of the tangent indicatrix of a closed space curve, we deduce the following classical result.

Theorem 3.3 (Fenchel)**.** *The total curvature of any closed space curve is at least 2π, and equality holds if and only if the curve is a (convex) planar curve.*

Proof. Let Γ be the tangent indicatrix of our space curve. If C is a closed plane curve, then Γ is a great circle on the sphere. As we shall see in the next section, convexity of the curve can be interpreted as saying $\kappa > 0$ everywhere, so the tangent indicatrix traverses the great circle exactly once and $\int_C \kappa ds = 2\pi$ (cf. Theorem 3.5 in the next section).

To prove the converse, note that, by our earlier remarks, Γ must cross $\boldsymbol{\xi}^\perp$ for almost every $\boldsymbol{\xi} \in \Sigma$ and hence must intersect it at least twice, and so it follows from Proposition 3.2 that $\int_C \kappa ds = \text{length}(\Gamma) \geq \frac{1}{4}(2)(4\pi) = 2\pi$. Now, we claim that if Γ is a connected, closed curve in Σ of length $\leq 2\pi$, then Γ lies in a closed hemisphere. It will follow, then, that if Γ is a tangent indicatrix of length 2π, it must be a great circle. (For if Γ lies in the hemisphere $\mathbf{A} \cdot \mathbf{x} \geq 0$, $\int_0^L \mathbf{T}(s) \cdot \mathbf{A} ds = 0$ forces $\mathbf{T} \cdot \mathbf{A} = 0$, so Γ is the great circle $\mathbf{A} \cdot \mathbf{x} = 0$.) It follows that the curve is planar and the tangent indicatrix traverses the great circle precisely one time, which means that $\kappa > 0$ and the curve is convex. (See the next section for more details on this.)

To prove the claim, we proceed as follows. Suppose $\text{length}(\Gamma) \leq 2\pi$. Choose P and Q in Γ so that the arcs $\Gamma_1 = \overset{\frown}{PQ}$ and $\Gamma_2 = \overset{\frown}{QP}$ have the same length. Choose N bisecting the shorter great circle arc from P to Q, as shown in Figure 3.2. For convenience, we rotate the picture so that N is the north pole of the sphere. Suppose now that the curve Γ_1 were to enter the southern hemisphere; let $\overline{\Gamma}_1$ denote the reflection of Γ_1

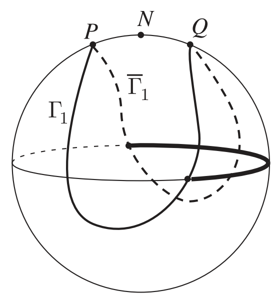


FIGURE 3.2

across the north pole (following arcs of great circle through N). Now, $\Gamma_1 \cup \overline{\Gamma}_1$ is a closed curve containing a pair of antipodal points and therefore is longer than a great circle. (See Exercise 1.) Since $\Gamma_1 \cup \overline{\Gamma}_1$ has the same length as Γ, we see that length$(\Gamma) > 2\pi$, which is a contradiction. Therefore Γ indeed lies in the northern hemisphere. $\square$

We now sketch the proof of a result that has led to many interesting questions in higher dimensions. We say a simple (non-self-intersecting) closed[4] space curve is *knotted* if we cannot fill it in with a disk.

Theorem 3.4 (Fáry-Milnor)**.** *If a simple closed space curve is knotted, then its total curvature is at least* 4π.

Sketch of proof. Suppose the total curvature of C is less than 4π. Then the average number $\#(\Gamma \cap \boldsymbol{\xi}^{\perp}) < 4$. Since this is generically an even number ≥ 2 (whenever the great circle isn't tangent to Γ), there must be an open set of $\boldsymbol{\xi}$'s for which we have $\#(\Gamma \cap \boldsymbol{\xi}^{\perp}) = 2$. Choose one such, $\boldsymbol{\xi}_0$. This means that the tangent vector to C is only perpendicular to $\boldsymbol{\xi}_0$ twice, so the function $f(\mathbf{x}) = \mathbf{x} \cdot \boldsymbol{\xi}_0$ has only two critical points. That is, the planes perpendicular to $\boldsymbol{\xi}_0$ will intersect C either in a single point (at the maximum and minimum points of f) or in exactly two points (by Rolle's Theorem). Now, by moving these planes from the bottom of C to the top, joining the two intersection points in each plane with a line segment, we fill in a disk, so C is unknotted. $\square$

3.2. Plane Curves. We conclude this chapter with some results on plane curves. Now we assign a sign to the curvature: Given an arclength-parametrized curve $\boldsymbol{\alpha}$, (re)define $\mathbf{N}(s)$ so that $\{\mathbf{T}(s), \mathbf{N}(s)\}$ is a right-handed basis for $\mathbb{R}^2$ (i.e., one turns counterclockwise from $\mathbf{T}(s)$ to $\mathbf{N}(s)$), and then set $\kappa(s) = \mathbf{T}'(s) \cdot \mathbf{N}(s)$, from which it follows that $\mathbf{T}'(s) = \kappa(s)\mathbf{N}(s)$ (why?), as before. So $\kappa > 0$ when $\mathbf{T}$ is twisting counterclockwise and $\kappa < 0$ when $\mathbf{T}$ is twisting clockwise. Although the total curvature $\displaystyle\int_C |\kappa(s)|ds$ of a simple closed plane curve may be quite a bit larger than 2π, it is intuitively plausible that the tangent vector must make precisely one full rotation, either counterclockwise or clockwise, and thus we have

Theorem 3.5 (Hopf Umlaufsatz)**.** *If C is a simple closed plane curve, then* $\displaystyle\int_C \kappa\, ds = \pm 2\pi$*, the $+$ occurring when C is oriented counterclockwise and $-$ when it's oriented clockwise.*

The crucial ingredient is to keep track of a *continuous* total angle through which the tangent vector has turned. That is, we need the following

---

[4]To be more careful here, if $\boldsymbol{\alpha}\colon [a,b] \to \mathbb{R}^3$ is a parametrization with $\boldsymbol{\alpha}(a) = \boldsymbol{\alpha}(b)$, then $\boldsymbol{\alpha}(t) = \boldsymbol{\alpha}(u)$ occurs *only* when $\{t,u\} = \{a,b\}$.

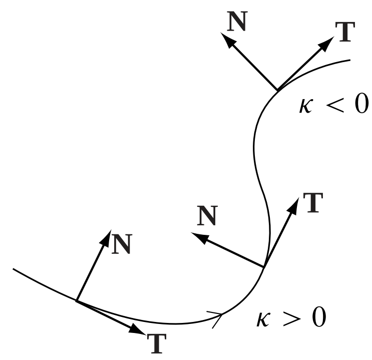


FIGURE 3.3

Lemma 3.6. *Let $\boldsymbol{\alpha}\colon [a,b] \to \mathbb{R}^2$ be a $\mathcal{C}^1$, regular parametrized plane curve. Then there is a $\mathcal{C}^1$ function $\theta\colon [a,b] \to \mathbb{R}$ so that $\mathbf{T}(t) = \big(\cos\theta(t), \sin\theta(t)\big)$ for all $t \in [a,b]$. Moreover, for any two such functions, θ and θ^*, we have $\theta(b) - \theta(a) = \theta^*(b) - \theta^*(a)$. The number $(\theta(b) - \theta(a))/2\pi$ is called the* rotation index *of $\boldsymbol{\alpha}$.*

Proof. Consider the four open semicircles $U_1 = \{(x,y) \in S^1 : x > 0\}$, $U_2 = \{(x,y) \in S^1 : x < 0\}$, $U_3 = \{(x,y) \in S^1 : y > 0\}$, and $U_4 = \{(x,y) \in S^1 : y < 0\}$. Then the functions

$$\begin{aligned}
\psi_{1,n}(x,y) &= \arctan(y/x) + 2n\pi \\
\psi_{2,n}(x,y) &= \arctan(y/x) + (2n+1)\pi \\
\psi_{3,n}(x,y) &= -\arctan(x/y) + (2n + \tfrac{1}{2})\pi \\
\psi_{4,n}(x,y) &= -\arctan(x/y) + (2n - \tfrac{1}{2})\pi
\end{aligned}$$

are smooth maps $\psi_{i,n}\colon U_i \to \mathbb{R}$ with the property that $\big(\cos(\psi_{i,n}(x,y)), \sin(\psi_{i,n}(x,y))\big) = (x,y)$ for every $i = 1,2,3,4$ and $n \in \mathbb{Z}$.

Define $\theta(a)$ so that $\mathbf{T}(a) = \big(\cos\theta(a), \sin\theta(a)\big)$. Let $S = \{t \in [a,b] : \theta$ is defined and $\mathcal{C}^1$ on $[a,t]\}$, and let $t_0 = \sup S$. Suppose first that $t_0 < b$. Choose i so that $\mathbf{T}(t_0) \in U_i$, and choose $n \in \mathbb{Z}$ so that $\psi_{i,n}(\mathbf{T}(t_0)) = \lim_{t \to t_0^-} \theta(t)$. Because $\mathbf{T}$ is continuous at t_0, there is $\delta > 0$ so that $\mathbf{T}(t) \in U_i$ for all t with $|t - t_0| < \delta$. Then setting $\theta(t) = \psi_{i,n}(\mathbf{T}(t))$ for all $t_0 \le t < t_0 + \delta$ gives us a $\mathcal{C}^1$ function θ defined on $[0, t_0 + \delta/2]$, so we cannot have $t_0 < b$. (Note that $\theta(t) = \psi_{i,n}(\mathbf{T}(t))$ for all $t_0 - \delta < t < t_0$. Why?) But the same argument shows that when $t_0 = b$, the function θ is $\mathcal{C}^1$ on all of $[a,b]$.

Now, since $\mathbf{T}(b) = \mathbf{T}(a)$, we know that $\theta(b) - \theta(a)$ must be an integral multiple of 2π. Moreover, for any other function θ^* with the same properties, we have $\theta^*(t) = \theta(t) + 2\pi n(t)$ for some integer $n(t)$. Since θ and θ^* are both continuous, n must be a continuous function as well; since it takes on only integer values, it must be a constant function. Therefore, $\theta^*(b) - \theta^*(a) = \theta(b) - \theta(a)$, as required. $\square$

Sketch of proof of Theorem 3.5. Note first that if $\mathbf{T}(s) = \big(\cos\theta(s), \sin\theta(s)\big)$, then $\mathbf{T}'(s) = \theta'(s)\big(-\sin\theta(s), \cos\theta(s)\big)$, so $\kappa(s) = \theta'(s)$ and $\displaystyle\int_0^L \kappa(s)ds = \int_0^L \theta'(s)ds = \theta(L) - \theta(0)$ is 2π times the rotation index of the closed curve $\boldsymbol{\alpha}$.

Let $\Delta = \{(s,t) : 0 \le s \le t \le L\}$. Consider the secant map $\mathbf{h}\colon \Delta \to S^1$ defined by

$$\mathbf{h}(s,t) = \begin{cases} \mathbf{T}(s), & s = t \\ -\mathbf{T}(0), & (s,t) = (0,L) \\ \dfrac{\boldsymbol{\alpha}(t) - \boldsymbol{\alpha}(s)}{\|\boldsymbol{\alpha}(t) - \boldsymbol{\alpha}(s)\|}, & \text{otherwise} \end{cases} .$$

Then it follows from Proposition 2.6 (using Taylor's Theorem to calculate $\boldsymbol{\alpha}(t) = \boldsymbol{\alpha}(s) + (t-s)\boldsymbol{\alpha}'(s) + \dots$) that $\mathbf{h}$ is continuous. A more sophisticated version of the proof of Lemma 3.6 will establish (see Exercise 13) that there is a continuous function $\tilde{\theta}\colon \Delta \to \mathbb{R}$ so that $\mathbf{h}(s,t) = \big(\cos\tilde{\theta}(s,t), \sin\tilde{\theta}(s,t)\big)$ for all $(s,t) \in \Delta$. It then follows from Lemma 3.6 that

$$\int_C \kappa\, ds = \theta(L) - \theta(0) = \tilde{\theta}(L,L) - \tilde{\theta}(0,0) = \underbrace{\tilde{\theta}(0,L) - \tilde{\theta}(0,0)}_{N_1} + \underbrace{\tilde{\theta}(L,L) - \tilde{\theta}(0,L)}_{N_2}.$$

Rotating the curve as required, we assume that $\boldsymbol{\alpha}(0)$ is the lowest point on the curve (i.e., the one whose y-coordinate is smallest) and, then, that $\boldsymbol{\alpha}(0)$ is the origin and $\mathbf{T}(0) = \mathbf{e}_1$, as shown in Figure 3.4. (The

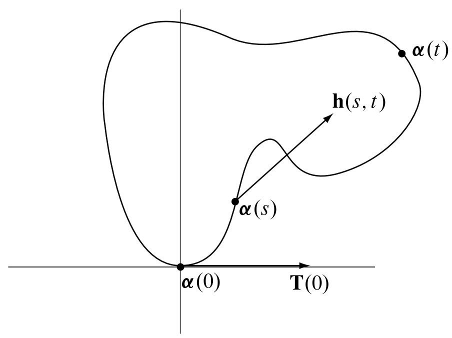


FIGURE 3.4

last may require reversing the orientation of the curve.) Now, N_1 is the angle through which the position vector of the curve turns, starting at 0 and ending at π; since the curve lies in the upper half-plane, we must have $N_1 = \pi$. But N_2 is likewise the angle through which the negative of the position vector turns, so $N_2 = N_1 = \pi$. With these assumptions, we see that the rotation index of the curve is 1. Allowing for the possible change in orientation, the rotation index must therefore be ± 1, as required. ☐

Corollary 3.7. *If C is any closed curve with nonzero rotation index (e.g., a simple closed curve), for any point $P \in C$ there is a point $Q \in C$ where the unit tangent vector is opposite that at P.*

Proof. Let $\mathbf{T}(s) = \big(\cos\theta(s), \sin\theta(s)\big)$ for a $\mathcal{C}^1$ function $\theta\colon [0,L] \to \mathbb{R}$, as in Lemma 3.6. Say $P = \boldsymbol{\alpha}(s_0)$, and let $\theta(s_0) = \theta_0$. Since $\theta(L) - \theta(0)$ is an integer multiple of 2π, there must be $s_1 \in [0,L]$ with either $\theta(s_1) = \theta_0 + \pi$ or $\theta(s_1) = \theta_0 - \pi$. Take $Q = \boldsymbol{\alpha}(s_1)$. ☐

Recall that one of the ways of characterizing a convex function $f\colon \mathbb{R} \to \mathbb{R}$ is that its graph lie on one side of each of its tangent lines. So we make the following

Definition. The regular closed plane curve $\boldsymbol{\alpha}$ is *convex* if it lies on one side of its tangent line at each point.

Proposition 3.8. *A simple closed regular plane curve C is convex if and only if we can choose the orientation of the curve so that $\kappa \geq 0$ everywhere.*

Remark. We leave it to the reader in Exercise 2 to give a non-simple closed curve for which this result is false.

Proof. Assume, without loss of generality, that $\mathbf{T}(0) = (1, 0)$ and the curve is oriented counterclockwise. Using the function θ constructed in Lemma 3.6, the condition that $\kappa \geq 0$ is equivalent to the condition that θ is a nondecreasing function with $\theta(L) = 2\pi$.

Suppose first that θ is nondecreasing and C is not convex. Then we can find a point $P = \boldsymbol{\alpha}(s_0)$ on the curve and values s_1', s_2' so that $\boldsymbol{\alpha}(s_1')$ and $\boldsymbol{\alpha}(s_2')$ lie on opposite sides of the tangent line to C at P. Then, by the maximum value theorem, there are values s_1 and s_2 so that $\boldsymbol{\alpha}(s_1)$ is the greatest distance "above" the tangent line and $\boldsymbol{\alpha}(s_2)$ is the greatest distance "below." Consider the unit tangent vectors $\mathbf{T}(s_0)$, $\mathbf{T}(s_1)$, and $\mathbf{T}(s_2)$. Since these vectors are either parallel or anti-parallel, some pair must be identical. Letting the respective values of s be s^* and s^{**} with $s^* < s^{**}$, we have $\theta(s^*) = \theta(s^{**})$ (since θ is nondecreasing and $\theta(L) = 2\pi$, the values cannot differ by a multiple of 2π), and therefore $\theta(s) = \theta(s^*)$ for all $s \in [s^*, s^{**}]$. This means that that portion of C between $\boldsymbol{\alpha}(s^*)$ and $\boldsymbol{\alpha}(s^{**})$ is a line segment parallel to the tangent line of C at P; this is a contradiction.

Conversely, suppose C is convex and $\theta(s_1) = \theta(s_2)$ for some $s_1 < s_2$. By Corollary 3.7 there must be s_3 with $\mathbf{T}(s_3) = -\mathbf{T}(s_1) = -\mathbf{T}(s_2)$. Since C is convex, the tangent line at two of $\boldsymbol{\alpha}(s_1)$, $\boldsymbol{\alpha}(s_2)$, and $\boldsymbol{\alpha}(s_3)$ must be the same, say at $\boldsymbol{\alpha}(s^*) = P$ and $\boldsymbol{\alpha}(s^{**}) = Q$. If $\overline{PQ}$ does not lie entirely in C, choose $R \in \overline{PQ}$, $R \notin C$. Since C is convex, the line through R perpendicular to $\overleftrightarrow{PQ}$ must intersect C in at least two points, say M and N, with N farther from $\overleftrightarrow{PQ}$ than M. Since M lies in the interior of $\triangle NPQ$, all three vertices of the triangle can never lie on the same side of any line through M. In particular, N, P, and Q cannot lie on the same side of the tangent line to C at M. Thus, it must be that $\overline{PQ} \subset C$, so $\theta(s) = \theta(s_1) = \theta(s_2)$ for all $s \in [s_1, s_2]$. Therefore, θ is nondecreasing, and we are done. $\square$

Definition. A critical point of κ is called a *vertex* of the curve C.

A closed curve must have at least two vertices: the maximum and minimum points of κ. Every point of a circle is a vertex. We conclude with the following

Proposition 3.9 (Four Vertex Theorem)**.** *A closed convex plane curve has at least four vertices.*

Proof. Suppose that C has fewer than four vertices. As we see from Figure 3.5, either κ must have two critical points (maximum and minimum) or κ must have three critical points (maximum, minimum, and inflection point). More precisely, suppose that κ increases from P to Q and decreases from Q to P. Without loss of generality, we may take P to be at the origin. The equation of $\overleftrightarrow{PQ}$ is $\mathbf{A} \cdot \mathbf{x} = 0$, where we choose $\mathbf{A}$ so that $\kappa'(s) \geq 0$ precisely when $\mathbf{A} \cdot \boldsymbol{\alpha}(s) \geq 0$. Then $\displaystyle\int_C \kappa'(s)(\mathbf{A} \cdot \boldsymbol{\alpha}(s))ds > 0$. Integrating by parts, we have

$$\int_C \kappa'(s)(\mathbf{A} \cdot \boldsymbol{\alpha}(s))ds = -\int_C \kappa(s)(\mathbf{A} \cdot \mathbf{T}(s))ds = \int_C \mathbf{A} \cdot \mathbf{N}'(s)ds = \mathbf{A} \cdot \int_C \mathbf{N}'(s)ds = 0.$$

From this contradiction, we infer that C must have at least four vertices. $\square$

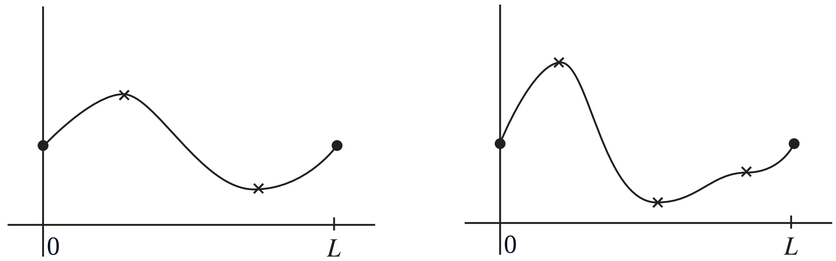


FIGURE 3.5

3.3. The Isoperimetric Inequality. One of the classic questions in mathematics is the following: Given a closed curve of length L, what shape will enclose the most area? A little experimentation will most likely lead the reader to the

Theorem 3.10 (Isoperimetric Inequality)**.** *If a simple closed plane curve C has length L and encloses area A, then*

$$L^2 \geq 4\pi A,$$

and equality holds if and only if C is a circle.

Proof. There are a number of different proofs, but we give one (due to E. Schmidt, 1939) based on Green's Theorem, Theorem 2.6 of the Appendix, and—not surprisingly—relying heavily on the geometric-arithmetic mean inequality and the Cauchy-Schwarz inequality (see Exercise A.1.2). We choose parallel

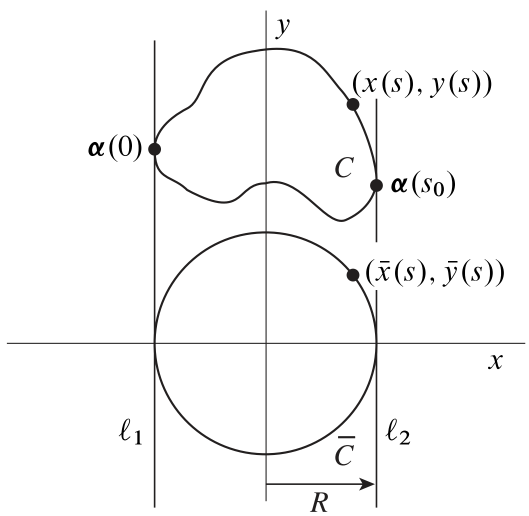


FIGURE 3.6

lines ℓ_1 and ℓ_2 tangent to, and enclosing, C, as pictured in Figure 3.6. We draw a circle $\overline{C}$ of radius R with those same tangent lines and put the origin at its center, with the y-axis parallel to ℓ_i. We now parametrize C by arclength by $\boldsymbol{\alpha}(s) = (x(s), y(s))$, $s \in [0, L]$, taking $\boldsymbol{\alpha}(0) \in \ell_1$ and $\boldsymbol{\alpha}(s_0) \in \ell_2$. We then consider $\overline{\boldsymbol{\alpha}}\colon [0, L] \to \mathbb{R}^2$ given by

$$\overline{\boldsymbol{\alpha}}(s) = \big(\overline{x}(s), \overline{y}(s)\big) = \begin{cases} \big(x(s), -\sqrt{R^2 - x(s)^2}\big), & 0 \leq s \leq s_0 \\ \big(x(s), \sqrt{R^2 - x(s)^2}\big), & s_0 \leq s \leq L \end{cases}.$$

($\overline{\boldsymbol{\alpha}}$ needn't be a parametrization of the circle $\overline{C}$, since it may cover certain portions multiple times, but that's no problem.) Letting A denote the area enclosed by C and $\overline{A} = \pi R^2$ that enclosed by $\overline{C}$, we have (by Exercise A.2.5)

$$A = \int_0^L x(s)y'(s)ds$$
$$\overline{A} = \pi R^2 = -\int_0^L \overline{y}(s)\overline{x}'(s)ds = -\int_0^L \overline{y}(s)x'(s)ds.$$

Adding these equations and applying the Cauchy-Schwarz inequality, we have

$$A + \pi R^2 = \int_0^L \big(x(s)y'(s) - \overline{y}(s)x'(s)\big)ds = \int_0^L \big(x(s), \overline{y}(s)\big) \cdot \big(y'(s), -x'(s)\big)ds$$
$$(*) \qquad \le \int_0^L \|\big(x(s), \overline{y}(s)\big)\| \|\big(y'(s), -x'(s)\big)\| ds = RL,$$

inasmuch as $\|(y'(s), -x'(s))\| = \|(x'(s), y'(s))\| = 1$ since $\boldsymbol{\alpha}$ is arclength-parametrized. We now recall the arithmetic-geometric mean inequality:

$$\sqrt{ab} \le \frac{a+b}{2} \qquad \text{for positive numbers } a \text{ and } b,$$

with equality holding if and only if $a = b$. We therefore have

$$\sqrt{A}\sqrt{\pi R^2} \le \frac{A + \pi R^2}{2} \le \frac{RL}{2},$$

so $4\pi A \le L^2$.

Now suppose equality holds here. Then we must have $A = \pi R^2$ and $L = 2\pi R$. It follows that the curve C has the same breadth in all directions (since L now determines R). But equality must also hold in $(*)$, so the vectors $\overline{\boldsymbol{\alpha}}(s) = \big(x(s), \overline{y}(s)\big)$ and $\big(y'(s), -x'(s)\big)$ must be everywhere parallel. Since the first vector has length R and the second has length 1, we infer that

$$\big(x(s), \overline{y}(s)\big) = R\big(y'(s), -x'(s)\big),$$

and so $x(s) = Ry'(s)$. By our remark at the beginning of this paragraph, the same result will hold if we rotate the axes $\pi/2$; let $y = y_0$ be the line halfway between the enclosing *horizontal* lines ℓ_i. Now, substituting $y - y_0$ for x and $-x$ for y, so we have $y(s) - y_0 = -Rx'(s)$, as well. Therefore, $x(s)^2 + \big(y(s) - y_0\big)^2 = R^2(x'(s)^2 + y'(s)^2) = R^2$, and C is indeed a circle of radius R. $\square$

EXERCISES 1.3

1. a. Prove that the shortest path between two points on the unit sphere is an arc of a great circle connecting them. (Hint: Without loss of generality, take one point to be $(0,0,1)$ and the other to be $(\sin u_0, 0, \cos u_0)$. Let $\boldsymbol{\alpha}(t) = (\sin u(t)\cos v(t), \sin u(t)\sin v(t), \cos u(t))$, $a \le t \le b$, be an arbitrary curve with $u(a) = 0$, $v(a) = 0$, $u(b) = u_0$, $v(b) = 0$, calculate the arclength of $\boldsymbol{\alpha}$, and show that it is smallest when $v(t) = 0$ for all t.)

   b. Prove that if P and Q are points on the unit sphere, then the shortest path between them has length $\arccos(P \cdot Q)$.

2. Give a closed plane curve C with $\kappa > 0$ that is not convex.

3. Draw closed plane curves with rotation indices 0, 2, -2, and 3, respectively.

*4. Suppose C is a simple closed plane curve with $0 < \kappa \le c$. Prove that length$(C) \ge 2\pi/c$.

5. Give an alternative proof of the latter part of Theorem 3.1 by considering instead the function
$$f(s) = \|\tilde{\mathbf{T}}(s) - \mathbf{T}^*(s)\|^2 + \|\tilde{\mathbf{N}}(s) - \mathbf{N}^*(s)\|^2 + \|\tilde{\mathbf{B}}(s) - \mathbf{B}^*(s)\|^2.$$

6. (See Exercise 1.2.15.) Prove that if C is a simple closed (convex) plane curve of constant breadth μ, then length$(C) = \pi\mu$.

7. Suppose C is a convex simple closed plane curve with maximum curvature κ_0. Prove that the distance between any pair of parallel tangent lines of C is at least $2/\kappa_0$.

8. A convex plane curve with the origin in its interior can be determined by its tangent lines $(\cos\theta)x + (\sin\theta)y = p(\theta)$, called its *support lines*, as shown in Figure 3.7. The function $p(\theta)$ is called the support function. (Here θ is the polar coordinate, and we assume $p(\theta) > 0$ for all $\theta \in [0, 2\pi]$.)

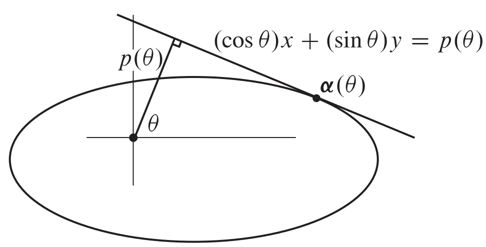


FIGURE 3.7

   a. Prove that the line given above is tangent to the curve at the point $\boldsymbol{\alpha}(\theta) = (p(\theta)\cos\theta - p'(\theta)\sin\theta,\, p(\theta)\sin\theta + p'(\theta)\cos\theta)$.
   b. Prove that the curvature of the curve at $\boldsymbol{\alpha}(\theta)$ is $1\big/\big(p(\theta) + p''(\theta)\big)$.
   c. Prove that the length of $\boldsymbol{\alpha}$ is given by $L = \displaystyle\int_0^{2\pi} p(\theta)\,d\theta$.
   d. Prove that the area enclosed by $\boldsymbol{\alpha}$ is given by $A = \dfrac{1}{2}\displaystyle\int_0^{2\pi} \big(p(\theta)^2 - p'(\theta)^2\big)\,d\theta$.
   e. Use the answer to part c to reprove the result of Exercise 6.

9. Let C be a $\mathcal{C}^2$ closed space curve, say parametrized by arclength by $\boldsymbol{\alpha}\colon [0, L] \to \mathbb{R}^3$. A unit *normal field* $\mathbf{X}$ on C is a $\mathcal{C}^1$ vector-valued function with $\mathbf{X}(0) = \mathbf{X}(L)$ and $\mathbf{X}(s)\cdot\mathbf{T}(s) = 0$ and $\|\mathbf{X}(s)\| = 1$ for all s. We define the *twist* of $\mathbf{X}$ to be
$$\mathrm{tw}(C, \mathbf{X}) = \frac{1}{2\pi}\int_0^L \mathbf{X}'(s)\cdot(\mathbf{T}(s)\times\mathbf{X}(s))\,ds.$$
   a. Show that if $\mathbf{X}$ and $\mathbf{X}^*$ are two unit normal fields on C, then $\mathrm{tw}(C, \mathbf{X})$ and $\mathrm{tw}(C, \mathbf{X}^*)$ differ by an integer. The fractional part of $\mathrm{tw}(C, \mathbf{X})$ (i.e., the twist mod 1) is called the *total twist* of C. (Hint: Write $\mathbf{X}(s) = \cos\theta(s)\mathbf{N}(s) + \sin\theta(s)\mathbf{B}(s)$.)
   b. Prove that the total twist of C equals the fractional part of $\dfrac{1}{2\pi}\displaystyle\int_0^L \tau\,ds$.

c. Prove that if a closed curve lies on a sphere, then its total twist is 0. (Hint: Choose an obvious candidate for **X**.)

Remark. W. Scherrer proved in 1940 that if the total twist of every closed curve on a surface is 0, then that surface must be a (subset of a) plane or sphere.

10. (See Exercise 1.2.24.) Under what circumstances does a closed space curve have a parallel curve that is also closed? (Hint: Exercise 8 should be relevant.)

11. (*The Bishop Frame*) Suppose $\boldsymbol{\alpha}$ is an arclength-parametrized $\mathcal{C}^2$ curve. Suppose we have $\mathcal{C}^1$ unit vector fields $\mathbf{N}_1$ and $\mathbf{N}_2 = \mathbf{T} \times \mathbf{N}_1$ along $\boldsymbol{\alpha}$ so that

$$\mathbf{T} \cdot \mathbf{N}_1 = \mathbf{T} \cdot \mathbf{N}_2 = \mathbf{N}_1 \cdot \mathbf{N}_2 = 0;$$

i.e., $\mathbf{T}, \mathbf{N}_1, \mathbf{N}_2$ will be a smoothly varying right-handed orthonormal frame as we move along the curve. (To this point, the Frenet frame would work just fine if the curve were $\mathcal{C}^3$ with $\kappa \neq 0$.) But now we want to impose the extra condition that $\mathbf{N}_1' \cdot \mathbf{N}_2 = 0$. We say the unit normal vector field $\mathbf{N}_1$ is *parallel* along $\boldsymbol{\alpha}$; this means that the only change of $\mathbf{N}_1$ is in the direction of $\mathbf{T}$. In this event, $\mathbf{T}, \mathbf{N}_1, \mathbf{N}_2$ is called a *Bishop frame* for $\boldsymbol{\alpha}$. A Bishop frame can be defined even when a Frenet frame cannot (e.g., when there are points with $\kappa = 0$).

a. Show that there are functions k_1 and k_2 so that

$$\begin{aligned} \mathbf{T}' &= & k_1\mathbf{N}_1 + k_2\mathbf{N}_2 \\ \mathbf{N}_1' &= -k_1\mathbf{T} & \\ \mathbf{N}_2' &= -k_2\mathbf{T} & \end{aligned}$$

b. Show that $\kappa^2 = k_1^2 + k_2^2$.

c. Show that if $\boldsymbol{\alpha}$ is $\mathcal{C}^3$ with $\kappa \neq 0$, then we can take $\mathbf{N}_1 = (\cos\theta)\mathbf{N} + (\sin\theta)\mathbf{B}$, where $\theta' = -\tau$. Check that $k_1 = \kappa\cos\theta$ and $k_2 = -\kappa\sin\theta$.

d. Show that $\boldsymbol{\alpha}$ lies on the surface of a sphere if and only if there are constants λ, μ so that $\lambda k_1 + \mu k_2 + 1 = 0$; moreover, if $\boldsymbol{\alpha}$ lies on a sphere of radius R, then $\lambda^2 + \mu^2 = R^2$. (Cf. Exercise 1.2.19.)

e. What condition is required to define a Bishop frame globally on a closed curve? (See Exercise 8.) How is this question related to Exercise 1.2.24?

12. Prove Proposition 3.2 as follows. Let $\boldsymbol{\alpha}\colon [0, L] \to \Sigma$ be the arclength parametrization of Γ, and define $\mathbf{F}\colon [0, L] \times [0, 2\pi) \to \Sigma$ by $\mathbf{F}(s, \phi) = \boldsymbol{\xi}$, where $\boldsymbol{\xi}^\perp$ is the great circle making angle ϕ with Γ at $\boldsymbol{\alpha}(s)$. Check that $\mathbf{F}$ takes on the value $\boldsymbol{\xi}$ precisely $\#(\Gamma \cap \boldsymbol{\xi}^\perp)$ times, so that $\mathbf{F}$ is a "multi-parametrization" of Σ that gives us

$$\int_\Sigma \#(\Gamma \cap \boldsymbol{\xi}^\perp)\, d\boldsymbol{\xi} = \int_0^L \int_0^{2\pi} \left\| \frac{\partial \mathbf{F}}{\partial s} \times \frac{\partial \mathbf{F}}{\partial \phi} \right\| d\phi\, ds.$$

Compute that $\left\| \dfrac{\partial \mathbf{F}}{\partial s} \times \dfrac{\partial \mathbf{F}}{\partial \phi} \right\| = |\sin\phi|$ (this is the hard part) and finish the proof. (Hints: As pictured in Figure 3.8, show $\mathbf{v}(s, \phi) = \cos\phi\,\mathbf{T}(s) + \sin\phi\,(\boldsymbol{\alpha}(s) \times \mathbf{T}(s))$ is the tangent vector to the great circle $\boldsymbol{\xi}^\perp$ and deduce that $\mathbf{F}(s, \phi) = \boldsymbol{\alpha}(s) \times \mathbf{v}(s, \phi)$. Show that $\dfrac{\partial \mathbf{F}}{\partial \phi}$ and $\boldsymbol{\alpha} \times \dfrac{\partial \mathbf{v}}{\partial s}$ are both multiples of $\mathbf{v}$.)

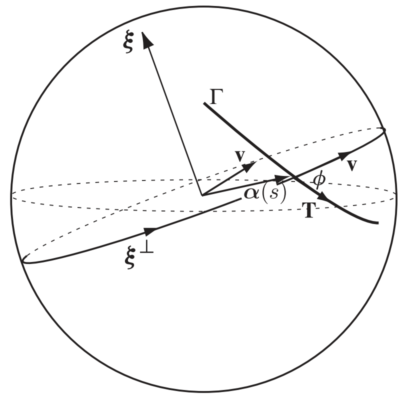


FIGURE 3.8

13. Generalize Theorem 3.5 to prove that if C is a *piecewise-smooth* simple closed plane curve with exterior angles ϵ_j, $j = 1, \dots, \ell$, then $\displaystyle\int_C \kappa\, ds + \sum_{j=1}^{\ell} \epsilon_j = \pm 2\pi$. (As shown in Figure 3.9, the exterior angle ϵ_j

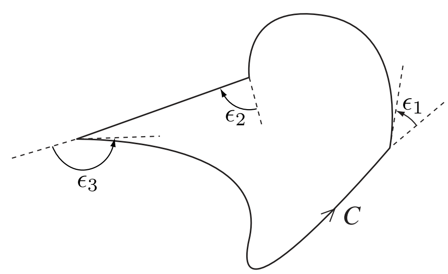


FIGURE 3.9

at $\boldsymbol{\alpha}(s_j)$ is defined to be the angle between $\boldsymbol{\alpha}'_-(s_j) = \lim\limits_{s \to s_j^-} \boldsymbol{\alpha}'(s)$ and $\boldsymbol{\alpha}'_+(s_j) = \lim\limits_{s \to s_j^+} \boldsymbol{\alpha}'(s)$, with the convention that $|\epsilon_j| \le \pi$.)

14. Complete the details of the proof of the indicated step in the proof of Theorem 3.5, as follows (following H. Hopf's original proof). Pick an interior point $\mathbf{s}_0 \in \Delta$.
    a. Choose $\tilde\theta(\mathbf{s}_0)$ so that $\mathbf{h}(\mathbf{s}_0) = \big(\cos\tilde\theta(\mathbf{s}_0), \sin\tilde\theta(\mathbf{s}_0)\big)$. Use Lemma 3.6, slightly modified, to determine $\tilde\theta$ uniquely as a function that is continuous on each ray $\overrightarrow{\mathbf{s}_0\mathbf{s}}$ for every $\mathbf{s} \in \Delta$.
    b. Since a continuous function on a compact (closed and bounded) set $\Delta \subset \mathbb{R}^2$ is uniformly continuous, given any $\varepsilon_0 > 0$, there is a number $\delta_0 > 0$ so that whenever $\mathbf{s}, \mathbf{s}' \in \Delta$ and $\|\mathbf{s} - \mathbf{s}'\| < \delta_0$, we will have $\|\mathbf{h}(\mathbf{s}) - \mathbf{h}(\mathbf{s}')\| < \varepsilon_0$. In particular, show that there is δ_0 so that whenever $\mathbf{s}, \mathbf{s}' \in \Delta$ and $\|\mathbf{s} - \mathbf{s}'\| < \delta_0$, the angle between the vectors $\mathbf{h}(\mathbf{s})$ and $\mathbf{h}(\mathbf{s}')$ is less than π.
    c. Consider the triangle formed by two radii of the unit circle making angle θ. Give an upper bound on θ in terms of the chord length ℓ. Using this, deduce that given $\varepsilon > 0$, there is $0 < \delta < \delta_0$ so that whenever $\|\mathbf{s} - \mathbf{s}'\| < \delta$, we have $|\tilde\theta(\mathbf{s}) - \tilde\theta(\mathbf{s}') + 2\pi n(\mathbf{s})| < \varepsilon$ for some integer $n(\mathbf{s})$.
    d. Now choose $\mathbf{s}' = \mathbf{s}_1 \in \Delta$ arbitrary. Consider the function $f(u) = \tilde\theta(\mathbf{s}_0 + u(\mathbf{s} - \mathbf{s}_0)) - \tilde\theta(\mathbf{s}_0 + u(\mathbf{s}_1 - \mathbf{s}_0))$. Show that f is continuous and $f(0) = 0$, and deduce that $|f(1)| < \pi$. Conclude that $n = 0$ in part c and, thus, that $\tilde\theta$ is continuous.

CHAPTER 2

Surfaces: Local Theory

1. Parametrized Surfaces and the First Fundamental Form

Let U be an open set in $\mathbb{R}^2$. A function $\mathbf{f}\colon U \to \mathbb{R}^m$ (for us, $m = 1$ and 3 will be most common) is called $\mathcal{C}^1$ if $\mathbf{f}$ and its partial derivatives $\dfrac{\partial \mathbf{f}}{\partial u}$ and $\dfrac{\partial \mathbf{f}}{\partial v}$ are all continuous. We will ordinarily use (u, v) as coordinates in our parameter space, and (x, y, z) as coordinates in $\mathbb{R}^3$. Similarly, for any $k \geq 2$, we say $\mathbf{f}$ is $\mathcal{C}^k$ if all its partial derivatives of order up to k exist and are continuous. We say $\mathbf{f}$ is *smooth* if $\mathbf{f}$ is $\mathcal{C}^k$ for every positive integer k. We will henceforth assume all our functions are $\mathcal{C}^k$ for $k \geq 3$. One of the crucial results for differential geometry is that if $\mathbf{f}$ is $\mathcal{C}^2$, then $\dfrac{\partial^2 \mathbf{f}}{\partial u \partial v} = \dfrac{\partial^2 \mathbf{f}}{\partial v \partial u}$ (and similarly for higher-order derivatives).

Notation: We will often also use subscripts to indicate partial derivatives, as follows:

$$
\begin{array}{lcc}
\mathbf{f}_u & \leftrightarrow & \dfrac{\partial \mathbf{f}}{\partial u} \\[2ex]
\mathbf{f}_v & \leftrightarrow & \dfrac{\partial \mathbf{f}}{\partial v} \\[2ex]
\mathbf{f}_{uu} & \leftrightarrow & \dfrac{\partial^2 \mathbf{f}}{\partial u^2} \\[2ex]
\mathbf{f}_{uv} = (\mathbf{f}_u)_v & \leftrightarrow & \dfrac{\partial^2 \mathbf{f}}{\partial v \partial u}
\end{array}
$$

Definition. A *regular parametrization* of a subset $M \subset \mathbb{R}^3$ is a ($\mathcal{C}^3$) one-to-one function

$$\mathbf{x}\colon U \to M \subset \mathbb{R}^3 \qquad \text{so that} \qquad \mathbf{x}_u \times \mathbf{x}_v \neq \mathbf{0}$$

for some open set $U \subset \mathbb{R}^2$.[1] A connected subset $M \subset \mathbb{R}^3$ is called a *surface* if each point has a neighborhood that is regularly parametrized.

We might consider the curves on M obtained by fixing $v = v_0$ and varying u, called a u-curve, and obtained by fixing $u = u_0$ and varying v, called a v-curve; these are depicted in Figure 1.1. At the point $P = \mathbf{x}(u_0, v_0)$, we see that $\mathbf{x}_u(u_0, v_0)$ is tangent to the u-curve and $\mathbf{x}_v(u_0, v_0)$ is tangent to the v-curve. We are requiring that these vectors span a plane, whose normal vector is given by $\mathbf{x}_u \times \mathbf{x}_v$.

Example 1. We give some basic examples of parametrized surfaces. Note that our parameters do not necessarily range over an open set of values.

(a) The graph of a function $f\colon U \to \mathbb{R}$, $z = f(x, y)$, is parametrized by $\mathbf{x}(u, v) = (u, v, f(u, v))$. Note that $\mathbf{x}_u \times \mathbf{x}_v = (-f_u, -f_v, 1) \neq \mathbf{0}$, so this is always a regular parametrization.

[1] For technical reasons with which we shall not concern ourselves in this course, we should also require that the inverse function $\mathbf{x}^{-1}\colon \mathbf{x}(U) \to U$ be continuous. We shall also often be sloppy and use subsets U that are not quite open. The interested reader can easily repair things by adding some companion parametrizations.

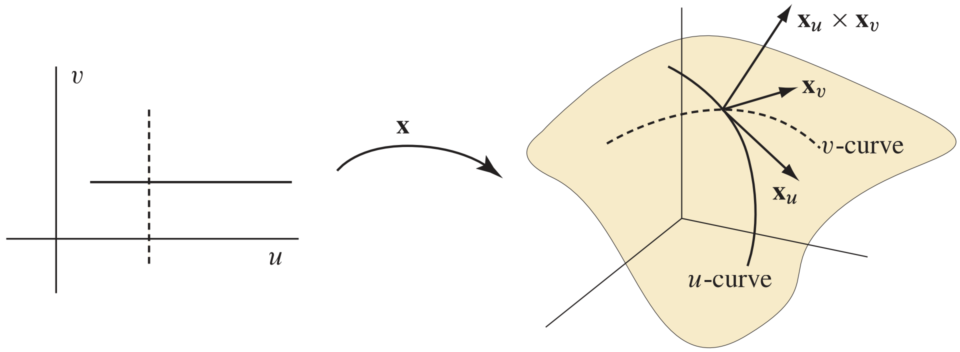


FIGURE 1.1

(b) The *helicoid*, as shown in Figure 1.2, is the surface formed by drawing horizontal rays from the axis

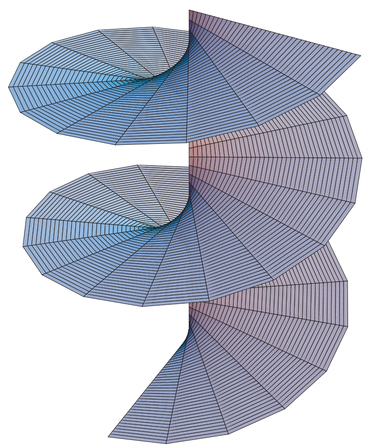


FIGURE 1.2

of the helix $\boldsymbol{\alpha}(t) = (\cos t, \sin t, bt)$ to points on the helix:

$$\mathbf{x}(u, v) = (u \cos v, u \sin v, bv), \qquad u > 0,\ v \in \mathbb{R}.$$

Note that $\mathbf{x}_u \times \mathbf{x}_v = (b \sin v, -b \cos v, u) \neq \mathbf{0}$. The u-curves are rays and the v-curves are helices.

(c) The *torus* (surface of a doughnut) is formed by rotating a circle of radius b about a circle of radius $a > b$ lying in an orthogonal plane, as pictured in Figure 1.3. The regular parametrization is given

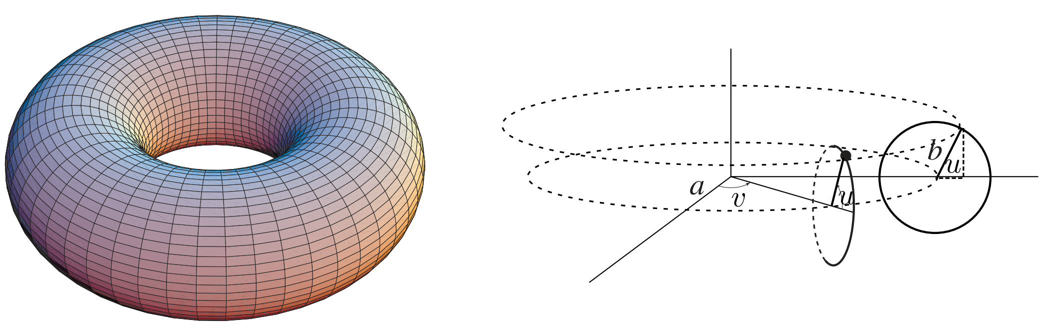


FIGURE 1.3

by

$$\mathbf{x}(u,v) = ((a + b\cos u)\cos v, (a + b\cos u)\sin v, b\sin u), \qquad 0 \le u, v < 2\pi.$$

Then $\mathbf{x}_u \times \mathbf{x}_v = -b(a + b\cos u)\big(\cos u\cos v, \cos u\sin v, \sin u\big)$, which is never $\mathbf{0}$.

(d) The standard parametrization of the unit sphere Σ is given by spherical coordinates $(\phi,\theta) \leftrightarrow (u,v)$:

$$\mathbf{x}(u,v) = (\sin u\cos v, \sin u\sin v, \cos u), \qquad 0 < u < \pi,\ 0 \le v < 2\pi.$$

Since $\mathbf{x}_u \times \mathbf{x}_v = \sin u(\sin u\cos v, \sin u\sin v, \cos u) = (\sin u)\mathbf{x}(u,v)$, the parametrization is regular away from $u = 0, \pi$, which we've excluded anyhow because $\mathbf{x}$ fails to be one-to-one at such points. The u-curves are the so-called lines of longitude and the v-curves are the lines of latitude on the sphere.

(e) Another interesting parametrization of the sphere is given by *stereographic projection.* (Cf. Exercise 1.1.1.) We parametrize the unit sphere less the north pole $(0,0,1)$ by the xy-plane, assigning to each

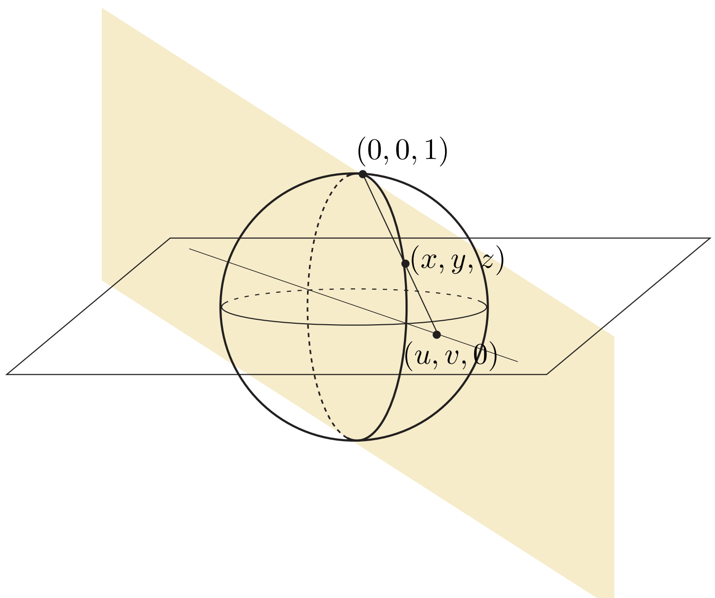


FIGURE 1.4

(u,v) the point $(\neq (0,0,1))$ where the line through $(0,0,1)$ and $(u,v,0)$ intersects the unit sphere, as pictured in Figure 1.4. We leave it to the reader to derive the following formula in Exercise 1:

$$\mathbf{x}(u,v) = \left(\frac{2u}{u^2+v^2+1}, \frac{2v}{u^2+v^2+1}, \frac{u^2+v^2-1}{u^2+v^2+1}\right). \qquad \nabla$$

For our last examples, we give two general classes of surfaces that will appear throughout our work.

Example 2. Let $I \subset \mathbb{R}$ be an interval, and let $\boldsymbol{\alpha}(u) = (0, f(u), g(u))$, $u \in I$, be a regular parametrized plane curve[2] with $f > 0$. Then the *surface of revolution* obtained by rotating $\boldsymbol{\alpha}$ about the z-axis is parametrized by

$$\mathbf{x}(u,v) = \big(f(u)\cos v, f(u)\sin v, g(u)\big), \quad u \in I,\ 0 \le v < 2\pi.$$

[2]Throughout, we assume regular parametrized curves to be one-to-one.

Note that $\mathbf{x}_u \times \mathbf{x}_v = f(u)\big(-g'(u)\cos v, -g'(u)\sin v, f'(u)\big)$, so this is a regular parametrization. The u-curves are often called *profile curves* or *meridians*; these are copies of $\boldsymbol{\alpha}$ rotated an angle v around the z-axis. The v-curves are circles, called *parallels*. ∇

Example 3. Let $I \subset \mathbb{R}$ be an interval, let $\boldsymbol{\alpha}\colon I \to \mathbb{R}^3$ be a regular parametrized curve, and let $\boldsymbol{\beta}\colon I \to \mathbb{R}^3$ be an arbitrary smooth function with $\boldsymbol{\beta}(u) \neq \mathbf{0}$ for all $u \in I$. We define a parametrized surface by

$$\mathbf{x}(u,v) = \boldsymbol{\alpha}(u) + v\boldsymbol{\beta}(u), \quad u \in I,\ v \in \mathbb{R}.$$

This is called a *ruled surface* with *rulings* $\boldsymbol{\beta}(u)$ and *directrix* $\boldsymbol{\alpha}$. It is easy to check that $\mathbf{x}_u \times \mathbf{x}_v = (\boldsymbol{\alpha}'(u) + v\boldsymbol{\beta}'(u)) \times \boldsymbol{\beta}(u)$, which may or may not be everywhere nonzero.

As particular examples, we have the helicoid (see Figure 1.2) and the following (see Figure 1.5):

(1) Cylinder: Here $\boldsymbol{\beta}$ is a constant vector, and the surface is regular as long as $\boldsymbol{\alpha}$ is one-to-one with $\boldsymbol{\alpha}' \neq \boldsymbol{\beta}$.

(2) Cone: Here we fix a point (say the origin) as the vertex, let $\boldsymbol{\alpha}$ be a curve with $\boldsymbol{\alpha} \times \boldsymbol{\alpha}' \neq \mathbf{0}$, and let $\boldsymbol{\beta} = -\boldsymbol{\alpha}$. Obviously, this fails to be a regular surface at the vertex (when $v = 1$), but $\mathbf{x}_u \times \mathbf{x}_v = (v-1)\boldsymbol{\alpha}(u) \times \boldsymbol{\alpha}'(u)$ is nonzero otherwise. (Note that another way to parametrize this surface would be to take $\boldsymbol{\alpha}^* = \mathbf{0}$ and $\boldsymbol{\beta}^* = \boldsymbol{\alpha}$.)

(3) Tangent developable: Let $\boldsymbol{\alpha}$ be a regular parametrized curve with nonzero curvature, and let $\boldsymbol{\beta} = \boldsymbol{\alpha}'$; that is, the rulings are the tangent lines of the curve $\boldsymbol{\alpha}$. Then $\mathbf{x}_u \times \mathbf{x}_v = -v\boldsymbol{\alpha}'(u) \times \boldsymbol{\alpha}''(u)$, so (at least locally) this is a regular parametrized surface away from the directrix. ∇

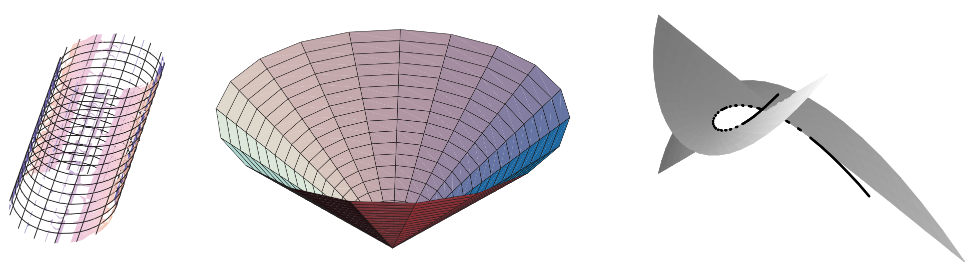


FIGURE 1.5

In calculus, we learn that, given a differentiable function f, the best linear approximation to the graph $y = f(x)$ "near" $x = a$ is given by the tangent line $y = f'(a)(x-a) + f(a)$, and similarly in higher dimensions. In the case of a regular parametrized surface, it seems reasonable that the tangent plane at $P = \mathbf{x}(u_0, v_0)$ should contain the tangent vector to the u-curve $\boldsymbol{\alpha}_1(u) = \mathbf{x}(u, v_0)$ at $u = u_0$ and the tangent vector to the v-curve $\boldsymbol{\alpha}_2(v) = \mathbf{x}(u_0, v)$ at $v = v_0$. That is, the tangent plane should contain the vectors $\mathbf{x}_u$ and $\mathbf{x}_v$, each evaluated at (u_0, v_0). Now, since $\mathbf{x}_u \times \mathbf{x}_v \neq \mathbf{0}$ by hypothesis, the vectors $\mathbf{x}_u$ and $\mathbf{x}_v$ are linearly independent and must therefore span a plane. We now make this an official

Definition. Let M be a regular parametrized surface, and let $P \in M$. Then choose a regular parametrization $\mathbf{x}\colon U \to M \subset \mathbb{R}^3$ with $P = \mathbf{x}(u_0, v_0)$. We define the *tangent plane* of M at P to be the subspace $T_P M$ spanned by $\mathbf{x}_u$ and $\mathbf{x}_v$ (evaluated at (u_0, v_0)).

Remark. The alert reader may wonder what happens if two people pick two different such local parametrizations of M near P. Do they both provide the same plane $T_P M$? This sort of question is very common in differential geometry, and is not one we intend to belabor in this introductory course. However, to get a feel for how such arguments go, the reader may work Exercise 15.

There are two unit vectors orthogonal to the tangent plane $T_P M$. Given a regular parametrization $\mathbf{x}$, we know that $\mathbf{x}_u \times \mathbf{x}_v$ is a nonzero vector orthogonal to the plane spanned by $\mathbf{x}_u$ and $\mathbf{x}_v$; we obtain the corresponding unit vector by taking

$$\mathbf{n} = \frac{\mathbf{x}_u \times \mathbf{x}_v}{\|\mathbf{x}_u \times \mathbf{x}_v\|}.$$

This is called the *unit normal* of the parametrized surface.

Example 4. We know from basic geometry and vector calculus that the unit normal of the unit sphere centered at the origin should be the position vector itself. This is in fact what we discovered in Example 1(d). ∇

Example 5. Consider the helicoid given in Example 1(b). Then, as we saw, $\mathbf{x}_u \times \mathbf{x}_v = (b\sin v, -b\cos v, u)$, and $\mathbf{n} = \dfrac{1}{\sqrt{u^2+b^2}}(b\sin v, -b\cos v, u)$. As we move along a ruling $v = v_0$, the normal starts horizontal at $u = 0$ (where the surface becomes vertical) and rotates in the plane orthogonal to the ruling, becoming more and more vertical as we move out the ruling. ∇

We saw in Chapter 1 that the geometry of a space curve is best understood by calculating (at least in principle) with an arclength parametrization. It would be nice, analogously, if we could find a parametrization $\mathbf{x}(u,v)$ of a surface so that $\mathbf{x}_u$ and $\mathbf{x}_v$ form an orthonormal basis at each point. We'll see later that this can happen only very rarely. But it makes it natural to introduce what is classically called the *first fundamental form*, $\mathrm{I}_P(\mathbf{U},\mathbf{V}) = \mathbf{U}\cdot\mathbf{V}$, for $\mathbf{U},\mathbf{V} \in T_P M$. Working in a parametrization, we have the natural basis $\{\mathbf{x}_u, \mathbf{x}_v\}$, and so we define

$$\begin{aligned}
E &= \mathrm{I}_P(\mathbf{x}_u,\mathbf{x}_u) = \mathbf{x}_u\cdot\mathbf{x}_u \\
F &= \mathrm{I}_P(\mathbf{x}_u,\mathbf{x}_v) = \mathbf{x}_u\cdot\mathbf{x}_v = \mathbf{x}_v\cdot\mathbf{x}_u = \mathrm{I}_P(\mathbf{x}_v,\mathbf{x}_u) \\
G &= \mathrm{I}_P(\mathbf{x}_v,\mathbf{x}_v) = \mathbf{x}_v\cdot\mathbf{x}_v,
\end{aligned}$$

and it is often convenient to put these in as entries of a (symmetric) matrix:

$$\mathrm{I}_P = \begin{bmatrix} E & F \\ F & G \end{bmatrix}.$$

Then, given tangent vectors $\mathbf{U} = a\mathbf{x}_u + b\mathbf{x}_v$ and $\mathbf{V} = c\mathbf{x}_u + d\mathbf{x}_v \in T_P M$, we have

$$\mathbf{U}\cdot\mathbf{V} = \mathrm{I}_P(\mathbf{U},\mathbf{V}) = (a\mathbf{x}_u + b\mathbf{x}_v)\cdot(c\mathbf{x}_u + d\mathbf{x}_v) = E(ac) + F(ad+bc) + G(bd).$$

In particular, $\|\mathbf{U}\|^2 = \mathrm{I}_P(\mathbf{U},\mathbf{U}) = Ea^2 + 2Fab + Gb^2$.

Suppose M and M^* are surfaces. We say they are *locally isometric* if for each $P \in M$ there are a regular parametrization $\mathbf{x}\colon U \to M$ with $\mathbf{x}(u_0,v_0) = P$ and a regular parametrization $\mathbf{x}^*\colon U \to M^*$ (using the same domain $U \subset \mathbb{R}^2$) with the property that $\mathrm{I}_P = \mathrm{I}^*_{P^*}$ whenever $P = \mathbf{x}(u,v)$ and $P^* = \mathbf{x}^*(u,v)$ for some $(u,v) \in U$. That is, the function $\mathbf{f} = \mathbf{x}^*\circ\mathbf{x}^{-1}\colon \mathbf{x}(U) \to \mathbf{x}^*(U)$ is a one-to-one correspondence that preserves the first fundamental form and is therefore distance-preserving (see Exercise 2).

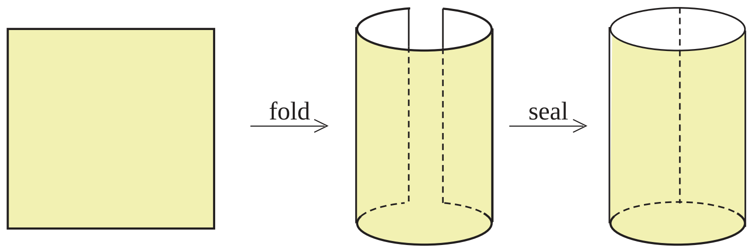


FIGURE 1.6

Example 6. Parametrize a portion of the plane (say, a piece of paper) by $\mathbf{x}(u,v) = (u,v,0)$ and a portion of a cylinder by $\mathbf{x}^*(u,v) = (\cos u, \sin u, v)$. Then it is easy to calculate that $E = E^* = 1$, $F = F^* = 0$, and $G = G^* = 1$, so these surfaces, pictured in Figure 1.6, are locally isometric. On the other hand, if we let u vary from 0 to 2π, the rectangle and the cylinder are not *globally* isometric because points far away in the rectangle can become very close (or identical) in the cylinder. ∇

If $\boldsymbol{\alpha}(t) = \mathbf{x}(u(t), v(t))$ is a curve on the parametrized surface M with $\boldsymbol{\alpha}(t_0) = \mathbf{x}(u_0, v_0) = P$, then it is an immediate consequence of the chain rule, Theorem 2.2 of the Appendix, that

$$\boldsymbol{\alpha}'(t_0) = u'(t_0)\mathbf{x}_u(u_0,v_0) + v'(t_0)\mathbf{x}_v(u_0,v_0).$$

(Customarily we will write simply $\mathbf{x}_u$, the point (u_0, v_0) at which it is evaluated being assumed.) That is, if the tangent vector $(u'(t_0), v'(t_0))$ back in the "parameter space" is (a,b), then the tangent vector to $\boldsymbol{\alpha}$ at P is the corresponding linear combination $a\mathbf{x}_u + b\mathbf{x}_v$. In fancy terms, this is merely a consequence of the linearity of the derivative of $\mathbf{x}$. We say a parametrization $\mathbf{x}(u,v)$ is *conformal* if angles measured in the

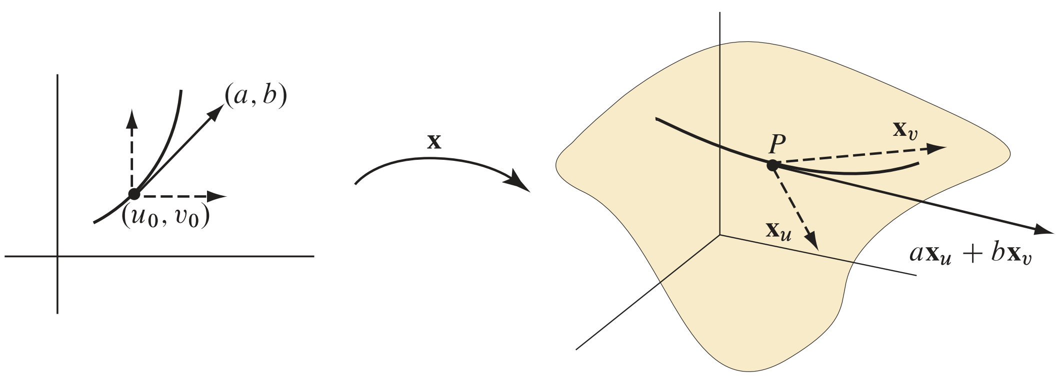


FIGURE 1.7

uv-plane agree with corresponding angles in T_PM for all P. We leave it to the reader to check in Exercise 6 that this is equivalent to the conditions $E = G$, $F = 0$.

Since

$$\begin{bmatrix} E & F \\ F & G \end{bmatrix} = \begin{bmatrix} \mathbf{x}_u\cdot\mathbf{x}_u & \mathbf{x}_u\cdot\mathbf{x}_v \\ \mathbf{x}_v\cdot\mathbf{x}_u & \mathbf{x}_v\cdot\mathbf{x}_v \end{bmatrix} = \begin{bmatrix} | & | \\ \mathbf{x}_u & \mathbf{x}_v \\ | & | \end{bmatrix}^{\mathsf{T}} \begin{bmatrix} | & | \\ \mathbf{x}_u & \mathbf{x}_v \\ | & | \end{bmatrix},$$

we have

$$EG - F^2 = \det\left(\begin{bmatrix} \mathbf{x}_u \cdot \mathbf{x}_u & \mathbf{x}_u \cdot \mathbf{x}_v \\ \mathbf{x}_v \cdot \mathbf{x}_u & \mathbf{x}_v \cdot \mathbf{x}_v \end{bmatrix}\right) = \det\left(\begin{bmatrix} \mathbf{x}_u \cdot \mathbf{x}_u & \mathbf{x}_u \cdot \mathbf{x}_v & 0 \\ \mathbf{x}_v \cdot \mathbf{x}_u & \mathbf{x}_v \cdot \mathbf{x}_v & 0 \\ 0 & 0 & 1 \end{bmatrix}\right)$$

$$= \det\left(\begin{bmatrix} | & | & | \\ \mathbf{x}_u & \mathbf{x}_v & \mathbf{n} \\ | & | & | \end{bmatrix}^{\mathsf{T}} \begin{bmatrix} | & | & | \\ \mathbf{x}_u & \mathbf{x}_v & \mathbf{n} \\ | & | & | \end{bmatrix}\right) = \left(\det\begin{bmatrix} | & | & | \\ \mathbf{x}_u & \mathbf{x}_v & \mathbf{n} \\ | & | & | \end{bmatrix}\right)^2,$$

which is the square of the volume of the parallelepiped spanned by $\mathbf{x}_u$, $\mathbf{x}_v$, and $\mathbf{n}$. Since $\mathbf{n}$ is a unit vector orthogonal to the plane spanned by $\mathbf{x}_u$ and $\mathbf{x}_v$, this is, in turn, the square of the area of the parallelogram spanned by $\mathbf{x}_u$ and $\mathbf{x}_v$. That is,

$$EG - F^2 = \|\mathbf{x}_u \times \mathbf{x}_v\|^2 > 0.$$

We remind the reader that we obtain the *surface area* of the parametrized surface $\mathbf{x}\colon U \to M$ by calculating the double integral

$$\int_U \|\mathbf{x}_u \times \mathbf{x}_v\| du dv = \int_U \sqrt{EG - F^2} du dv.$$

EXERCISES 2.1

1. Derive the formula given in Example 1(e) for the parametrization of the unit sphere.

♯2. Suppose $\boldsymbol{\alpha}(t) = \mathbf{x}(u(t), v(t))$, $a \le t \le b$, is a parametrized curve on a surface M. Show that

$$\begin{aligned} \text{length}(\boldsymbol{\alpha}) &= \int_a^b \sqrt{\mathrm{I}_{\boldsymbol{\alpha}(t)}\big(\boldsymbol{\alpha}'(t), \boldsymbol{\alpha}'(t)\big)}\, dt \\ &= \int_a^b \sqrt{E(u(t), v(t))(u'(t))^2 + 2F(u(t), v(t))u'(t)v'(t) + G(u(t), v(t))(v'(t))^2}\, dt. \end{aligned}$$

Conclude that if $\boldsymbol{\alpha} \subset M$ and $\boldsymbol{\alpha}^* \subset M^*$ are corresponding paths in locally isometric surfaces, then $\text{length}(\boldsymbol{\alpha}) = \text{length}(\boldsymbol{\alpha}^*)$.

3. Compute I (i.e., E, F, and G) for the following parametrized surfaces.
   *a. the sphere of radius a: $\mathbf{x}(u, v) = a(\sin u \cos v, \sin u \sin v, \cos u)$
   b. the torus: $\mathbf{x}(u, v) = ((a + b\cos u)\cos v, (a + b\cos u)\sin v, b\sin u)$ $(0 < b < a)$
   c. the helicoid: $\mathbf{x}(u, v) = (u\cos v, u\sin v, bv)$
   *d. the catenoid: $\mathbf{x}(u, v) = a(\cosh u \cos v, \cosh u \sin v, u)$

4. Find the surface area of the following parametrized surfaces.
   *a. the torus: $\mathbf{x}(u, v) = ((a + b\cos u)\cos v, (a + b\cos u)\sin v, b\sin u)$ $(0 < b < a)$, $0 \le u, v \le 2\pi$
   b. a portion of the helicoid: $\mathbf{x}(u, v) = (u\cos v, u\sin v, bv)$, $1 < u < 3$, $0 \le v \le 2\pi$
   c. a zone of a sphere[3]: $\mathbf{x}(u, v) = a(\sin u \cos v, \sin u \sin v, \cos u)$, $0 \le u_0 \le u \le u_1 \le \pi$, $0 \le v \le 2\pi$

[3] You should obtain the remarkable result that the surface area of the portion of a sphere between two parallel planes depends only on the distance between the planes, not on where you locate them.

*5. Show that if all the normal lines to a surface pass through a fixed point, then the surface is (a portion of) a sphere. (By the normal line to M at P we mean the line passing through P with direction vector the unit normal at P.)

6. Check that the parametrization $\mathbf{x}(u,v)$ is conformal if and only if $E = G$ and $F = 0$. (Hint: For $\Longrightarrow$, choose *two* convenient pairs of orthogonal directions.)

*7. Check that a parametrization preserves area and is conformal if and only if it is a local isometry.

*8. Check that the parametrization of the unit sphere by stereographic projection (see Example 1(e)) is conformal.

9. *(Lambert's cylindrical projection)* Project the unit sphere (except for the north and south poles) radially outward to the cylinder of radius 1 by sending (x,y,z) to $(x/\sqrt{x^2+y^2}, y/\sqrt{x^2+y^2}, z)$. Check that this map preserves area locally, but is neither a local isometry nor conformal. (Hint: Let $\mathbf{x}(u,v)$ be the spherical coordinates parametrization of the sphere, and consider $\mathbf{x}^*(u,v) = (\cos v, \sin v, \cos u)$. Compare the parallelogram formed by $\mathbf{x}_u$ and $\mathbf{x}_v$ with the parallelogram formed by $\mathbf{x}^*_u$ and $\mathbf{x}^*_v$.)

♯10. Consider the "pacman" region M given by $\mathbf{x}(u,v) = (u\cos v, u\sin v, 0)$, $0 \le u \le R$, $0 \le v \le V$, with $V < 2\pi$. Let $c = V/2\pi$. Let M^* be given by the parametrization

$$\mathbf{x}^*(u,v) = \big(cu\cos(v/c), cu\sin(v/c), \sqrt{1-c^2}u\big), \quad 0 \le u \le R,\ 0 \le v \le V.$$

Compute that $E = E^*$, $F = F^*$, and $G = G^*$, and conclude that the mapping $\mathbf{f} = \mathbf{x}^* \circ \mathbf{x}^{-1}\colon M \to M^*$ is a local isometry. Describe this mapping in concrete geometric terms.

11. Consider the hyperboloid of one sheet, M, given by the equation $x^2 + y^2 - z^2 = 1$.
   a. Show that $\mathbf{x}(u,v) = (\cosh u\cos v, \cosh u\sin v, \sinh u)$, $u \in \mathbb{R}$, $0 \le v < 2\pi$, gives a parametrization of M as a surface of revolution.
   *b. Find two parametrizations of M as a ruled surface $\boldsymbol{\alpha}(u) + v\boldsymbol{\beta}(u)$.
   c. Show that $\mathbf{x}(u,v) = \left(\dfrac{uv+1}{uv-1}, \dfrac{u-v}{uv-1}, \dfrac{u+v}{uv-1}\right)$ gives a parametrization of M where *both* sets of parameter curves are rulings.

♯12. Given a ruled surface M parametrized by $\mathbf{x}(u,v) = \boldsymbol{\alpha}(u) + v\boldsymbol{\beta}(u)$ with $\boldsymbol{\alpha}' \neq 0$ and $\|\boldsymbol{\beta}\| = 1$.
   a. Check that we may assume that $\boldsymbol{\alpha}'(u) \cdot \boldsymbol{\beta}(u) = 0$ for all u. (Hint: Replace $\boldsymbol{\alpha}(u)$ with $\boldsymbol{\alpha}(u) + t(u)\boldsymbol{\beta}(u)$ for a suitable function t.)
   b. Suppose, moreover, that $\boldsymbol{\alpha}'(u)$, $\boldsymbol{\beta}(u)$, and $\boldsymbol{\beta}'(u)$ are linearly dependent for every u. Conclude that $\boldsymbol{\beta}'(u) = \lambda(u)\boldsymbol{\alpha}'(u)$ for some function λ. Prove that:
      (i) If $\lambda(u) = 0$ for all u, then M is a cylinder.
      (ii) If λ is a nonzero constant, then M is a cone.
      (iii) If λ and λ' are both nowhere zero, then M is a tangent developable. (Hint: Find the directrix.)

13. *(The Mercator projection)* Mercator developed his system for mapping the earth, as pictured in Figure 1.8, in 1569, about a century before the advent of calculus. We want a parametrization $\mathbf{x}(u,v)$ of the sphere, $u \in \mathbb{R}$, $v \in (-\pi,\pi)$, so that the u-curves are the longitudes and so that the parametrization is conformal. Letting (ϕ,θ) be the usual spherical coordinates, write $\phi = f(u)$ and $\theta = v$. Show that

FIGURE 1.8

conformality and symmetry about the equator will dictate $f(u) = 2\arctan(e^{-u})$. Deduce that

$$\mathbf{x}(u,v) = (\operatorname{sech} u \cos v, \operatorname{sech} u \sin v, \tanh u).$$

(Cf. Example 2 in Section 1 of Chapter 1.)

14. A parametrization $\mathbf{x}(u,v)$ is called a *Tschebyschev net* if the opposite sides of any quadrilateral formed by the coordinate curves have equal length.
    a. Prove that this occurs if and only if $\dfrac{\partial E}{\partial v} = \dfrac{\partial G}{\partial u} = 0$. (Hint: Express the length of the u-curves, $u_0 \le u \le u_1$, as an integral and use the fact that this length is independent of v.)
    b. Prove that we can locally reparametrize by $\tilde{\mathbf{x}}(\tilde{u},\tilde{v})$ so as to obtain $\tilde{E} = \tilde{G} = 1$, $\tilde{F} = \cos\theta(\tilde{u},\tilde{v})$ (so that the $\tilde{u}$- and $\tilde{v}$-curves are parametrized by arclength and meet at angle θ). (Hint: Choose $\tilde{u}$ as a function of u so that $\tilde{\mathbf{x}}_{\tilde{u}} = \mathbf{x}_u \big/ (d\tilde{u}/du)$ has unit length.)

15. Suppose $\mathbf{x}$ and $\mathbf{y}$ are two parametrizations of a surface M near P. Say $\mathbf{x}(u_0,v_0) = P = \mathbf{y}(s_0,t_0)$. Prove that $\operatorname{Span}(\mathbf{x}_u,\mathbf{x}_v) = \operatorname{Span}(\mathbf{y}_s,\mathbf{y}_t)$ (where the partial derivatives are all evaluated at the obvious points). (Hint: $\mathbf{f} = \mathbf{x}^{-1}\circ\mathbf{y}$ gives a $\mathcal{C}^1$ map from an open set around (s_0,t_0) to an open set around (u_0,v_0). Apply the chain rule to show $\mathbf{y}_s, \mathbf{y}_t \in \operatorname{Span}(\mathbf{x}_u,\mathbf{x}_v)$.)

16. (A programmable calculator, Maple, or Mathematica will be needed for parts of this problem.) A catenoid, as pictured in Figure 1.9, is parametrized by

$$\mathbf{x}(u,v) = (a\cosh u\cos v, a\cosh u \sin v, au), \quad u \in \mathbb{R},\ 0 \le v < 2\pi \quad (a > 0 \text{ fixed}).$$

    *a. Compute the surface area of that portion of the catenoid given by $|u| \le 1/a$. (Hint: $\cosh^2 u = \frac{1}{2}(1+\cosh 2u)$.)
    b. Find the number $R_0 > 0$ so that for every $R \ge R_0$, there is at least one catenoid whose boundary is the pair of parallel circles $x^2 + y^2 = R^2$, $|z| = 1$. (Hint: Graph $f(t) = t\cosh(1/t)$.)
    c. For $R \ge R_0$, compare the area of the catenoid(s) with $2\pi R^2$ (the area of the pair of disks filling in the circles). For what values of R does the pair of disks have the least area? (You should display the results of your investigation in either a graph or a table.)

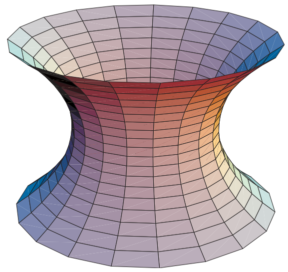


FIGURE 1.9

d. (For extra credit) Show that as $R \to \infty$, the area of the inner catenoid is asymptotic to $2\pi R^2$ and the area of the outer catenoid is asymptotic to $4\pi R$.

17. There are two obvious families of circles on a torus. Find a third family. (Hint: Look for a plane that is tangent to the torus at *two* points. Using the parametrization of the torus, you should be able to find equations (either parametric or cartesian) for the curve in which the bitangent plane intersects the torus.)

2. The Gauss Map and the Second Fundamental Form

Given a regular parametrized surface M, the function $\mathbf{n}: M \to \Sigma$ that assigns to each point $P \in M$ the unit normal $\mathbf{n}(P)$, as pictured in Figure 2.1, is called the *Gauss map* of M. As we shall see in this chapter,

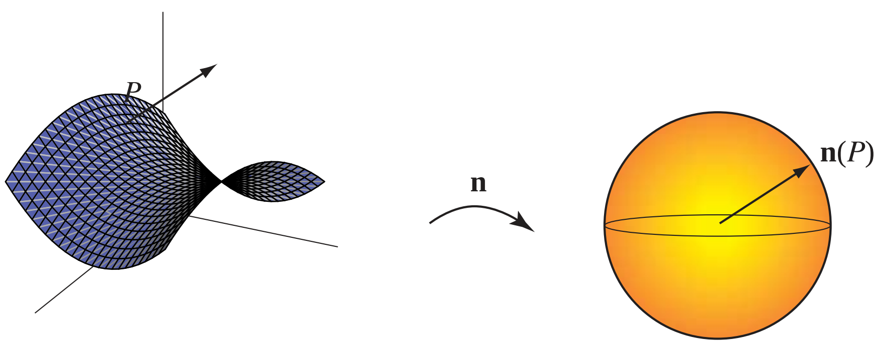


FIGURE 2.1

most of the geometric information about our surface M is encapsulated in the mapping $\mathbf{n}$.

Example 1. A few basic examples are these.

(a) On a plane, the tangent plane never changes, so the Gauss map is a constant.

(b) On a cylinder, the tangent plane is constant along the rulings, so the Gauss map sends the entire surface to an equator of the sphere.

(c) On a sphere centered at the origin, the Gauss map is merely the (normalized) position vector.

(d) On a saddle surface (as pictured in Figure 2.1), the Gauss map appears to "reverse orientation": As we move counterclockwise in a small circle around P, we see that the unit vector $\mathbf{n}$ turns clockwise around $\mathbf{n}(P)$. ∇

Recall from the Appendix that for any function f on M (scalar- or vector-valued) and any tangent vector $\mathbf{V} \in T_P M$, we can compute the directional derivative $D_{\mathbf{V}} f(P)$ by choosing a curve $\boldsymbol{\alpha}\colon (-\varepsilon, \varepsilon) \to M$ with $\boldsymbol{\alpha}(0) = P$ and $\boldsymbol{\alpha}'(0) = \mathbf{V}$ and computing $(f \circ \boldsymbol{\alpha})'(0)$.

To understand the shape of M at the point P, we might try to understand the curvature at P of various curves in M. Perhaps the most obvious thing to try is various *normal* slices of M. That is, we slice M with the plane through P spanned by $\mathbf{n}(P)$ and a *unit* vector $\mathbf{V} \in T_P M$. Various such normal slices are shown for a saddle surface in Figure 2.2. Let $\boldsymbol{\alpha}$ be the arclength-parametrized curve obtained by taking such

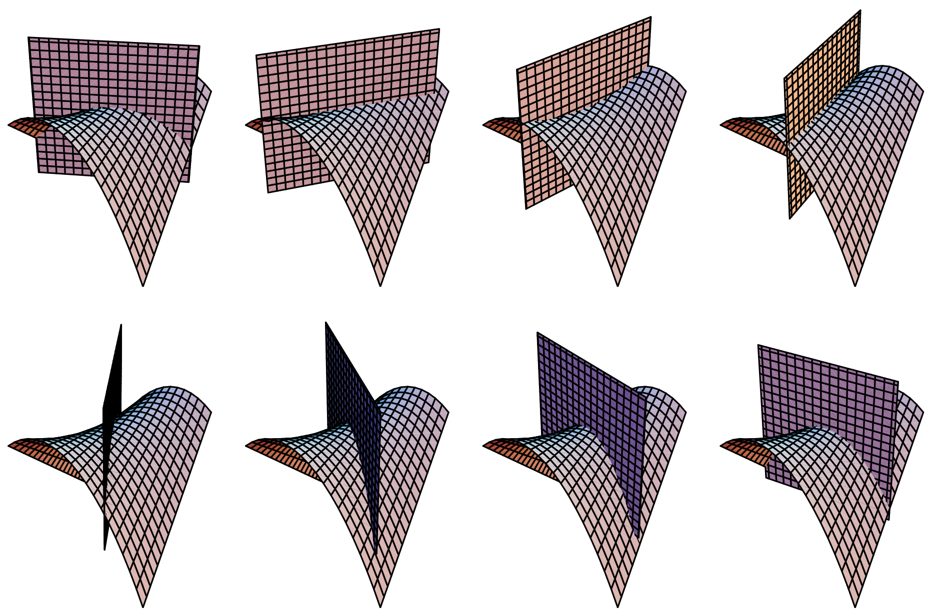


FIGURE 2.2

a normal slice. We have $\boldsymbol{\alpha}(0) = P$ and $\boldsymbol{\alpha}'(0) = \mathbf{V}$. Then since the curve lies in the plane spanned by $\mathbf{n}(P)$ and $\mathbf{V}$, the principal normal of the curve at P must be $\pm\mathbf{n}(P)$ ($+$ if the curve is curving towards $\mathbf{n}$, $-$ if it's curving away). Since $(\mathbf{n} \circ \boldsymbol{\alpha}(s)) \cdot \mathbf{T}(s) = 0$ for all s near 0, applying Lemma 2.1 of Chapter 1 yet again, we have:

$$(\dagger) \qquad \pm\kappa(P) = \kappa \mathbf{N} \cdot \mathbf{n}(P) = \mathbf{T}'(0) \cdot \mathbf{n}(P) = -\mathbf{T}(0) \cdot (\mathbf{n} \circ \boldsymbol{\alpha})'(0) = -D_{\mathbf{V}}\mathbf{n}(P) \cdot \mathbf{V}.$$

This leads us to study the directional derivative $D_{\mathbf{V}}\mathbf{n}(P)$ more carefully.

Proposition 2.1. *For any $\mathbf{V} \in T_P M$, the directional derivative $D_{\mathbf{V}}\mathbf{n}(P) \in T_P M$. Moreover, the linear map $S_P\colon T_P M \to T_P M$ defined by*

$$S_P(\mathbf{V}) = -D_{\mathbf{V}}\mathbf{n}(P)$$

is a symmetric *linear map; i.e., for any $\mathbf{U}, \mathbf{V} \in T_P M$, we have*

$$(*) \qquad S_P(\mathbf{U}) \cdot \mathbf{V} = \mathbf{U} \cdot S_P(\mathbf{V})$$

S_P is called the shape operator *at P.*

Proof. For any curve $\boldsymbol{\alpha}\colon (-\varepsilon, \varepsilon) \to M$ with $\boldsymbol{\alpha}(0) = P$ and $\boldsymbol{\alpha}'(0) = \mathbf{V}$, we observe that $\mathbf{n} \circ \boldsymbol{\alpha}$ has constant length 1. Thus, by Lemma 2.1 of Chapter 1, $D_{\mathbf{V}}\mathbf{n}(P) \cdot \mathbf{n}(P) = (\mathbf{n} \circ \boldsymbol{\alpha})'(0) \cdot (\mathbf{n} \circ \boldsymbol{\alpha})(0) = 0$, so $D_{\mathbf{V}}\mathbf{n}(P)$ is in the tangent plane to M at P. That S_P is a linear map is an immediate consequence of Proposition 2.3 of the Appendix.

Symmetry is our first important application of the equality of mixed partial derivatives. First we verify $(*)$ when $\mathbf{U} = \mathbf{x}_u$, $\mathbf{V} = \mathbf{x}_v$. Note that $\mathbf{n}\cdot\mathbf{x}_v = 0$, so $0 = \big(\mathbf{n}\cdot\mathbf{x}_v\big)_u = \mathbf{n}_u\cdot\mathbf{x}_v + \mathbf{n}\cdot\mathbf{x}_{vu}$. (Remember that we're writing $\mathbf{n}_u$ for $D_{\mathbf{x}_u}\mathbf{n}$.) Thus,

$$\begin{aligned} S_P(\mathbf{x}_u)\cdot\mathbf{x}_v &= -D_{\mathbf{x}_u}\mathbf{n}(P)\cdot\mathbf{x}_v = -\mathbf{n}_u\cdot\mathbf{x}_v = \mathbf{n}\cdot\mathbf{x}_{vu} \\ &= \mathbf{n}\cdot\mathbf{x}_{uv} = -\mathbf{n}_v\cdot\mathbf{x}_u = -D_{\mathbf{x}_v}\mathbf{n}(P)\cdot\mathbf{x}_u = S_P(\mathbf{x}_v)\cdot\mathbf{x}_u. \end{aligned}$$

Next, knowing this, we just write out general vectors $\mathbf{U}$ and $\mathbf{V}$ as linear combinations of $\mathbf{x}_u$ and $\mathbf{x}_v$: If $\mathbf{U} = a\mathbf{x}_u + b\mathbf{x}_v$ and $\mathbf{V} = c\mathbf{x}_u + d\mathbf{x}_v$, then

$$\begin{aligned} S_P(\mathbf{U})\cdot\mathbf{V} &= S_P(a\mathbf{x}_u + b\mathbf{x}_v)\cdot(c\mathbf{x}_u + d\mathbf{x}_v) \\ &= \big(aS_P(\mathbf{x}_u) + bS_P(\mathbf{x}_v)\big)\cdot(c\mathbf{x}_u + d\mathbf{x}_v) \\ &= acS_P(\mathbf{x}_u)\cdot\mathbf{x}_u + adS_P(\mathbf{x}_u)\cdot\mathbf{x}_v + bcS_P(\mathbf{x}_v)\cdot\mathbf{x}_u + bdS_P(\mathbf{x}_v)\cdot\mathbf{x}_v \\ &= acS_P(\mathbf{x}_u)\cdot\mathbf{x}_u + adS_P(\mathbf{x}_v)\cdot\mathbf{x}_u + bcS_P(\mathbf{x}_u)\cdot\mathbf{x}_v + bdS_P(\mathbf{x}_v)\cdot\mathbf{x}_v \\ &= (a\mathbf{x}_u + b\mathbf{x}_v)\cdot\big(cS_P(\mathbf{x}_u) + dS_P(\mathbf{x}_v)\big) = \mathbf{U}\cdot S_P(\mathbf{V}), \end{aligned}$$

as required. □

Proposition 2.2. *If the shape operator S_P is O for all $P \in M$, then M is a subset of a plane.*

Proof. Since the directional derivative of the unit normal $\mathbf{n}$ is $\mathbf{0}$ in every direction at every point P, we have $\mathbf{n}_u = \mathbf{n}_v = \mathbf{0}$ for any (local) parametrization $\mathbf{x}(u,v)$ of M. By Proposition 2.4 of the Appendix, it follows that $\mathbf{n}$ is constant. (This is why we assume our surfaces are connected.) □

Example 2. Let M be a sphere of radius a centered at the origin. Then $\mathbf{n} = \dfrac{1}{a}\mathbf{x}(u,v)$, so for any P, we have $S_P(\mathbf{x}_u) = -\mathbf{n}_u = -\dfrac{1}{a}\mathbf{x}_u$ and $S_P(\mathbf{x}_v) = -\mathbf{n}_v = -\dfrac{1}{a}\mathbf{x}_v$, so S_P is $-1/a$ times the identity map on the tangent plane T_PM. ▽

It does not seem an easy task to give the matrix of the shape operator with respect to the basis $\{\mathbf{x}_u, \mathbf{x}_v\}$. But, in general, the proof of Proposition 2.1 suggests that we define the second fundamental form, as follows. If $\mathbf{U}, \mathbf{V} \in T_PM$, we set

$$\mathrm{II}_P(\mathbf{U},\mathbf{V}) = S_P(\mathbf{U})\cdot\mathbf{V}.$$

Note that the formula (†) on p. 45 shows that the curvature of the normal slice in direction $\mathbf{V}$ (with $\|\mathbf{V}\| = 1$) is, in our new notation, given by

$$\pm\kappa = -D_{\mathbf{V}}\mathbf{n}(P)\cdot\mathbf{V} = S_P(\mathbf{V})\cdot\mathbf{V} = \mathrm{II}_P(\mathbf{V},\mathbf{V}).$$

As we did at the end of the previous section, we wish to give a matrix representation when we're working with a parametrized surface. As we saw in the proof of Proposition 2.1, we have

$$\begin{aligned} \ell &= \mathrm{II}_P(\mathbf{x}_u,\mathbf{x}_u) = -D_{\mathbf{x}_u}\mathbf{n}\cdot\mathbf{x}_u = \mathbf{x}_{uu}\cdot\mathbf{n} \\ m &= \mathrm{II}_P(\mathbf{x}_u,\mathbf{x}_v) = -D_{\mathbf{x}_u}\mathbf{n}\cdot\mathbf{x}_v = \mathbf{x}_{vu}\cdot\mathbf{n} = \mathbf{x}_{uv}\cdot\mathbf{n} = \mathrm{II}_P(\mathbf{x}_v,\mathbf{x}_u) \\ n &= \mathrm{II}_P(\mathbf{x}_v,\mathbf{x}_v) = -D_{\mathbf{x}_v}\mathbf{n}\cdot\mathbf{x}_v = \mathbf{x}_{vv}\cdot\mathbf{n}. \end{aligned}$$

(By the way, this explains the presence of the minus sign in the original definition of the shape operator.) We then write

$$\mathrm{II}_P = \begin{bmatrix} \ell & m \\ m & n \end{bmatrix} = \begin{bmatrix} \mathbf{x}_{uu} \cdot \mathbf{n} & \mathbf{x}_{uv} \cdot \mathbf{n} \\ \mathbf{x}_{uv} \cdot \mathbf{n} & \mathbf{x}_{vv} \cdot \mathbf{n} \end{bmatrix}.$$

If, as before, $\mathbf{U} = a\mathbf{x}_u + b\mathbf{x}_v$ and $\mathbf{V} = c\mathbf{x}_u + d\mathbf{x}_v$, then

$$\begin{aligned} \mathrm{II}_P(\mathbf{U},\mathbf{V}) &= \mathrm{II}_P(a\mathbf{x}_u + b\mathbf{x}_v, c\mathbf{x}_u + d\mathbf{x}_v) \\ &= ac\,\mathrm{II}_P(\mathbf{x}_u,\mathbf{x}_u) + ad\,\mathrm{II}_P(\mathbf{x}_u,\mathbf{x}_v) + bc\,\mathrm{II}_P(\mathbf{x}_v,\mathbf{x}_u) + bd\,\mathrm{II}_P(\mathbf{x}_v,\mathbf{x}_v) \\ &= \ell(ac) + m(bc + ad) + n(bd). \end{aligned}$$

In the event that $\{\mathbf{x}_u, \mathbf{x}_v\}$ is an orthonormal basis for $T_P M$, we see that the matrix II_P represents the shape operator S_P. But it is not difficult to check (see Exercise 2) that, in general, the matrix of the linear map S_P with respect to the basis $\{\mathbf{x}_u, \mathbf{x}_v\}$ is given by

$$\mathrm{I}_P^{-1}\mathrm{II}_P = \begin{bmatrix} E & F \\ F & G \end{bmatrix}^{-1} \begin{bmatrix} \ell & m \\ m & n \end{bmatrix}.$$

Remark. We proved in Proposition 2.1 that S_P is a symmetric linear map. This means that its matrix representation with respect to an orthonormal basis (or, more generally, orthogonal basis with vectors of equal length) *will* be symmetric: In this case the matrix I_P is a scalar multiple of the identity matrix and the matrix product remains symmetric.

By the Spectral Theorem, Theorem 1.3 of the Appendix, S_P has two real eigenvalues, traditionally denoted $k_1(P), k_2(P)$.

Definition. The eigenvalues of S_P are called the *principal curvatures* of M at P. Corresponding eigenvectors are called *principal directions*. A curve in M is called a *line of curvature* if its tangent vector at each point is a principal direction.

Recall that it also follows from the Spectral Theorem that the principal directions are orthogonal, so we can always choose an orthonormal basis for $T_P M$ consisting of principal directions. Having done so, we can then easily determine the curvatures of normal slices in arbitrary directions, as follows.

Proposition 2.3 (Euler's Formula). *Let $\mathbf{e}_1, \mathbf{e}_2$ be unit vectors in the principal directions at P with corresponding principal curvatures k_1 and k_2. Suppose $\mathbf{V} = \cos\theta\mathbf{e}_1 + \sin\theta\mathbf{e}_2$ for some $\theta \in [0, 2\pi)$, as pictured in Figure 2.3. Then $\mathrm{II}_P(\mathbf{V},\mathbf{V}) = k_1\cos^2\theta + k_2\sin^2\theta$.*

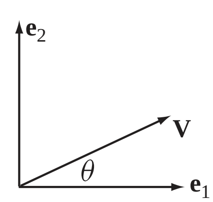


FIGURE 2.3

Proof. This is a straightforward computation: Since $S_P(\mathbf{e}_i) = k_i\mathbf{e}_i$ for $i = 1, 2$, we have

$$\mathrm{II}_P(\mathbf{V},\mathbf{V}) = S_P(\mathbf{V}) \cdot \mathbf{V} = S_P(\cos\theta\mathbf{e}_1 + \sin\theta\mathbf{e}_2) \cdot (\cos\theta\mathbf{e}_1 + \sin\theta\mathbf{e}_2)$$

$$= (\cos\theta k_1\mathbf{e}_1 + \sin\theta k_2\mathbf{e}_2)\cdot(\cos\theta\mathbf{e}_1 + \sin\theta\mathbf{e}_2) = k_1\cos^2\theta + k_2\sin^2\theta,$$

as required. □

On a sphere, all normal slices have the same (nonzero) curvature. On the other hand, if we look carefully at Figure 2.2, we see that certain normal slices of a saddle surface are true lines. This leads us to make the following

Definition. If the normal slice in direction $\mathbf{V}$ has zero curvature, i.e., if $\mathrm{II}_P(\mathbf{V},\mathbf{V}) = 0$, then we call $\mathbf{V}$ an *asymptotic direction*.[4] A curve in M is called an *asymptotic curve* if its tangent vector at each point is an asymptotic direction.

Example 3. If a surface M contains a line, that line is an asymptotic curve. For the normal slice in the direction of the line contains the line (and perhaps other things far away), which, of course, has zero curvature. ▽

Corollary 2.4. *There is an asymptotic direction at* P *if and only if* $k_1k_2 \le 0$.

Proof. $k_2 = 0$ if and only if $\mathbf{e}_2$ is an asymptotic direction. Now suppose $k_2 \ne 0$. If $\mathbf{V}$ is a unit asymptotic vector making angle θ with $\mathbf{e}_1$, then we have $k_1\cos^2\theta + k_2\sin^2\theta = 0$, and so $\tan^2\theta = -k_1/k_2$, so $k_1k_2 \le 0$. Conversely, if $k_1k_2 < 0$, take θ with $\tan\theta = \pm\sqrt{-k_1/k_2}$, and then $\mathbf{V}$ is an asymptotic direction. □

Example 4. We consider the helicoid, as pictured in Figure 1.2. It is a ruled surface and so the rulings are asymptotic curves. What is quite less obvious is that the family of helices on the surface are also asymptotic curves. But, as we see in Figure 2.4, the normal slice tangent to the helix at P has an inflection

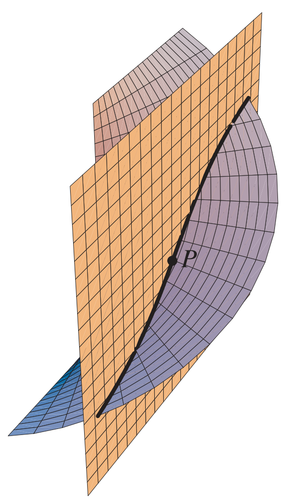


FIGURE 2.4

point at P, and therefore the helix is an asymptotic curve. We ask the reader to check this by calculation in Exercise 5. ▽

[4]Of course, $\mathbf{V} \ne \mathbf{0}$ here. See Exercise 22 for an explanation of this terminology.

It is also an immediate consequence of Proposition 2.3 that the principal curvatures are the maximum and minimum (signed) curvatures of the various normal slices. Assume $k_2 \le k_1$. Then

$$k_1\cos^2\theta + k_2\sin^2\theta = k_1(1-\sin^2\theta) + k_2\sin^2\theta = k_1 + (k_2-k_1)\sin^2\theta \le k_1$$

(and, similarly, $\ge k_2$). Moreover, as the Spectral Theorem tells us, the maximum and minimum occur at right angles to one another. Looking back at Figure 2.2, where the slices are taken at angles in increments of $\pi/8$, we see that the normal slices that are "most curved" appear in the third and seventh frames; the asymptotic directions appear in the second and fourth frames. (Cf. Exercise 8.)

Next we come to one of the most important concepts in the geometry of surfaces:

Definition. The product of the principal curvatures is called the *Gaussian curvature*: $K = \det S_P = k_1k_2$. The average of the principal curvatures is called the *mean curvature*: $H = \frac{1}{2}\operatorname{tr} S_P = \frac{1}{2}(k_1+k_2)$. We say M is a *minimal surface* if $H = 0$ and *flat* if $K = 0$.

Note that whereas the signs of the principal curvatures change if we reverse the direction of the unit normal $\mathbf{n}$, the Gaussian curvature K, being the product of both, is independent of the choice of unit normal. (And the sign of the mean curvature depends on the choice.)

Example 5. It follows from our comments in Example 1 that both a plane and a cylinder are flat surfaces: In the former case, $S_P = \mathrm{O}$ for all P, and, in the latter, $\det S_P = 0$ for all P since the shape operator is singular. ∇

Example 6. Consider the saddle surface $\mathbf{x}(u,v) = (u,v,uv)$. We compute:

$$\begin{aligned} \mathbf{x}_u &= (1,0,v) & \mathbf{x}_{uu} &= (0,0,0)\\ \mathbf{x}_v &= (0,1,u) & \mathbf{x}_{uv} &= (0,0,1)\\ \mathbf{n} &= \frac{1}{\sqrt{1+u^2+v^2}}(-v,-u,1) & \mathbf{x}_{vv} &= (0,0,0), \end{aligned}$$

and so

$$E = 1+v^2, \quad F = uv, \quad G = 1+u^2, \quad \text{and} \quad \ell = n = 0, m = \frac{1}{\sqrt{1+u^2+v^2}}.$$

Thus, with $P = \mathbf{x}(u,v)$, we have

$$\mathrm{I}_P = \begin{bmatrix} 1+v^2 & uv \\ uv & 1+u^2 \end{bmatrix} \quad \text{and} \quad \mathrm{II}_P = \frac{1}{\sqrt{1+u^2+v^2}}\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix},$$

so the matrix of the shape operator with respect to the basis $\{\mathbf{x}_u, \mathbf{x}_v\}$ is given by

$$S_P = \mathrm{I}_P^{-1}\mathrm{II}_P = \frac{1}{(1+u^2+v^2)^{3/2}}\begin{bmatrix} -uv & 1+u^2 \\ 1+v^2 & -uv \end{bmatrix}.$$

(Note that this matrix is, in general, not symmetric.)

With a bit of calculation, we determine that the principal curvatures (eigenvalues) are

$$k_1 = \frac{-uv + \sqrt{(1+u^2)(1+v^2)}}{(1+u^2+v^2)^{3/2}} \quad \text{and} \quad k_2 = \frac{-uv - \sqrt{(1+u^2)(1+v^2)}}{(1+u^2+v^2)^{3/2}},$$

and $K = \det S_P = -1/(1+u^2+v^2)^2$. Note from the form of II_P that the u- and v-curves are asymptotic curves, as should be evident from the fact that these are lines. With a bit more work, we determine that the principal directions, i.e., the eigenvectors of S_P, are the vectors

$$\sqrt{1+u^2}\mathbf{x}_u \pm \sqrt{1+v^2}\mathbf{x}_v .$$

(It is worth checking that these vectors are, in fact, orthogonal.) The corresponding curves in the uv-plane have tangent vectors $\left(\sqrt{1+u^2}, \pm\sqrt{1+v^2}\right)$ and must therefore be solutions of the differential equation

$$\frac{dv}{du} = \pm\frac{\sqrt{1+v^2}}{\sqrt{1+u^2}}.$$

If we substitute $v = \sinh q$, $\int dv/\sqrt{1+v^2} = \int dq = q = \operatorname{arcsinh} v$, so, separating variables, we obtain

$$\int \frac{dv}{\sqrt{1+v^2}} = \pm \int \frac{du}{\sqrt{1+u^2}}; \quad \text{i.e.,} \quad \operatorname{arcsinh} v = \pm \operatorname{arcsinh} u + c.$$

Since $\sinh(x+y) = \sinh x \cosh y + \cosh x \sinh y$, we obtain

$$v = \sinh(\pm \operatorname{arcsinh} u + c) = \pm(\cosh c)u + (\sinh c)\sqrt{1+u^2}.$$

When $c = 0$, we get $v = \pm u$ (as should be expected on geometric grounds). As c varies through nonzero values, we obtain a family of hyperbolas. Some typical lines of curvature on the saddle surface are indicated in Figure 2.5. ∇

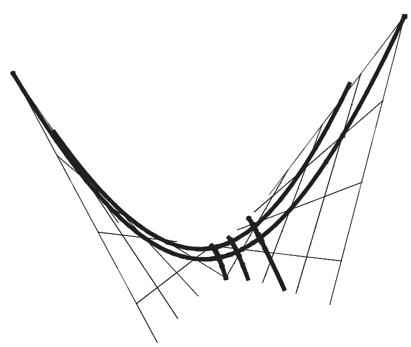


FIGURE 2.5

Definition. Fix $P \in M$. We say P is an *umbilic*[5] if $k_1 = k_2$. If $k_1 = k_2 = 0$, we say P is a *planar point*. If $K = 0$ but P is not a planar point, we say P is a *parabolic point*. If $K > 0$, we say P is an *elliptic point*, and if $K < 0$, we say P is a *hyperbolic point*.

Example 7. On the "outside" of a torus (see Figure 1.3), all the normal slices curve in the same direction, so these are elliptic points. Now imagine laying a plane on top of a torus; it is tangent to the torus along the "top circle," and so the unit normal to the surface stays constant as we move around this curve. For any point P on this circle and $\mathbf{V}$ tangent to the circle, we have $S_P(\mathbf{V}) = -D_{\mathbf{V}}\mathbf{n} = \mathbf{0}$, so $\mathbf{V}$ is a principal direction with corresponding principal curvature 0. Thus, these are parabolic points. On the other hand, consider a point P on the innermost band of the torus. At such a point the surface looks saddle-like; that is, with the unit normal as pictured in Figure 2.6, the horizontal circle (going around the inside of the torus) is a

[5]From the Latin *umbilīcus*, navel.

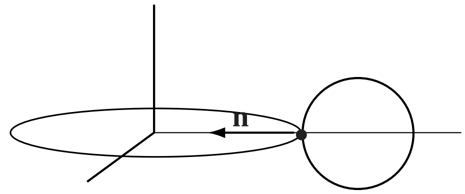


FIGURE 2.6

line of curvature with positive principal curvature, and the vertical circle is a line of curvature with negative principal curvature. Thus, the points on the inside of the torus are hyperbolic points. ∇

Remark. Gauss's original interpretation of Gaussian curvature was the following: Imagine a small curvilinear rectangle $\mathcal{P}$ at $P \in M$ with sides h_1 and h_2 along principal directions. Then, since the principal directions are eigenvectors of the shape operator, the image of $\mathcal{P}$ under the Gauss map is nearly a small curvilinear rectangle at $\mathbf{n}(P) \in \Sigma$ with sides k_1h_1 and k_2h_2. Thus, $K = k_1k_2$ is the factor by which $\mathbf{n}$ distorts signed area as it maps M to Σ. (Note that for a cylinder, the rectangle collapses to a line segment; for a saddle surface, orientation is reversed by $\mathbf{n}$ and so the Gaussian curvature is negative.)

Let's close this section by revisiting our discussion of the curvature of normal slices. Suppose $\boldsymbol{\alpha}$ is an arclength-parametrized curve lying on M with $\boldsymbol{\alpha}(0) = P$ and $\boldsymbol{\alpha}'(0) = \mathbf{V}$. Then the calculation in formula (†) on p. 45 shows that

$$\mathrm{II}_P(\mathbf{V},\mathbf{V}) = \kappa\mathbf{N}\cdot\mathbf{n};$$

i.e., $\mathrm{II}_P(\mathbf{V},\mathbf{V})$ gives the component of the curvature vector $\kappa\mathbf{N}$ of $\boldsymbol{\alpha}$ *normal* to the surface M at P, which we denote by κ_n and call the *normal curvature* of $\boldsymbol{\alpha}$ at P. What is remarkable about this formula is that it shows that the normal curvature depends only on the direction of $\boldsymbol{\alpha}$ at P and otherwise not on the curve. (For the case of the normal slice, the normal curvature is, up to a sign, all the curvature.) What's more, κ_n can be computed just from the second fundamental form II of M. We immediately deduce the following

Proposition 2.5 (Meusnier's Formula)**.** *Let $\boldsymbol{\alpha}$ be a curve on M passing through P with unit tangent vector $\mathbf{V}$. Then*

$$\mathrm{II}_P(\mathbf{V},\mathbf{V}) = \kappa_n = \kappa\cos\phi,$$

where ϕ is the angle between the principal normal, $\mathbf{N}$, of $\boldsymbol{\alpha}$ and the surface normal, $\mathbf{n}$, at P.

In particular, if $\boldsymbol{\alpha}$ is an asymptotic curve, then its normal curvature is 0 at each point. This means that, so long as $\kappa \neq 0$, its principal normal is always orthogonal to the surface normal, i.e., always tangent to the surface.

Example 8. Let's now investigate a very interesting surface, called the *pseudosphere*, as shown in Figure 2.7. It is the surface of revolution obtained by rotating the tractrix (see Example 2 of Chapter 1, Section 1) about the x-axis, and so it is parametrized by

$$\mathbf{x}(u,v) = (u - \tanh u, \operatorname{sech} u\cos v, \operatorname{sech} u\sin v), \qquad u > 0,\ v \in [0,2\pi).$$

Note that the circles (of revolution) are lines of curvature: Either apply Exercise 15 or observe, directly, that the only component of the surface normal that changes as we move around the circle is normal to the circle

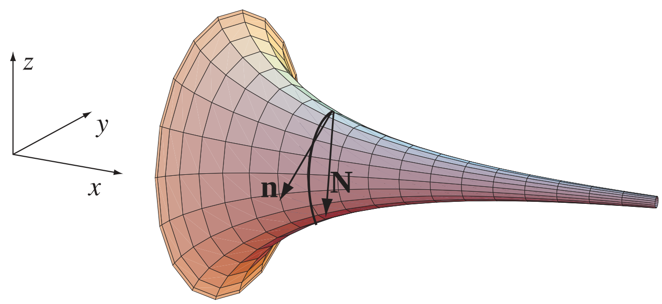


FIGURE 2.7

in the plane of the circle. Similarly, the various tractrices are lines of curvature: In the plane of one tractrix, the surface normal and the curve normal agree.

Now, by Exercise 1.2.5, the curvature of the tractrix is $\kappa = 1/\sinh u$; since $\mathbf{N} = -\mathbf{n}$ along this curve, we have $k_1 = \kappa_n = -1/\sinh u$. Now what about the circles? Here we have $\kappa = 1/\operatorname{sech} u = \cosh u$, but this is not the normal curvature. The angle ϕ between $\mathbf{N}$ and $\mathbf{n}$ is the supplement of the angle θ we see in Figure 1.9 of Chapter 1 (to see why, see Figure 2.8). Thus, by Meusnier's Formula, Proposition 2.5,

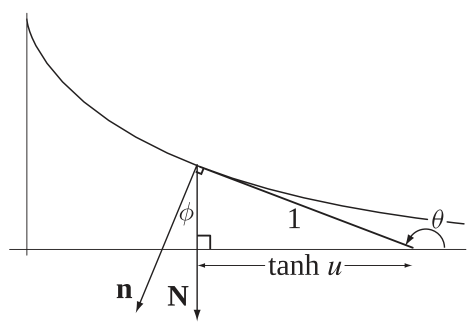


FIGURE 2.8

we have $k_2 = \kappa_n = \kappa \cos\phi = (\cosh u)(\tanh u) = \sinh u$. Amazingly, then, we have $K = k_1 k_2 = (-1/\sinh u)(\sinh u) = -1$. ∇

Example 9. Let's now consider the case of a general surface of revolution, parametrized as in Example 2 of Section 1, by

$$\mathbf{x}(u,v) = \big(f(u)\cos v, f(u)\sin v, g(u)\big),$$

where $f'(u)^2 + g'(u)^2 = 1$. Recall that the u-curves are called *meridians* and the v-curves are called *parallels*. Then

$$\begin{aligned}
\mathbf{x}_u &= \big(f'(u)\cos v, f'(u)\sin v, g'(u)\big)\\
\mathbf{x}_v &= \big(-f(u)\sin v, f(u)\cos v, 0\big)\\
\mathbf{n} &= \big(-g'(u)\cos v, -g'(u)\sin v, f'(u)\big)\\
\mathbf{x}_{uu} &= \big(f''(u)\cos v, f''(u)\sin v, g''(u)\big)\\
\mathbf{x}_{uv} &= \big(-f'(u)\sin v, f'(u)\cos v, 0\big)\\
\mathbf{x}_{vv} &= \big(-f(u)\cos v, -f(u)\sin v, 0\big),
\end{aligned}$$

and so we have

$$E = 1, \quad F = 0, \quad G = f(u)^2, \quad \text{and} \quad \ell = f'(u)g''(u) - f''(u)g'(u), \quad m = 0, \quad n = f(u)g'(u).$$

By Exercise 2.2.1, then $k_1 = f'(u)g''(u) - f''(u)g'(u)$ and $k_2 = g'(u)/f(u)$. Thus,

$$K = k_1k_2 = \big(f'(u)g''(u) - f''(u)g'(u)\big)\frac{g'(u)}{f(u)} = -\frac{f''(u)}{f(u)},$$

since from $f'(u)^2 + g'(u)^2 = 1$ we deduce that $f'(u)f''(u) + g'(u)g''(u) = 0$, and so

$$f'(u)g'(u)g''(u) - f''(u)g'(u)^2 = -(f'(u)^2 + g'(u)^2)f''(u) = -f''(u).$$

Note, as we observed in the special case of Example 8, that on every surface of revolution, the meridians and the parallels are lines of curvature. ∇

EXERCISES 2.2

*1. Check that if there are no umbilic points and the parameter curves are lines of curvature, then $F = m = 0$ and we have the principal curvatures $k_1 = \ell/E$ and $k_2 = n/G$. Conversely, prove that if $F = m = 0$, then the parameter curves are lines of curvature.

$\sharp$2. a. Show that the matrix representing the linear map $S_P : T_P M \to T_P M$ with respect to the basis $\{\mathbf{x}_u, \mathbf{x}_v\}$ is

$$\mathrm{I}_P^{-1}\mathrm{II}_P = \begin{bmatrix} E & F \\ F & G \end{bmatrix}^{-1} \begin{bmatrix} \ell & m \\ m & n \end{bmatrix}.$$

(Hint: Write $S_P(\mathbf{x}_u) = a\mathbf{x}_u + b\mathbf{x}_v$ and $S_P(\mathbf{x}_v) = c\mathbf{x}_u + d\mathbf{x}_v$, and use the definition of ℓ, m, and n to get a system of linear equations for a, b, c, and d.)

b. Deduce that $K = \dfrac{\ell n - m^2}{EG - F^2}$.

3. Compute the second fundamental form II_P of the following parametrized surfaces. Then calculate the matrix of the shape operator, and determine H and K.
   a. the cylinder: $\mathbf{x}(u, v) = (a\cos u, a\sin u, v)$
   *b. the torus: $\mathbf{x}(u, v) = ((a + b\cos u)\cos v, (a + b\cos u)\sin v, b\sin u)$ $(0 < b < a)$
   c. the helicoid: $\mathbf{x}(u, v) = (u\cos v, u\sin v, bv)$
   *d. the catenoid: $\mathbf{x}(u, v) = a(\cosh u\cos v, \cosh u\sin v, u)$
   e. the Mercator parametrization of the sphere: $\mathbf{x}(u, v) = (\operatorname{sech} u\cos v, \operatorname{sech} u\sin v, \tanh u)$
   f. Enneper's surface: $\mathbf{x}(u, v) = (u - u^3/3 + uv^2, v - v^3/3 + u^2v, u^2 - v^2)$

4. Find the principal curvatures, the principal directions, and asymptotic directions (when they exist) for each of the surfaces in Exercise 3. Identify the lines of curvature and asymptotic curves when possible.

*5. Prove by calculation that any one of the helices $\boldsymbol{\alpha}(t) = (a\cos t, a\sin t, bt)$ is an asymptotic curve on the helicoid given in Example 1(b) of Section 1. Also, calculate how the surface normal $\mathbf{n}$ changes as one moves along a ruling, and use this to explain why the rulings are asymptotic curves as well.

*6. Calculate the first and second fundamental forms of the pseudosphere (see Example 8) and check our computations of the principal curvatures and Gaussian curvature.

7. Show that a ruled surface has Gaussian curvature $K \le 0$.

8. a. Prove that the principal directions bisect the asymptotic directions at a hyperbolic point. (Hint: Euler's Formula.)
   b. Prove that if the asymptotic directions of M are orthogonal, then M is minimal. Prove the converse assuming M has no planar points.

9. Let $\kappa_n(\theta)$ denote the normal curvature in the direction making angle θ with the first principal direction.
   a. Show that $H = \dfrac{1}{2\pi}\displaystyle\int_0^{2\pi} \kappa_n(\theta)\,d\theta$.
   b. Show that $H = \dfrac{1}{2}\Big(\kappa_n(\theta) + \kappa_n\big(\theta + \dfrac{\pi}{2}\big)\Big)$ for any θ.
   c. (More challenging) Show that, more generally, for any θ and $m \ge 3$, we have $H = \dfrac{1}{m}\left(\kappa_n(\theta) + \kappa_n\big(\theta + \dfrac{2\pi}{m}\big) + \dots + \kappa_n\big(\theta + \dfrac{2\pi(m-1)}{m}\big)\right)$.

10. Consider the ruled surface M given by $\mathbf{x}(u,v) = (v\cos u, v\sin u, uv)$, $v > 0$.
   a. Describe this surface geometrically.
   b. Find the first and second fundamental forms and the Gaussian curvature of M.
   c. Check that the v-curves are lines of curvature.
   d. Proceeding somewhat as in Example 6, show that the other lines of curvature are given by the equation $v\sqrt{1+u^2} = c$ for various constants c. Show that these curves are the intersection of M with the spheres $x^2 + y^2 + z^2 = c^2$. (It might be fun to use Mathematica to see this explicitly.)

11. The curve $\boldsymbol{\alpha}(t) = \mathbf{x}(u(t), v(t))$ may arise by writing $\dfrac{dv}{du} = \dfrac{v'(t)}{u'(t)}$ and solving a differential equation to relate u and v either explicitly or implicitly.
   a. Show that $\boldsymbol{\alpha}$ is an asymptotic curve if and only if $\ell(u')^2 + 2mu'v' + n(v')^2 = 0$. Thus, if $\ell + 2m\frac{dv}{du} + n\big(\frac{dv}{du}\big)^2 = 0$, then $\boldsymbol{\alpha}$ is an asymptotic curve.
   b. Show that $\boldsymbol{\alpha}$ is a line of curvature if and only if $\begin{vmatrix} Eu' + Fv' & Fu' + Gv' \\ \ell u' + mv' & mu' + nv' \end{vmatrix} = 0$. Give the appropriate condition in terms of dv/du.
   c. Deduce that an alternative condition for $\boldsymbol{\alpha}$ to be a line of curvature is that
$$\begin{vmatrix} (v')^2 & -u'v' & (u')^2 \\ E & F & G \\ \ell & m & n \end{vmatrix} = 0.$$

12. a. Apply Meusnier's Formula to a latitude circle on a sphere of radius a to calculate the normal curvature.
   b. Apply Meusnier's Formula to prove that the curvature of any curve lying on a sphere of radius a satisfies $\kappa \ge 1/a$.

13. Prove or give a counterexample: If M is a surface with Gaussian curvature $K > 0$, then the curvature of any curve $C \subset M$ is everywhere positive. (Remember that, by definition, $\kappa \ge 0$.)

♯14. Suppose that for every $P \in M$, the shape operator S_P is some scalar multiple of the identity, i.e., $S_P(\mathbf{V}) = k(P)\mathbf{V}$ for all $\mathbf{V} \in T_P M$. (Here the scalar $k(P)$ may well depend on the point P.)

a. Differentiate the equations

$$D_{\mathbf{x}_u}\mathbf{n} = \mathbf{n}_u = -k\mathbf{x}_u$$
$$D_{\mathbf{x}_v}\mathbf{n} = \mathbf{n}_v = -k\mathbf{x}_v$$

appropriately to determine k_u and k_v and deduce that k must be constant.

b. We showed in Proposition 2.2 that M is planar when $k = 0$. Show that when $k \neq 0$, M is (a portion of) a sphere.

15. a. Prove that $\boldsymbol{\alpha}$ is a line of curvature in M if and only if $(\mathbf{n}\circ\boldsymbol{\alpha})'(t) = -k(t)\boldsymbol{\alpha}'(t)$, where $k(t)$ is the principal curvature at $\boldsymbol{\alpha}(t)$ in the direction $\boldsymbol{\alpha}'(t)$. (More colloquially, differentiating along the curve $\boldsymbol{\alpha}$, we just write $\mathbf{n}' = -k\boldsymbol{\alpha}'$.)

b. Suppose two surfaces M and M^* intersect along a curve C. Suppose C is a line of curvature in M. Prove that C is a line of curvature in M^* if and only if the angle between M and M^* is constant along C. (In the proof of $\Longleftarrow$, be sure to include the case that M and M^* intersect tangentially along C.)

16. Prove or give a counterexample:

a. If a curve is both an asymptotic curve and a line of curvature, then it must be planar. (Hint: Along an asymptotic curve that is not a line, how is the Frenet frame related to the surface normal?)

b. If a curve is planar and an asymptotic curve, then it must be a line.

17. a. How is the Frenet frame along an asymptotic curve related to the geometry of the surface?

b. Suppose $K(P) < 0$. If C is an asymptotic curve with $\kappa(P) \neq 0$, prove that its torsion satisfies $|\tau(P)| = \sqrt{-K(P)}$. (Hint: If we choose an orthonormal basis $\{\mathbf{U}, \mathbf{V}\}$ for $T_P(M)$ with $\mathbf{U}$ tangent to C, what is the matrix for S_P? See the Remark on p. 47.)

18. Continuing Exercise 17, show that if $K(P) < 0$, then the two asymptotic curves have torsion of opposite signs at P.

19. Suppose $U \subset \mathbb{R}^3$ is open and $\mathbf{x}\colon U \to \mathbb{R}^3$ is a smooth map (of rank 3) so that $\mathbf{x}_u$, $\mathbf{x}_v$, and $\mathbf{x}_w$ are always orthogonal. Then the level surfaces $u = \text{const}$, $v = \text{const}$, $w = \text{const}$ form a *triply orthogonal system* of surfaces.

a. Show that the spherical coordinate mapping $\mathbf{x}(u, v, w) = (u \sin v \cos w, u \sin v \sin w, u \cos v)$ ($u > 0$, $0 < v < \pi$, $0 < w < 2\pi$) furnishes an example.

b. Prove that the curves of intersection of any pair of surfaces from different systems (e.g., $v = \text{const}$ and $w = \text{const}$) are lines of curvature in each of the respective surfaces. (Hint: Differentiate the various equations $\mathbf{x}_u \cdot \mathbf{x}_v = 0$, $\mathbf{x}_v \cdot \mathbf{x}_w = 0$, $\mathbf{x}_u \cdot \mathbf{x}_w = 0$ with respect to the missing variable. What are the shape operators of the various surfaces?)

20. In this exercise we analyze the surfaces of revolution that are minimal. It will be convenient to work with a meridian as a graph ($y = h(u)$, $z = u$) when using the parametrization of surfaces of revolution given in Example 9.

a. Use Exercise 1.2.4 and Proposition 2.5 to show that the principal curvatures are

$$k_1 = -\frac{h''}{(1+h'^2)^{3/2}} \qquad \text{and} \qquad k_2 = \frac{1}{h}\cdot\frac{1}{\sqrt{1+h'^2}}.$$

b. Deduce that $H = 0$ if and only if $h(u)h''(u) = 1 + h'(u)^2$.

c. Solve the differential equation. (Hint: Either substitute $z(u) = \ln h(u)$ or introduce $w(u) = h'(u)$, find dw/dh, and integrate by separating variables.) You should find that $h(u) = \frac{1}{c}\cosh(cu + b)$ for some constants b and c.

21. By choosing coordinates in $\mathbb{R}^3$ appropriately, we may arrange that P is the origin, the tangent plane T_PM is the xy-plane, and the x- and y-axes are in the principal directions at P.

a. Show that in these coordinates M is locally the graph $z = f(x,y) = \frac{1}{2}(k_1x^2 + k_2y^2) + \epsilon(x,y)$, where $\displaystyle\lim_{x,y\to 0}\frac{\epsilon(x,y)}{x^2+y^2} = 0$. (You may start with Taylor's Theorem: If f is $\mathcal{C}^2$, we have

$$\begin{aligned} f(x,y) = f(0,0) + f_x(0,0)x + f_y(0,0)y +& \\ &\tfrac{1}{2}\big(f_{xx}(0,0)x^2 + 2f_{xy}(0,0)xy + f_{yy}(0,0)y^2\big) + \epsilon(x,y), \end{aligned}$$

where $\displaystyle\lim_{x,y\to 0}\frac{\epsilon(x,y)}{x^2+y^2} = 0$.)

b. Show that if P is an elliptic point, then a neighborhood of P in $M \cap T_PM$ is just the origin itself. What happens in the case of a parabolic point?

c. (More challenging) Show that if P is a hyperbolic point, a neighborhood of P in $M \cap T_PM$ is a curve that crosses itself at P and whose tangent directions at P are the asymptotic directions. (Hints: Work in coordinates (x,u) with $y = ux$. Show that in the xu-plane the curve has the equation $0 = g(x,u) = \frac{1}{2}(k_1 + k_2u^2) + h(x,u)$, where $h(0,u) = 0$ for all u, so it consists of *two* ($\mathcal{C}^1$) curves, one passing through $(0, \sqrt{-k_1/k_2})$ and the other through $(0, -\sqrt{-k_1/k_2})$. Show, moreover, that if two curves pass through the same point $(0,u_0)$ in the xu-plane, then the corresponding curves in the xy-plane are tangent at $(0,0)$.[6])

22. Let $P \in M$ be a non-planar point, and if $K \geq 0$, choose the unit normal so that $\ell, n \geq 0$.

a. We define the *Dupin indicatrix* to be the conic in T_PM defined by the equation $\mathrm{II}_P(\mathbf{V},\mathbf{V}) = 1$. Show that if P is an elliptic point, the Dupin indicatrix is an ellipse; if P is a hyperbolic point, the Dupin indicatrix is a hyperbola; and if P is a parabolic point, the Dupin indicatrix is a pair of parallel lines.

b. Show that if P is a hyperbolic point, the asymptotes of the Dupin indicatrix are given by $\mathrm{II}_P(\mathbf{V},\mathbf{V}) = 0$, i.e., the set of asymptotic directions.

c. Suppose M is represented locally near P as in Exercise 21. Show that for small positive values of c, the intersection of M with the plane $z = c$ "looks like" the Dupin indicatrix. How can you make this statement more precise?

23. Suppose the surface M is given near P as a level surface of a smooth function $F\colon \mathbb{R}^3 \to \mathbb{R}$ with $\nabla F(P) \neq 0$. A line $L \subset \mathbb{R}^3$ is said to have (at least) *k-point contact* with M at P if, given any linear parametrization $\boldsymbol{\alpha}$ of L with $\boldsymbol{\alpha}(0) = P$, the function $\mathcal{F} = F\circ\boldsymbol{\alpha}$ vanishes to order $k-1$, i.e.,

[6] Here we have "blown up" the origin in order to keep track of the different tangent directions. The *blowing-up* construction is widely used in topology and algebraic geometry.

$\mathcal{F}(0) = \mathcal{F}'(0) = \cdots = \mathcal{F}^{(k-1)}(0) = 0$. (Such a line is to be visualized as the limit of lines that intersect M at P and at $k-1$ other points that approach P.)

a. Show that L has 2-point contact with M at P if and only if L is tangent to M at P, i.e., $L \subset T_P M$.

b. Show that L has 3-point contact with M at P if and only if L is an asymptotic direction at P. (Hint: It may be helpful to follow the setup of Exercise 21.)

c. (Challenge) Assume P is a hyperbolic point. What does it mean for L to have 4-point contact with M at P?

3. The Codazzi and Gauss Equations and the Fundamental Theorem of Surface Theory

We now wish to proceed towards a deeper understanding of Gaussian curvature. We have to this point considered only the normal components of the second derivatives $\mathbf{x}_{uu}$, $\mathbf{x}_{uv}$, and $\mathbf{x}_{vv}$. Now let's consider them *in toto*. Since $\{\mathbf{x}_u, \mathbf{x}_v, \mathbf{n}\}$ gives a basis for $\mathbb{R}^3$, there are functions Γ^u_{uu}, Γ^v_{uu}, $\Gamma^u_{uv} = \Gamma^u_{vu}$, $\Gamma^v_{uv} = \Gamma^v_{vu}$, Γ^u_{vv}, and Γ^v_{vv} so that

$$(\dagger) \qquad \begin{aligned} \mathbf{x}_{uu} &= \Gamma^u_{uu}\mathbf{x}_u + \Gamma^v_{uu}\mathbf{x}_v + \ell\mathbf{n} \\ \mathbf{x}_{uv} &= \Gamma^u_{uv}\mathbf{x}_u + \Gamma^v_{uv}\mathbf{x}_v + m\mathbf{n} \\ \mathbf{x}_{vv} &= \Gamma^u_{vv}\mathbf{x}_u + \Gamma^v_{vv}\mathbf{x}_v + n\mathbf{n}. \end{aligned}$$

(Note that $\mathbf{x}_{uv} = \mathbf{x}_{vu}$ dictates the symmetries $\Gamma^{\bullet}_{uv} = \Gamma^{\bullet}_{vu}$.) The functions $\Gamma^{\bullet}_{\bullet\bullet}$ are called *Christoffel symbols*.

Example 1. Let's compute the Christoffel symbols for the usual parametrization of the sphere (see Example 1(d) on p. 37). By straightforward calculation we obtain

$$\begin{aligned} \mathbf{x}_u &= (\cos u \cos v, \cos u \sin v, -\sin u) \\ \mathbf{x}_v &= (-\sin u \sin v, \sin u \cos v, 0) \\ \mathbf{x}_{uu} &= (-\sin u \cos v, -\sin u \sin v, -\cos u) = -\mathbf{x}(u,v) \\ \mathbf{x}_{uv} &= (-\cos u \sin v, \cos u \cos v, 0) \\ \mathbf{x}_{vv} &= (-\sin u \cos v, -\sin u \sin v, 0) = -\sin u(\cos v, \sin v, 0). \end{aligned}$$

(Note that the u-curves are great circles, parametrized by arclength, so it is no surprise that the acceleration vector $\mathbf{x}_{uu}$ is inward-pointing of length 1. The v-curves are latitude circles of radius $\sin u$, so, similarly, the acceleration vector $\mathbf{x}_{vv}$ points inwards towards the center of the respective circle.)

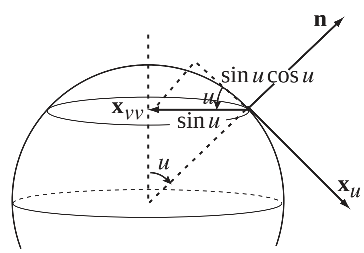


FIGURE 3.1

Since $\mathbf{x}_{uu}$ lies entirely in the direction of $\mathbf{n}$, we have $\Gamma^{u}_{uu} = \Gamma^{v}_{uu} = 0$. Now, by inspection, $\mathbf{x}_{uv} = \cot u\,\mathbf{x}_v$, so $\Gamma^{u}_{uv} = 0$ and $\Gamma^{v}_{uv} = \cot u$. Last, as we can see in Figure 3.1, we have $\mathbf{x}_{vv} = -\sin u \cos u\,\mathbf{x}_u - \sin^2 u\,\mathbf{n}$, so $\Gamma^{u}_{vv} = -\sin u \cos u$ and $\Gamma^{v}_{vv} = 0$. ∇

Now, dotting the equations in (†) with $\mathbf{x}_u$ and $\mathbf{x}_v$ gives

$$\begin{aligned}
\mathbf{x}_{uu}\cdot\mathbf{x}_u &= \Gamma^{u}_{uu}E + \Gamma^{v}_{uu}F\\
\mathbf{x}_{uu}\cdot\mathbf{x}_v &= \Gamma^{u}_{uu}F + \Gamma^{v}_{uu}G\\[1ex]
\mathbf{x}_{uv}\cdot\mathbf{x}_u &= \Gamma^{u}_{uv}E + \Gamma^{v}_{uv}F\\
\mathbf{x}_{uv}\cdot\mathbf{x}_v &= \Gamma^{u}_{uv}F + \Gamma^{v}_{uv}G\\[1ex]
\mathbf{x}_{vv}\cdot\mathbf{x}_u &= \Gamma^{u}_{vv}E + \Gamma^{v}_{vv}F\\
\mathbf{x}_{vv}\cdot\mathbf{x}_v &= \Gamma^{u}_{vv}F + \Gamma^{v}_{vv}G.
\end{aligned}$$

Now observe that

$$(\spadesuit)\qquad \begin{aligned}
\mathbf{x}_{uu}\cdot\mathbf{x}_u &= \tfrac12(\mathbf{x}_u\cdot\mathbf{x}_u)_u = \tfrac12 E_u\\
\mathbf{x}_{uv}\cdot\mathbf{x}_u &= \tfrac12(\mathbf{x}_u\cdot\mathbf{x}_u)_v = \tfrac12 E_v\\
\mathbf{x}_{uv}\cdot\mathbf{x}_v &= \tfrac12(\mathbf{x}_v\cdot\mathbf{x}_v)_u = \tfrac12 G_u\\
\mathbf{x}_{uu}\cdot\mathbf{x}_v &= (\mathbf{x}_u\cdot\mathbf{x}_v)_u - \mathbf{x}_u\cdot\mathbf{x}_{uv} = F_u - \tfrac12 E_v\\
\mathbf{x}_{vv}\cdot\mathbf{x}_u &= (\mathbf{x}_u\cdot\mathbf{x}_v)_v - \mathbf{x}_{uv}\cdot\mathbf{x}_v = F_v - \tfrac12 G_u\\
\mathbf{x}_{vv}\cdot\mathbf{x}_v &= \tfrac12(\mathbf{x}_v\cdot\mathbf{x}_v)_v = \tfrac12 G_v
\end{aligned}$$

Thus, we can rewrite our equations as follows:

$$(\ddagger)\qquad \begin{aligned}
\begin{bmatrix} E & F\\ F & G\end{bmatrix}\begin{bmatrix}\Gamma^{u}_{uu}\\ \Gamma^{v}_{uu}\end{bmatrix} = \begin{bmatrix}\tfrac12 E_u\\ F_u - \tfrac12 E_v\end{bmatrix} &\implies \begin{bmatrix}\Gamma^{u}_{uu}\\ \Gamma^{v}_{uu}\end{bmatrix} = \begin{bmatrix} E & F\\ F & G\end{bmatrix}^{-1}\begin{bmatrix}\tfrac12 E_u\\ F_u - \tfrac12 E_v\end{bmatrix}\\
\begin{bmatrix} E & F\\ F & G\end{bmatrix}\begin{bmatrix}\Gamma^{u}_{uv}\\ \Gamma^{v}_{uv}\end{bmatrix} = \begin{bmatrix}\tfrac12 E_v\\ \tfrac12 G_u\end{bmatrix} &\implies \begin{bmatrix}\Gamma^{u}_{uv}\\ \Gamma^{v}_{uv}\end{bmatrix} = \begin{bmatrix} E & F\\ F & G\end{bmatrix}^{-1}\begin{bmatrix}\tfrac12 E_v\\ \tfrac12 G_u\end{bmatrix}\\
\begin{bmatrix} E & F\\ F & G\end{bmatrix}\begin{bmatrix}\Gamma^{u}_{vv}\\ \Gamma^{v}_{vv}\end{bmatrix} = \begin{bmatrix}F_v - \tfrac12 G_u\\ \tfrac12 G_v\end{bmatrix} &\implies \begin{bmatrix}\Gamma^{u}_{vv}\\ \Gamma^{v}_{vv}\end{bmatrix} = \begin{bmatrix} E & F\\ F & G\end{bmatrix}^{-1}\begin{bmatrix}F_v - \tfrac12 G_u\\ \tfrac12 G_v\end{bmatrix}.
\end{aligned}$$

What is quite remarkable about these formulas is that the Christoffel symbols, which tell us about the tangential component of the second derivatives $\mathbf{x}_{\bullet\bullet}$, can be computed *just* from knowing E, F, and G, i.e., the first fundamental form.

Example 2. Let's now recompute the Christoffel symbols of the unit sphere and compare our answers with Example 1. Since $E = 1$, $F = 0$, and $G = \sin^2 u$, we have

$$\begin{aligned}
\begin{bmatrix}\Gamma^{u}_{uu}\\ \Gamma^{v}_{uu}\end{bmatrix} &= \begin{bmatrix}1 & 0\\ 0 & \csc^2 u\end{bmatrix}\begin{bmatrix}0\\ 0\end{bmatrix} = \begin{bmatrix}0\\ 0\end{bmatrix}\\
\begin{bmatrix}\Gamma^{u}_{uv}\\ \Gamma^{v}_{uv}\end{bmatrix} &= \begin{bmatrix}1 & 0\\ 0 & \csc^2 u\end{bmatrix}\begin{bmatrix}0\\ \sin u\cos u\end{bmatrix} = \begin{bmatrix}0\\ \cot u\end{bmatrix}
\end{aligned}$$

$$\begin{bmatrix} \Gamma^u_{vv} \\ \Gamma^v_{vv} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & \csc^2 u \end{bmatrix} \begin{bmatrix} -\sin u \cos u \\ 0 \end{bmatrix} = \begin{bmatrix} -\sin u \cos u \\ 0 \end{bmatrix}.$$

Thus, the only nonzero Christoffel symbols are $\Gamma^v_{uv} = \Gamma^v_{vu} = \cot u$ and $\Gamma^u_{vv} = -\sin u \cos u$, as before. ∇

By Exercise 2.2.2, the matrix of the shape operator S_P with respect to the basis $\{\mathbf{x}_u, \mathbf{x}_v\}$ is

$$\begin{bmatrix} a & c \\ b & d \end{bmatrix} = \begin{bmatrix} E & F \\ F & G \end{bmatrix}^{-1} \begin{bmatrix} \ell & m \\ m & n \end{bmatrix} = \frac{1}{EG - F^2} \begin{bmatrix} \ell G - mF & mG - nF \\ -\ell F + mE & -mF + nE \end{bmatrix}.$$

Note that these coefficients tell us the derivatives of $\mathbf{n}$ with respect to u and v:

$$(\dagger\dagger) \qquad \begin{aligned} \mathbf{n}_u &= D_{\mathbf{x}_u}\mathbf{n} = -S_P(\mathbf{x}_u) = -(a\mathbf{x}_u + b\mathbf{x}_v) \\ \mathbf{n}_v &= D_{\mathbf{x}_v}\mathbf{n} = -S_P(\mathbf{x}_v) = -(c\mathbf{x}_u + d\mathbf{x}_v). \end{aligned}$$

We now differentiate the equations ($\dagger$) again and use equality of mixed partial derivatives. To start, we have

$$\begin{aligned} \mathbf{x}_{uuv} &= (\Gamma^u_{uu})_v\mathbf{x}_u + \Gamma^u_{uu}\mathbf{x}_{uv} + (\Gamma^v_{uu})_v\mathbf{x}_v + \Gamma^v_{uu}\mathbf{x}_{vv} + \ell_v\mathbf{n} + \ell\mathbf{n}_v \\ &= (\Gamma^u_{uu})_v\mathbf{x}_u + \Gamma^u_{uu}\big(\Gamma^u_{uv}\mathbf{x}_u + \Gamma^v_{uv}\mathbf{x}_v + m\mathbf{n}\big) + (\Gamma^v_{uu})_v\mathbf{x}_v + \Gamma^v_{uu}\big(\Gamma^u_{vv}\mathbf{x}_u + \Gamma^v_{vv}\mathbf{x}_v + n\mathbf{n}\big) \\ &\qquad + \ell_v\mathbf{n} - \ell(c\mathbf{x}_u + d\mathbf{x}_v) \\ &= \big((\Gamma^u_{uu})_v + \Gamma^u_{uu}\Gamma^u_{uv} + \Gamma^v_{uu}\Gamma^u_{vv} - \ell c\big)\mathbf{x}_u + \big((\Gamma^v_{uu})_v + \Gamma^u_{uu}\Gamma^v_{uv} + \Gamma^v_{uu}\Gamma^v_{vv} - \ell d\big)\mathbf{x}_v \\ &\qquad + \big(\Gamma^u_{uu}m + \Gamma^v_{uu}n + \ell_v\big)\mathbf{n}, \end{aligned}$$

and, similarly,

$$\begin{aligned} \mathbf{x}_{uvu} &= \big((\Gamma^u_{uv})_u + \Gamma^u_{uv}\Gamma^u_{uu} + \Gamma^v_{uv}\Gamma^u_{uv} - ma\big)\mathbf{x}_u + \big((\Gamma^v_{uv})_u + \Gamma^u_{uv}\Gamma^v_{uu} + \Gamma^v_{uv}\Gamma^v_{uv} - mb\big)\mathbf{x}_v \\ &\qquad + \big(\ell\Gamma^u_{uv} + m\Gamma^v_{uv} + m_u\big)\mathbf{n}. \end{aligned}$$

Since $\mathbf{x}_{uuv} = \mathbf{x}_{uvu}$, we compare the indicated components and obtain:

$$(\diamondsuit) \qquad \begin{aligned} (\mathbf{x}_u)&: \quad (\Gamma^u_{uu})_v + \Gamma^v_{uu}\Gamma^u_{vv} - \ell c = (\Gamma^u_{uv})_u + \Gamma^v_{uv}\Gamma^u_{uv} - ma \\ (\mathbf{x}_v)&: \quad (\Gamma^v_{uu})_v + \Gamma^u_{uu}\Gamma^v_{uv} + \Gamma^v_{uu}\Gamma^v_{vv} - \ell d = (\Gamma^v_{uv})_u + \Gamma^u_{uv}\Gamma^v_{uu} + \Gamma^v_{uv}\Gamma^v_{uv} - mb \\ (\mathbf{n})&: \quad \ell_v + m\Gamma^u_{uu} + n\Gamma^v_{uu} = m_u + \ell\Gamma^u_{uv} + m\Gamma^v_{uv}. \end{aligned}$$

Analogously, comparing the indicated components of $\mathbf{x}_{uvv} = \mathbf{x}_{vvu}$, we find:

$$\begin{aligned} (\mathbf{x}_u)&: \quad (\Gamma^u_{uv})_v + \Gamma^u_{uv}\Gamma^u_{uv} + \Gamma^v_{uv}\Gamma^u_{vv} - mc = (\Gamma^u_{vv})_u + \Gamma^u_{vv}\Gamma^u_{uu} + \Gamma^v_{vv}\Gamma^u_{uv} - na \\ (\mathbf{x}_v)&: \quad (\Gamma^v_{uv})_v + \Gamma^u_{uv}\Gamma^v_{uv} - md = (\Gamma^v_{vv})_u + \Gamma^u_{vv}\Gamma^v_{uu} - nb \\ (\mathbf{n})&: \quad m_v + m\Gamma^u_{uv} + n\Gamma^v_{uv} = n_u + \ell\Gamma^u_{vv} + m\Gamma^v_{vv}. \end{aligned}$$

The two equations coming from the normal component give us the

Codazzi equations

$$\boxed{\begin{aligned} \ell_v - m_u &= \ell\Gamma^u_{uv} + m\big(\Gamma^v_{uv} - \Gamma^u_{uu}\big) - n\Gamma^v_{uu} \\ m_v - n_u &= \ell\Gamma^u_{vv} + m\big(\Gamma^v_{vv} - \Gamma^u_{uv}\big) - n\Gamma^v_{uv}. \end{aligned}}$$

Using $K = \dfrac{\ell n - m^2}{EG - F^2}$ and the formulas above for a, b, c, and d, the four equations involving the $\mathbf{x}_u$ and $\mathbf{x}_v$ components yield the

Gauss equations

$$\begin{aligned}
EK &= \big(\Gamma^v_{uu}\big)_v - \big(\Gamma^v_{uv}\big)_u + \Gamma^u_{uu}\Gamma^v_{uv} + \Gamma^v_{uu}\Gamma^v_{vv} - \Gamma^u_{uv}\Gamma^v_{uu} - \big(\Gamma^v_{uv}\big)^2\\
FK &= \big(\Gamma^u_{uv}\big)_u - \big(\Gamma^u_{uu}\big)_v + \Gamma^v_{uv}\Gamma^u_{uv} - \Gamma^v_{uu}\Gamma^u_{vv}\\
FK &= \big(\Gamma^v_{uv}\big)_v - \big(\Gamma^v_{vv}\big)_u + \Gamma^u_{uv}\Gamma^v_{uv} - \Gamma^u_{vv}\Gamma^v_{uu}\\
GK &= \big(\Gamma^u_{vv}\big)_u - \big(\Gamma^u_{uv}\big)_v + \Gamma^u_{vv}\Gamma^u_{uu} + \Gamma^v_{vv}\Gamma^u_{uv} - \big(\Gamma^u_{uv}\big)^2 - \Gamma^v_{uv}\Gamma^u_{vv}.
\end{aligned}$$

For example, to derive the first, we use the equation ($\diamondsuit$) above:

$$\begin{aligned}
\big(\Gamma^v_{uu}\big)_v - \big(\Gamma^v_{uv}\big)_u + \Gamma^u_{uu}\Gamma^v_{uv} + \Gamma^v_{uu}\Gamma^v_{vv} - \Gamma^u_{uv}\Gamma^v_{uu} - \big(\Gamma^v_{uv}\big)^2 &= \ell d - mb\\
&= \frac{1}{EG-F^2}\big(\ell(-mF+nE) + m(\ell F - mE)\big) = \frac{E(\ell n - m^2)}{EG-F^2} = EK.
\end{aligned}$$

In an orthogonal parametrization ($F = 0$), we leave it to the reader to check in Exercise 3 that

$$(*) \qquad K = -\frac{1}{2\sqrt{EG}}\left(\Big(\frac{E_v}{\sqrt{EG}}\Big)_v + \Big(\frac{G_u}{\sqrt{EG}}\Big)_u\right).$$

One of the crowning results of local differential geometry is the following

Theorem 3.1 (Gauss's Theorema Egregium)**.** *The Gaussian curvature is determined by only the first fundamental form* I. *That is, K can be computed from just E, F, G, and their first and second partial derivatives.*

Proof. From any of the Gauss equations, we see that K can be computed by knowing any one of E, F, and G, together with the Christoffel symbols and their derivatives. But the equations (‡) show that the Christoffel symbols (and hence any of their derivatives) can be calculated in terms of E, F, and G and their partial derivatives. $\square$

Corollary 3.2. *If two surfaces are locally isometric, their Gaussian curvatures at corresponding points are equal.*

For example, the plane and cylinder are locally isometric, and hence the cylinder (as we well know) is flat. We now conclude that since the Gaussian curvature of a sphere is nonzero, a sphere cannot be locally isometric to a plane. Thus, there is no way to map the earth "faithfully" (preserving distance)—even locally—on a piece of paper. In some sense, the Mercator projection (see Exercise 2.1.13) is the best we can do, for, although it distorts distances, it does preserve angles.

The Codazzi and Gauss equations are rather opaque, to say the least. We obtained the convenient equation ($*$) for the Gaussian curvature from the Gauss equations. To give a bit more insight into the meaning of the Codazzi equations, we have the following

Lemma 3.3. *Suppose* $\mathbf{x}$ *is a parametrization for which the u- and v-curves are lines of curvature, with respective principal curvatures k_1 and k_2. Then we have*

$$(\star) \qquad (k_1)_v = \frac{E_v}{2E}(k_2 - k_1) \qquad \text{and} \qquad (k_2)_u = \frac{G_u}{2G}(k_1 - k_2).$$

Proof. By Exercise 2.2.1, $\ell = k_1 E$, $n = k_2 G$, and $F = m = 0$. By the first Codazzi equation and the equations (‡) on p. 58, we have

$$(k_1)_v E + k_1 E_v = \ell_v = k_1 E \Gamma^u_{uv} - k_2 G \Gamma^v_{uu} = \tfrac{1}{2} E_v (k_1 + k_2),$$

and so

$$(k_1)_v = \frac{E_v}{2E}(k_2 - k_1).$$

The other formula follows similarly from the second Codazzi equation. □

Let's now apply the Codazzi equations to prove a rather striking result about the general surface with $K = 0$ everywhere.

Proposition 3.4. *Suppose M is a flat surface with no planar points. Then M is a ruled surface whose tangent plane is constant along the rulings.*

Proof. Since M has no planar points, we can choose $k_1 = 0$ and $k_2 \neq 0$ everywhere. Then by Theorem 3.3 of the Appendix, there is a local parametrization of M so that the u-curves are the first lines of curvature and the v-curves are the second lines of curvature. This means first of all that $F = m = 0$. (See Exercise 2.2.1.) Now, since $k_1 = 0$, for any $P \in M$ we have $S_P(\mathbf{x}_u) = \mathbf{0}$, and so $\mathbf{n}_u = \mathbf{0}$ everywhere and $\mathbf{n}$ is constant along the u-curves. We also observe that $\ell = \mathrm{II}(\mathbf{x}_u, \mathbf{x}_u) = -S_P(\mathbf{x}_u) \cdot \mathbf{x}_u = 0$.

We now want to show that the u-curves are in fact *lines*. Since $k_1 = 0$ everywhere, $(k_1)_v = 0$ and, since $k_2 \neq k_1$, we infer from Lemma 3.3 that $E_v = 0$. From the equations (‡) it now follows that $\Gamma^v_{uu} = 0$. Thus,

$$\mathbf{x}_{uu} = \Gamma^u_{uu}\mathbf{x}_u + \Gamma^v_{uu}\mathbf{x}_v + \ell\mathbf{n} = \Gamma^u_{uu}\mathbf{x}_u$$

is just a multiple of $\mathbf{x}_u$. Thus, the tangent vector $\mathbf{x}_u$ never changes direction as we move along the u-curves, and this means that the u-curves must be lines. In conclusion, we have a ruled surface whose tangent plane is constant along rulings. □

Remark. Flat ruled surfaces are often called *developable*. (See Exercise 10 and Exercise 2.1.12.) The terminology comes from the fact that they can be rolled out—or "developed"—onto a plane.

Next we prove a striking *global* result about compact surfaces. (Recall that a subset of $\mathbb{R}^3$ is *compact* if it is closed and bounded. The salient feature of compact sets is the maximum value theorem: A continuous real-valued function on a compact set achieves its maximum and minimum values.) We begin with a straightforward

Proposition 3.5. *Suppose $M \subset \mathbb{R}^3$ is a compact surface. Then there is a point $P \in M$ with $K(P) > 0$.*

Proof. Because M is compact, the continuous function $f(\mathbf{x}) = \|\mathbf{x}\|$ achieves its maximum at some point of M, and so there is a point $P \in M$ farthest from the origin (which may or may not be inside M), as indicated in Figure 3.2. Let $f(P) = R$. As Exercise 1.2.7 shows, the curvature of any curve $\boldsymbol{\alpha} \subset M$ at P is at least $1/R$. Applying this to any normal section of M at P and choosing the unit normal $\mathbf{n}$ to be inward-pointing, we deduce that every normal curvature of M at P is at least $1/R$. It follows that $K(P) \geq 1/R^2 > 0$. (That is, M is at least as curved at P as the circumscribed sphere of radius R tangent to M at P.) □

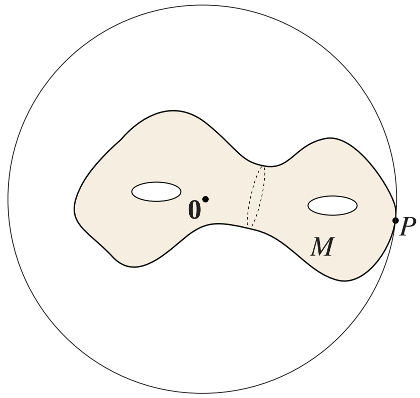


FIGURE 3.2

The reader is asked in Exercise 19 to find surfaces of revolution of constant curvature. There are, interestingly, many nonobvious examples. However, if we restrict ourselves to smooth, compact surfaces, we have the following beautiful

Theorem 3.6 (Liebmann)**.** *If M is a smooth, compact surface of constant Gaussian curvature K, then $K > 0$ and M must be a sphere of radius $1/\sqrt{K}$.*

We will need the following

Lemma 3.7 (Hilbert)**.** *Suppose P is not an umbilic point and $k_1(P) > k_2(P)$. Suppose k_1 has a local maximum at P and k_2 has a local minimum at P. Then $K(P) \leq 0$.*

Proof. We work in a "principal" coordinate parametrization[7] near P, so that the u-curves are lines of curvature with principal curvature k_1 and the v-curves are lines of curvature with principal curvature k_2. Since $k_1 \neq k_2$ and $(k_1)_v = (k_2)_u = 0$ at P, it follows from Lemma 3.3 that $E_v = G_u = 0$ at P.

Differentiating the equations ($\star$), and remembering that $(k_1)_u = (k_2)_v = 0$ at P as well, we have *at* P:

$$(k_1)_{vv} = \frac{E_{vv}}{2E}(k_2 - k_1) \leq 0 \qquad \text{(because } k_1 \text{ has a local maximum at } P\text{)}$$
$$(k_2)_{uu} = \frac{G_{uu}}{2G}(k_1 - k_2) \geq 0 \qquad \text{(because } k_2 \text{ has a local minimum at } P\text{),}$$

and so $E_{vv} \geq 0$ and $G_{uu} \geq 0$ at P. Using the equation ($*$) for the Gaussian curvature on p. 60, we see similarly that *at* P

$$K = -\frac{1}{2EG}\big(E_{vv} + G_{uu}\big),$$

as all the remaining terms involve E_v and G_u. So we conclude that $K(P) \leq 0$, as desired. $\square$

Proof of Theorem 3.6. By Proposition 3.5, there is a point where M is positively curved, and since the Gaussian curvature is constant, we must have $K > 0$. If every point is umbilic, then by Exercise 2.2.14, we know that M is a sphere. If there is some non-umbilic point, the larger principal curvature, k_1, achieves its maximum value at some point P because M is compact. Then, since $K = k_1 k_2$ is constant, the function $k_2 = K/k_1$ must achieve its minimum at P. Since P is necessarily a non-umbilic point (why?), it follows from Lemma 3.7 that $K(P) \leq 0$, which is a contradiction. $\square$

[7]Since locally there are no umbilic points, the existence of such a parametrization is an immediate consequence of Theorem 3.3 of the Appendix.

Remark. H. Hopf proved a stronger result, which requires techniques from complex analysis: If M is a compact surface topologically equivalent to a sphere and having constant *mean* curvature, then M must be a sphere.

We conclude this section with the analogue of Theorem 3.1 of Chapter 1.

Theorem 3.8 (Fundamental Theorem of Surface Theory)**.** Uniqueness: *Two parametrized surfaces* $\mathbf{x}, \mathbf{x}^*: U \to \mathbb{R}^3$ *are congruent (i.e., differ by a rigid motion) if and only if* $\mathrm{I} = \mathrm{I}^*$ *and* $\mathrm{II} = \pm\mathrm{II}^*$. Existence: *Moreover, given differentiable functions E, F, G, ℓ, m, and n with $E > 0$ and $EG - F^2 > 0$ and satisfying the Codazzi and Gauss equations, there exists (locally) a parametrized surface $\mathbf{x}(u, v)$ with the respective* I *and* II.

Proof. The existence statement requires some theorems from partial differential equations beyond our reach at this stage. The uniqueness statement, however, is much like the proof of Theorem 3.1 of Chapter 1. (The main technical difference is that we no longer are lucky enough to be working with an *orthonormal* basis at each point, as we were with the Frenet frame.)

First, suppose $\mathbf{x}^* = \Psi \circ \mathbf{x}$ for some rigid motion $\Psi: \mathbb{R}^3 \to \mathbb{R}^3$ (i.e., $\Psi(\mathbf{x}) = A\mathbf{x} + \mathbf{b}$ for some $\mathbf{b} \in \mathbb{R}^3$ and some 3×3 orthogonal matrix A). Since a translation doesn't change partial derivatives, we may assume that $\mathbf{b} = \mathbf{0}$. Now, since orthogonal matrices preserve length and dot product, we have $E^* = \|\mathbf{x}_u^*\|^2 = \|A\mathbf{x}_u\|^2 = \|\mathbf{x}_u\|^2 = E$, etc., so $\mathrm{I} = \mathrm{I}^*$. If $\det A > 0$, then $\mathbf{n}^* = A\mathbf{n}$, whereas if $\det A < 0$, then $\mathbf{n}^* = -A\mathbf{n}$. Thus, $\ell^* = \mathbf{x}_{uu}^* \cdot \mathbf{n}^* = A\mathbf{x}_{uu} \cdot (\pm A\mathbf{n}) = \pm\ell$, the positive sign holding when $\det A > 0$ and the negative when $\det A < 0$. Thus, $\mathrm{II}^* = \mathrm{II}$ if $\det A > 0$ and $\mathrm{II}^* = -\mathrm{II}$ if $\det A < 0$.

Conversely, suppose $\mathrm{I} = \mathrm{I}^*$ and $\mathrm{II} = \pm\mathrm{II}^*$. By composing $\mathbf{x}^*$ with a reflection, if necessary, we may assume that $\mathrm{II} = \mathrm{II}^*$. Now we need the following

Lemma 3.9. *Suppose $\boldsymbol{\alpha}$ and $\boldsymbol{\alpha}^*$ are smooth functions on $[0, b]$, $\mathbf{v}_1\mathbf{v}_2\mathbf{v}_3$ and $\mathbf{v}_1^*\mathbf{v}_2^*\mathbf{v}_3^*$ are smoothly varying bases for $\mathbb{R}^3$, also defined on $[0, b]$, so that*

$$\mathbf{v}_i(t) \cdot \mathbf{v}_j(t) = \mathbf{v}_i^*(t) \cdot \mathbf{v}_j^*(t) = g_{ij}(t), \qquad i, j = 1, 2, 3,$$

$$\boldsymbol{\alpha}'(t) = \sum_{i=1}^{3} p_i(t)\mathbf{v}_i(t) \qquad \textit{and} \qquad \boldsymbol{\alpha}^{*\prime}(t) = \sum_{i=1}^{3} p_i(t)\mathbf{v}_i^*(t),$$

$$\mathbf{v}_j'(t) = \sum_{i=1}^{3} q_{ij}\mathbf{v}_i(t) \qquad \textit{and} \qquad \mathbf{v}_j^{*\prime}(t) = \sum_{i=1}^{3} q_{ij}\mathbf{v}_i^*(t), \quad j = 1, 2, 3.$$

(Note that the coefficient functions p_i and q_{ij} are the same for both the starred and unstarred equations.) If $\boldsymbol{\alpha}(0) = \boldsymbol{\alpha}^*(0)$ and $\mathbf{v}_i(0) = \mathbf{v}_i^*(0)$, $i = 1, 2, 3$, then $\boldsymbol{\alpha}(t) = \boldsymbol{\alpha}^*(t)$ and $\mathbf{v}_i(t) = \mathbf{v}_i^*(t)$ for all $t \in [0, b]$, $i = 1, 2, 3$.

Fix a point $\mathbf{u}_0 \in U$. By composing $\mathbf{x}^*$ with a rigid motion, we may assume that *at* $\mathbf{u}_0$ we have $\mathbf{x} = \mathbf{x}^*$, $\mathbf{x}_u = \mathbf{x}_u^*$, $\mathbf{x}_v = \mathbf{x}_v^*$, and $\mathbf{n} = \mathbf{n}^*$ (why?). Choose an arbitrary $\mathbf{u}_1 \in U$, and join $\mathbf{u}_0$ to $\mathbf{u}_1$ by a path $\mathbf{u}(t)$, $t \in [0, b]$, and apply the lemma with $\boldsymbol{\alpha} = \mathbf{x} \circ \mathbf{u}$, $\mathbf{v}_1 = \mathbf{x}_u \circ \mathbf{u}$, $\mathbf{v}_2 = \mathbf{x}_v \circ \mathbf{u}$, $\mathbf{v}_3 = \mathbf{n} \circ \mathbf{u}$, $p_i = u_i'$, and the q_{ij} prescribed by the equations (†) and (††). Since $\mathrm{I} = \mathrm{I}^*$ and $\mathrm{II} = \mathrm{II}^*$, the same equations hold for $\boldsymbol{\alpha}^* = \mathbf{x}^* \circ \mathbf{u}$, and so $\mathbf{x}(\mathbf{u}_1) = \mathbf{x}^*(\mathbf{u}_1)$ as desired. That is, the two parametrized surfaces are identical. $\square$

Proof of Lemma 3.9. Introduce the matrix function of t

$$M(t) = \begin{bmatrix} | & | & | \\ \mathbf{v}_1(t) & \mathbf{v}_2(t) & \mathbf{v}_3(t) \\ | & | & | \end{bmatrix},$$

and analogously for $M^*(t)$. Then the displayed equations in the statement of the Lemma can be written as

$$M'(t) = M(t)Q(t) \qquad \text{and} \qquad M^{*\prime}(t) = M^*(t)Q(t).$$

On the other hand, we have $M(t)^\mathsf{T}M(t) = G(t)$. Since the $\mathbf{v}_i(t)$ form a basis for $\mathbb{R}^3$ for each t, we know the matrix G is invertible. Now, differentiating the equation $G(t)G^{-1}(t) = I$ yields $(G^{-1})'(t) = -G^{-1}(t)G'(t)G^{-1}(t)$, and differentiating the equation $G(t) = M(t)^\mathsf{T}M(t)$ yields $G'(t) = M'(t)^\mathsf{T}M(t) + M(t)^\mathsf{T}M'(t) = Q(t)^\mathsf{T}G(t) + G(t)Q(t)$. Now consider

$$\begin{aligned}
(M^*G^{-1}M^\mathsf{T})'(t) &= M^{*\prime}(t)G(t)^{-1}M(t)^\mathsf{T} + M^*(t)(G^{-1})'(t)M(t)^\mathsf{T} + M^*(t)G(t)^{-1}M'(t)^\mathsf{T} \\
&= M^*(t)Q(t)G(t)^{-1}M(t)^\mathsf{T} + M^*(t)\big(-G(t)^{-1}G'(t)G(t)^{-1}\big)M(t)^\mathsf{T} \\
&\qquad + M^*(t)G(t)^{-1}Q(t)^\mathsf{T}M(t)^\mathsf{T} \\
&= M^*(t)Q(t)G(t)^{-1}M(t)^\mathsf{T} - M^*(t)G(t)^{-1}Q(t)^\mathsf{T}M(t)^\mathsf{T} - M^*(t)Q(t)G(t)^{-1}M(t)^\mathsf{T} \\
&\qquad + M^*(t)G(t)^{-1}Q(t)^\mathsf{T}M(t)^\mathsf{T} = \mathrm{O}.
\end{aligned}$$

Since $M(0) = M^*(0)$, we have $M^*(0)G(0)^{-1}M(0)^\mathsf{T} = M(0)M(0)^{-1}M(0)^{\mathsf{T}-1}M(0)^\mathsf{T} = I$, and so $M^*(t)G(t)^{-1}M(t)^\mathsf{T} = I$ for all $t \in [0,b]$. It follows that $M^*(t) = M(t)$ for all $t \in [0,b]$, and so $\boldsymbol{\alpha}^{*\prime}(t) - \boldsymbol{\alpha}'(t) = \mathbf{0}$ for all t as well. Since $\boldsymbol{\alpha}^*(0) = \boldsymbol{\alpha}(0)$, it follows that $\boldsymbol{\alpha}^*(t) = \boldsymbol{\alpha}(t)$ for all $t \in [0,b]$, as we wished to establish. ☐

EXERCISES 2.3

1. Calculate the Christoffel symbols for a cone, $\mathbf{x}(u,v) = (u\cos v, u\sin v, u)$, both directly (as in Example 1) and by using the formulas (‡).
2. Calculate the Christoffel symbols for the following parametrized surfaces. Then check in each case that the Codazzi equations and the first Gauss equation hold.
   a. the plane, parametrized by polar coordinates: $\mathbf{x}(u,v) = (u\cos v, u\sin v, 0)$
   b. a helicoid: $\mathbf{x}(u,v) = (u\cos v, u\sin v, v)$
   ♯c. a cone: $\mathbf{x}(u,v) = (u\cos v, u\sin v, cu)$, $c \neq 0$
   ♯*d. a surface of revolution: $\mathbf{x}(u,v) = \big(f(u)\cos v, f(u)\sin v, g(u)\big)$, with $f'(u)^2 + g'(u)^2 = 1$
3. Use the first Gauss equation to derive the formula $(*)$ given on p. 60 for Gaussian curvature.
4. Check the Gaussian curvature of the sphere using the formula $(*)$ on p. 60.
5. Check that for a parametrized surface with $E = G = \lambda(u,v)$ and $F = 0$, the Gaussian curvature is given by $K = -\dfrac{1}{2\lambda}\nabla^2(\ln\lambda)$. (Here $\nabla^2 f = \dfrac{\partial^2 f}{\partial u^2} + \dfrac{\partial^2 f}{\partial v^2}$ is the Laplacian of f.)
6. Prove there is no *compact* minimal surface $M \subset \mathbb{R}^3$.

7. Decide whether there is a parametrized surface $\mathbf{x}(u, v)$ with
   a. $E = G = 1$, $F = 0$, $\ell = 1 = -n$, $m = 0$
   b. $E = G = 1$, $F = 0$, $\ell = e^u = n$, $m = 0$
   c. $E = 1$, $F = 0$, $G = \cos^2 u$, $\ell = \cos^2 u$, $m = 0$, $n = 1$

8. a. Modify the proof of Theorem 3.6 to prove that a smooth, compact surface with $K > 0$ and constant mean curvature is a sphere.
   b. Give an example to show that the result of Lemma 3.7 fails if we assume k_1 has a local minimum and k_2 has a local maximum at P.

9. Give examples of (locally) *non-congruent* parametrized surfaces $\mathbf{x}$ and $\mathbf{x}^*$ with
   a. $\mathrm{I} = \mathrm{I}^*$
   b. $\mathrm{II} = \mathrm{II}^*$ (Hint: Try reparametrizing some of our simplest surfaces.)

10. Let $\mathbf{x}(u, v) = \boldsymbol{\alpha}(u) + v\boldsymbol{\beta}(u)$ be a parametrization of a ruled surface. Prove that the tangent plane is constant along rulings (i.e., the surface is flat) if and only if $\boldsymbol{\alpha}'(u)$, $\boldsymbol{\beta}(u)$, and $\boldsymbol{\beta}'(u)$ are linearly dependent for every u. (Hint: When is $S_P(\mathbf{x}_v) = \mathbf{0}$? Alternatively, consider $\mathbf{x}_u \times \mathbf{x}_v$ and apply Exercise A.2.1.)

11. Prove that $\boldsymbol{\alpha}$ is a line of curvature in M if and only if the ruled surface formed by the surface normals along $\boldsymbol{\alpha}$ is flat. (Hint: See Exercise 10.)

12. Show that the Gaussian curvature of the parametrized surfaces

$$\mathbf{x}(u, v) = (u \cos v, u \sin v, v)$$
$$\mathbf{y}(u, v) = (u \cos v, u \sin v, \ln u)$$

is the same for each (u, v), and yet the first fundamental forms $\mathrm{I}_{\mathbf{x}}$ and $\mathrm{I}_{\mathbf{y}}$ do not agree. (Thus, we might expect that the converse of Corollary 3.2 is false; it takes slightly more work to show that there can be *no* local isometry.)

13. Suppose that through each point of a surface M there is a planar asymptotic curve with nonzero curvature. Prove that M must be a (subset of a) plane. (Hint: Start with Exercise 2.2.17a and apply Proposition 3.4.)

14. Suppose that the surface M is doubly ruled by orthogonal lines (i.e., through each point of M there pass two orthogonal lines).
    a. Using the Gauss equations, prove that $K = 0$.
    b. Now deduce that M must be a plane.

    (Hint: As usual, assume that, locally, the families of lines are u- and v-curves.)

15. Prove that the only minimal ruled surface with no planar points is the helicoid. (Hint: Consider the curves orthogonal to the rulings. Use Exercises 2.2.8b, 14, and 1.2.20.)

16. Suppose M is a surface with no umbilic points and one constant principal curvature $k_1 \neq 0$. Prove that M is (a subset of) a tube of radius $r = 1/|k_1|$ about a curve. That is, there is a curve $\boldsymbol{\alpha}$ so that M is (a subset of) the union of circles of radius r in each normal plane, centered along the curve. (Hints: As

usual, work with a parametrization where the u-curves are lines of curvature with principal curvature k_1 and the v-curves are lines of curvature with principal curvature k_2. Use Lemma 3.3 to show that the u-curves have curvature $|k_1|$ and are planar. Then define $\boldsymbol{\alpha}$ appropriately and check that it is a regular curve.)

17. If M is a surface with both principal curvatures constant, prove that M is (a subset of) either a sphere, a plane, or a right circular cylinder. (Hint: See Exercise 2.2.14, Proposition 3.4, and Exercise 16.)

18. Consider the parametrized surfaces

$$\mathbf{x}(u,v) = (-\cosh u \sin v, \cosh u \cos v, u) \qquad \text{(a catenoid)}$$
$$\mathbf{y}(u,v) = (u \cos v, u \sin v, v) \qquad \text{(a helicoid).}$$

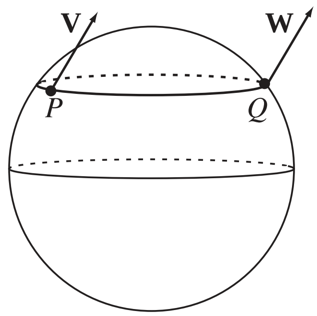
a. Compute the first and second fundamental forms of both surfaces, and check that both surfaces are minimal.
b. Find the asymptotic curves on both surfaces.
c. Show that we can locally reparametrize the helicoid in such a way as to make the first fundamental forms of the two surfaces agree; this means that the two surfaces are locally isometric. (Hint: See p. 39. Replace u with $\sinh u$ in the parametrization of the helicoid. Why is this legitimate?)
d. Why are they not globally isometric?
e. (for the student who's seen a bit of complex variables) As a hint to what's going on here, let $z = u + iv$ and $\mathbf{Z} = \mathbf{x} + i\mathbf{y}$, and check that, continuing to use the substitution from part c, $\mathbf{Z} = (\sin iz, \cos iz, z)$. Understand now how one can obtain a one-parameter family of isometric surfaces interpolating between the helicoid and the catenoid.

19. Find all the surfaces of revolution of constant curvature
a. $K = 0$
b. $K = 1$
c. $K = -1$

(Hint: There are more than you might suspect. But your answers will involve integrals you cannot express in terms of elementary functions.)

4. Covariant Differentiation, Parallel Translation, and Geodesics

Now we turn to the "intrinsic" geometry of a surface, i.e., the geometry that can be observed by an inhabitant (for example, a very thin ant) of the surface, who can only perceive what happens along (or, say, tangential to) the surface. Anyone who has studied Euclidean geometry knows how important the notion of *parallelism* is (and classical non-Euclidean geometry arises when one removes Euclid's parallel postulate, which stipulates that given any line L in the plane and any point P not lying on L, there is a unique line through P parallel to L). It seems quite intuitive to say that, working just in $\mathbb{R}^3$, two vectors $\mathbf{V}$ (thought of as being "tangent at P") and $\mathbf{W}$ (thought of as being "tangent at Q") are parallel provided that we obtain $\mathbf{W}$ when we move $\mathbf{V}$ "parallel to itself" from P to Q; in other words, if $\mathbf{W} = \mathbf{V}$. But what would an inhabitant of the sphere say? How should he compare a tangent vector at one point of the sphere to a tangent vector



Are **V** and **W** parallel?

FIGURE 4.1

at another and determine if they're "parallel"? (See Figure 4.1.) Perhaps a better question is this: Given a curve $\boldsymbol{\alpha}$ on the surface and a vector field $\mathbf{X}$ defined along $\boldsymbol{\alpha}$, should we say $\mathbf{X}$ is parallel if it has zero derivative along $\boldsymbol{\alpha}$?

We already know how an inhabitant differentiates a scalar function $f: M \to \mathbb{R}$, by considering the directional derivative $D_{\mathbf{V}} f$ for any tangent vector $\mathbf{V} \in T_P M$. We now begin with a

Definition. We say a function $\mathbf{X}: M \to \mathbb{R}^3$ is a *vector field* on M if

(1) $\mathbf{X}(P) \in T_P M$ for every $P \in M$, and
(2) for any parametrization $\mathbf{x}: U \to M$, the function $\mathbf{X}\circ\mathbf{x}: U \to \mathbb{R}^3$ is (continuously) differentiable.

Now, we can differentiate a vector field $\mathbf{X}$ on M in the customary fashion: If $\mathbf{V} \in T_P M$, we choose a curve $\boldsymbol{\alpha}$ with $\boldsymbol{\alpha}(0) = P$ and $\boldsymbol{\alpha}'(0) = \mathbf{V}$ and set $D_{\mathbf{V}}\mathbf{X} = (\mathbf{X}\circ\boldsymbol{\alpha})'(0)$. (As usual, the chain rule tells us this is well-defined.) But the inhabitant of the surface can only see that portion of this vector lying in the tangent plane. This brings us to the

Definition. Given a vector field $\mathbf{X}$ and $\mathbf{V} \in T_P M$, we define the *covariant derivative*

$$\begin{aligned} \nabla_{\mathbf{V}}\mathbf{X} &= (D_{\mathbf{V}}\mathbf{X})^{\parallel} = \text{the projection of } D_{\mathbf{V}}\mathbf{X} \text{ onto } T_P M \\ &= D_{\mathbf{V}}\mathbf{X} - (D_{\mathbf{V}}\mathbf{X}\cdot\mathbf{n})\mathbf{n}. \end{aligned}$$

Given a curve $\boldsymbol{\alpha}$ in M, we say the vector field $\mathbf{X}$ is *covariant constant* or *parallel* along $\boldsymbol{\alpha}$ if $\nabla_{\boldsymbol{\alpha}'(t)}\mathbf{X} = \mathbf{0}$ for all t. (This means that $D_{\boldsymbol{\alpha}'(t)}\mathbf{X} = (\mathbf{X}\circ\boldsymbol{\alpha})'(t)$ is a multiple of the normal vector $\mathbf{n}(\boldsymbol{\alpha}(t))$.)

Example 1. Let M be a sphere and let $\boldsymbol{\alpha}$ be a great circle in M. The derivative of the unit tangent vector of $\boldsymbol{\alpha}$ points towards the center of the circle, which is in this case the center of the sphere, and thus is completely normal to the sphere. Therefore, the unit tangent vector field of $\boldsymbol{\alpha}$ is parallel along $\boldsymbol{\alpha}$. Observe that the constant vector field $(0, 0, 1)$ *is* parallel along the equator $z = 0$ of a sphere centered at the origin. Is this true of any other constant vector field? ∇

Example 2. A fundamental example requires that we revisit the Christoffel symbols. Given a parametrized surface $\mathbf{x}: U \to M$, we have

$$\begin{aligned} \nabla_{\mathbf{x}_u}\mathbf{x}_u &= (\mathbf{x}_{uu})^{\parallel} = \Gamma_{uu}^{u}\mathbf{x}_u + \Gamma_{uu}^{v}\mathbf{x}_v \\ \nabla_{\mathbf{x}_v}\mathbf{x}_u &= (\mathbf{x}_{uv})^{\parallel} = \Gamma_{uv}^{u}\mathbf{x}_u + \Gamma_{uv}^{v}\mathbf{x}_v = \nabla_{\mathbf{x}_u}\mathbf{x}_v, \quad \text{and} \end{aligned}$$

$$\nabla_{\mathbf{x}_v}\mathbf{x}_v = (\mathbf{x}_{vv})^{\parallel} = \Gamma^u_{vv}\mathbf{x}_u + \Gamma^v_{vv}\mathbf{x}_v. \quad \triangledown$$

The first result we prove is the following

Proposition 4.1. *Let I be an interval in $\mathbb{R}$ with $0 \in I$. Given a curve $\boldsymbol{\alpha}\colon I \to M$ with $\boldsymbol{\alpha}(0) = P$ and $\mathbf{X}_0 \in T_P M$, there is a unique parallel vector field $\mathbf{X}$ defined along $\boldsymbol{\alpha}$ with $\mathbf{X}(P) = \mathbf{X}_0$.*

Proof. Assuming $\boldsymbol{\alpha}$ lies in a parametrized portion $\mathbf{x}\colon U \to M$, set $\boldsymbol{\alpha}(t) = \mathbf{x}(u(t), v(t))$ and write $\mathbf{X}(\boldsymbol{\alpha}(t)) = a(t)\mathbf{x}_u(u(t), v(t)) + b(t)\mathbf{x}_v(u(t), v(t))$. Then $\boldsymbol{\alpha}'(t) = u'(t)\mathbf{x}_u + v'(t)\mathbf{x}_v$ (where the the cumbersome argument $(u(t), v(t))$ is understood). So, by the product rule and chain rule, we have

$$\begin{aligned}
\nabla_{\boldsymbol{\alpha}'(t)}\mathbf{X} &= \big((\mathbf{X}\circ\boldsymbol{\alpha})'(t)\big)^{\parallel} = \left(\frac{d}{dt}\big(a(t)\mathbf{x}_u(u(t), v(t)) + b(t)\mathbf{x}_v(u(t), v(t))\big)\right)^{\parallel} \\
&= a'(t)\mathbf{x}_u + b'(t)\mathbf{x}_v + a(t)\left(\frac{d}{dt}\mathbf{x}_u(u(t), v(t))\right)^{\parallel} + b(t)\left(\frac{d}{dt}\mathbf{x}_v(u(t), v(t))\right)^{\parallel} \\
&= a'(t)\mathbf{x}_u + b'(t)\mathbf{x}_v + a(t)\big(u'(t)\mathbf{x}_{uu} + v'(t)\mathbf{x}_{uv}\big)^{\parallel} + b(t)\big(u'(t)\mathbf{x}_{vu} + v'(t)\mathbf{x}_{vv}\big)^{\parallel} \\
&= a'(t)\mathbf{x}_u + b'(t)\mathbf{x}_v + a(t)\big(u'(t)(\Gamma^u_{uu}\mathbf{x}_u + \Gamma^v_{uu}\mathbf{x}_v) + v'(t)(\Gamma^u_{uv}\mathbf{x}_u + \Gamma^v_{uv}\mathbf{x}_v)\big) \\
&\qquad + b(t)\big(u'(t)(\Gamma^u_{vu}\mathbf{x}_u + \Gamma^v_{vu}\mathbf{x}_v) + v'(t)(\Gamma^u_{vv}\mathbf{x}_u + \Gamma^v_{vv}\mathbf{x}_v)\big) \\
&= \big(a'(t) + a(t)(\Gamma^u_{uu}u'(t) + \Gamma^u_{uv}v'(t)) + b(t)(\Gamma^u_{vu}u'(t) + \Gamma^u_{vv}v'(t))\big)\mathbf{x}_u \\
&\qquad + \big(b'(t) + a(t)(\Gamma^v_{uu}u'(t) + \Gamma^v_{uv}v'(t)) + b(t)(\Gamma^v_{vu}u'(t) + \Gamma^v_{vv}v'(t))\big)\mathbf{x}_v.
\end{aligned}$$

Thus, to say $\mathbf{X}$ is parallel along the curve $\boldsymbol{\alpha}$ is to say that $a(t)$ and $b(t)$ are solutions of the linear system of first order differential equations

$$(\clubsuit) \qquad \begin{aligned}
a'(t) + a(t)(\Gamma^u_{uu}u'(t) + \Gamma^u_{uv}v'(t)) + b(t)(\Gamma^u_{vu}u'(t) + \Gamma^u_{vv}v'(t)) &= 0 \\
b'(t) + a(t)(\Gamma^v_{uu}u'(t) + \Gamma^v_{uv}v'(t)) + b(t)(\Gamma^v_{vu}u'(t) + \Gamma^v_{vv}v'(t)) &= 0.
\end{aligned}$$

By Theorem 3.2 of the Appendix, this system has a unique solution on I once we specify $a(0)$ and $b(0)$, and hence we obtain a unique parallel vector field $\mathbf{X}$ with $\mathbf{X}(P) = \mathbf{X}_0$. $\square$

Definition. If $\boldsymbol{\alpha}$ is a path from P to Q, we refer to $\mathbf{X}(Q)$ as the *parallel translate* of $\mathbf{X}(P) = \mathbf{X}_0 \in T_P M$ along $\boldsymbol{\alpha}$, or the result of *parallel translation* along $\boldsymbol{\alpha}$.

Remark. The system of differential equations ($\clubsuit$) that defines parallel translation shows that it is "intrinsic," i.e., depends only on the first fundamental form of M, despite our original extrinsic definition. In particular, parallel translation in locally isometric surfaces will be identical.

Example 3. Fix a latitude circle $u = u_0$ ($u_0 \neq 0, \pi$) on the unit sphere (see Example 1(d) on p. 37) and let's calculate the effect of parallel-translating the vector $\mathbf{X}_0 = \mathbf{x}_v$ starting at the point P given by $u = u_0$, $v = 0$, once around the circle, counterclockwise. We parametrize the curve by $u(t) = u_0$, $v(t) = t$, $0 \le t \le 2\pi$. Using our computation of the Christoffel symbols of the sphere in Example 1 or 2 of Section 3, we obtain from ($\clubsuit$) the differential equations

$$\begin{aligned}
a'(t) &= \sin u_0 \cos u_0 b(t), && a(0) = 0 \\
b'(t) &= -\cot u_0 a(t), && b(0) = 1.
\end{aligned}$$

We solve this system by differentiating the second equation again and substituting the first:

$$b''(t) = -\cot u_0 a'(t) = -\cos^2 u_0 b(t), \qquad b(0) = 1.$$

Recalling that every solution of the differential equation $y''(t) + k^2 y(t) = 0$ is of the form $y(t) = c_1 \cos(kt) + c_2 \sin(kt)$, $c_1, c_2 \in \mathbb{R}$, we see that the solution is

$$a(t) = \sin u_0 \sin\big((\cos u_0)t\big), \qquad b(t) = \cos\big((\cos u_0)t\big).$$

Note that $\|\mathbf{X}(\boldsymbol{\alpha}(t))\|^2 = Ea(t)^2 + 2Fa(t)b(t) + Gb(t)^2 = \sin^2 u_0$ for all t. That is, the original vector $\mathbf{X}_0$ rotates as we parallel translate it around the latitude circle, and its length is preserved. As we see in Figure 4.2, the vector rotates clockwise as we proceed around the latitude circle (in the upper hemisphere). But

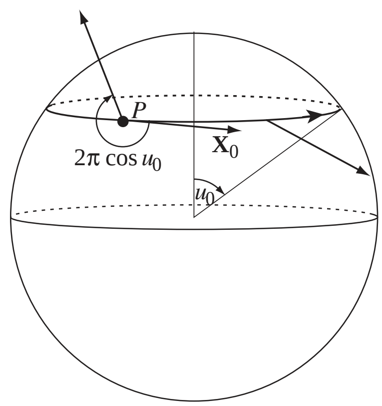


FIGURE 4.2

this makes sense: If we just take the covariant derivative of the vector field tangent to the circle, it points upwards (cf. Figure 3.1), so the vector field must rotate clockwise to counteract that effect in order to remain parallel. Since $b(2\pi) = \cos(2\pi \cos u_0)$, we see that the vector turns through an angle of $-2\pi \cos u_0$. ∇

Example 4 (Foucault pendulum)**.** Foucault observed in 1851 that the swing plane of a pendulum located on the latitude circle $u = u_0$ precesses with a period of $T = 24/\cos u_0$ hours. We can use the result of Example 3 to explain this. We imagine the earth as fixed and "transport" the swinging pendulum once around the circle in 24 hours. If we make the pendulum very long and the swing rather short, the motion will be "essentially" tangential to the surface of the earth. If we move slowly around the circle, the forces will be "essentially" normal to the sphere: In particular, letting R denote the radius of the earth (approximately 3960 mi), the tangential component of the centripetal acceleration is (cf. Figure 3.1)

$$(R\sin u_0)\cos u_0 \left(\frac{2\pi}{24}\right)^2 \le \frac{2\pi^2 R}{24^2} \approx 135.7 \text{ mi/hr}^2 \approx 0.0553 \text{ ft/sec}^2 \approx 0.17\% g.$$

Thus, the "swing vector field" is, for all practical purposes, parallel along the curve. Therefore, it turns through an angle of $2\pi \cos u_0$ in one trip around the circle, so it takes $\dfrac{2\pi}{(2\pi \cos u_0)/24} = \dfrac{24}{\cos u_0}$ hours to return to its original swing plane. ∇

Our experience in Example 3 suggests the following

Proposition 4.2. *Parallel translation preserves lengths and angles. That is, if* $\mathbf{X}$ *and* $\mathbf{Y}$ *are parallel vector fields along a curve* $\boldsymbol{\alpha}$ *from* P *to* Q*, then* $\|\mathbf{X}(P)\| = \|\mathbf{X}(Q)\|$ *and the angle between* $\mathbf{X}(P)$ *and* $\mathbf{Y}(P)$ *equals the angle between* $\mathbf{X}(Q)$ *and* $\mathbf{Y}(Q)$ *(assuming these are nonzero vectors).*

Proof. Consider $f(t) = \mathbf{X}(\boldsymbol{\alpha}(t)) \cdot \mathbf{Y}(\boldsymbol{\alpha}(t))$. Then

$$\begin{aligned} f'(t) &= (\mathbf{X}\circ\boldsymbol{\alpha})'(t) \cdot (\mathbf{Y}\circ\boldsymbol{\alpha})(t) + (\mathbf{X}\circ\boldsymbol{\alpha})(t) \cdot (\mathbf{Y}\circ\boldsymbol{\alpha})'(t) \\ &= D_{\boldsymbol{\alpha}'(t)}\mathbf{X} \cdot \mathbf{Y} + \mathbf{X} \cdot D_{\boldsymbol{\alpha}'(t)}\mathbf{Y} \overset{(1)}{=} \nabla_{\boldsymbol{\alpha}'(t)}\mathbf{X} \cdot \mathbf{Y} + \mathbf{X} \cdot \nabla_{\boldsymbol{\alpha}'(t)}\mathbf{Y} \overset{(2)}{=} 0. \end{aligned}$$

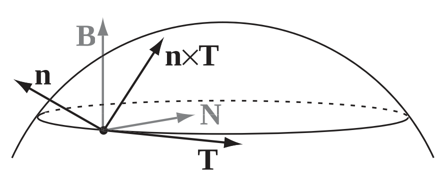
Note that equality (1) holds because $\mathbf{X}$ and $\mathbf{Y}$ are tangent to M and hence their dot product with any vector normal to the surface is 0. Equality (2) holds because $\mathbf{X}$ and $\mathbf{Y}$ are assumed parallel along $\boldsymbol{\alpha}$. It follows that the dot product $\mathbf{X} \cdot \mathbf{Y}$ remains constant along $\boldsymbol{\alpha}$. Taking $\mathbf{Y} = \mathbf{X}$, we infer that $\|\mathbf{X}\|$ (and similarly $\|\mathbf{Y}\|$) is constant. Knowing that, using the famous formula $\cos\theta = \mathbf{X} \cdot \mathbf{Y}/\|\mathbf{X}\|\|\mathbf{Y}\|$ for the angle θ between $\mathbf{X}$ and $\mathbf{Y}$, we infer that the angle remains constant. ☐

Now we change gears somewhat. We saw in Exercise 1.1.8 that the shortest path joining two points in $\mathbb{R}^3$ is a line segment and in Exercise 1.3.1 that the shortest path joining two points on the unit sphere is a great circle. One characterization of the line segment is that it never changes direction, so that its unit tangent vector is parallel (so no distance is wasted by turning). (What about the sphere?) It seems plausible that the mythical inhabitant of our general surface M might try to travel from one point to another in M, *staying in* M, by similarly not turning; that is, so that his unit tangent vector field is parallel along his path. Physically, this means that if he travels at constant speed, any acceleration should be normal to the surface. This leads us to the following

Definition. We say a parametrized curve $\boldsymbol{\alpha}$ in a surface M is a *geodesic* if its tangent vector is parallel along the curve, i.e., if $\nabla_{\boldsymbol{\alpha}'}\boldsymbol{\alpha}' = 0$.

Recall that since parallel translation preserves lengths, $\boldsymbol{\alpha}$ must have constant speed, although it may not be arclength-parametrized. In general, we refer to an unparametrized curve as a geodesic if its arclength parametrization is in fact a geodesic.

In general, given any arclength-parametrized curve $\boldsymbol{\alpha}$ lying on M, we defined its normal curvature at the end of Section 2. Instead of using the Frenet frame, it is natural to consider the *Darboux frame* for $\boldsymbol{\alpha}$, which takes into account the fact that $\boldsymbol{\alpha}$ lies on the surface M. (Both are illustrated in Figure 4.3.) We take



The Frenet and Darboux frames

FIGURE 4.3

the right-handed orthonormal basis $\{\mathbf{T}, \mathbf{n}\times\mathbf{T}, \mathbf{n}\}$; note that the first two vectors give a basis for $T_P M$. We can decompose the curvature vector

$$\kappa\mathbf{N} = \big(\underbrace{\kappa\mathbf{N}\cdot(\mathbf{n}\times\mathbf{T})}_{\kappa_g}\big)(\mathbf{n}\times\mathbf{T}) + \big(\underbrace{\kappa\mathbf{N}\cdot\mathbf{n}}_{\kappa_n}\big)\mathbf{n}.$$

As we saw before, κ_n gives the *normal* component of the curvature vector; κ_g gives the *tangential* component of the curvature vector and is called the *geodesic curvature*. This terminology arises from the fact that $\boldsymbol{\alpha}$ is a geodesic if and only if its geodesic curvature vanishes. (When $\kappa = 0$, the principal normal is not defined, and we really should write $\boldsymbol{\alpha}''$ in the place of $\kappa\mathbf{N}$. If the acceleration vanishes at a point, then certainly its normal and tangential components are both $\mathbf{0}$.)

Example 5. We saw in Example 1 that every great circle on a sphere is a geodesic. Are there others? Let $\boldsymbol{\alpha}$ be a geodesic on a sphere centered at the origin. Since $\kappa_g = 0$, the acceleration vector $\boldsymbol{\alpha}''(s)$ must be a multiple of $\boldsymbol{\alpha}(s)$ for every s, and so $\boldsymbol{\alpha}'' \times \boldsymbol{\alpha} = \mathbf{0}$. Therefore $\boldsymbol{\alpha}' \times \boldsymbol{\alpha} = \mathbf{A}$ is a constant vector, so $\boldsymbol{\alpha}$ lies in the plane passing through the origin with normal vector $\mathbf{A}$. That is, $\boldsymbol{\alpha}$ is a great circle. ∇

Remark. We saw in Example 3 that a vector rotates clockwise at a constant rate as we parallel translate along the latitude circle of the sphere. If we think about the unit tangent vector $\mathbf{T}$ moving counterclockwise along this curve, its covariant derivative along the curve points up the sphere, as shown in Figure 4.4, i.e., "to the left." Thus, we must compensate by steering "to the right" in order to have no net turning (i.e., to

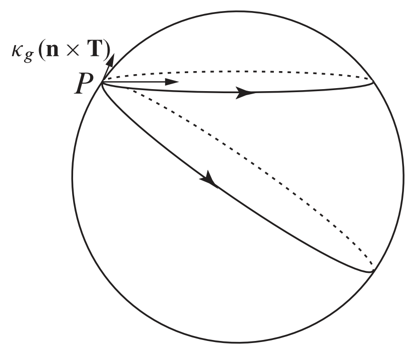


FIGURE 4.4

make the covariant derivative zero). Of course, this makes sense also because, according to Example 5, the geodesic that passes through P in the same direction heads "downhill," to the right.

Using the equations (♣), let's now give the equations for the curve $\boldsymbol{\alpha}(t) = \mathbf{x}(u(t), v(t))$ to be a geodesic. Since $\mathbf{X} = \boldsymbol{\alpha}'(t) = u'(t)\mathbf{x}_u + v'(t)\mathbf{x}_v$, we have $a(t) = u'(t)$ and $b(t) = v'(t)$, and the resulting equations are

(♣♣)
$$\begin{aligned} u''(t) + \Gamma^u_{uu}u'(t)^2 + 2\Gamma^u_{uv}u'(t)v'(t) + \Gamma^u_{vv}v'(t)^2 &= 0 \\ v''(t) + \Gamma^v_{uu}u'(t)^2 + 2\Gamma^v_{uv}u'(t)v'(t) + \Gamma^v_{vv}v'(t)^2 &= 0. \end{aligned}$$

The following result is a consequence of basic results on differential equations (see Theorem 3.1 of the Appendix).

Proposition 4.3. *Given a point $P \in M$ and $\mathbf{V} \in T_PM$, $\mathbf{V} \neq \mathbf{0}$, there exist $\varepsilon > 0$ and a* unique *geodesic $\boldsymbol{\alpha}\colon (-\varepsilon, \varepsilon) \to M$ with $\boldsymbol{\alpha}(0) = P$ and $\boldsymbol{\alpha}'(0) = \mathbf{V}$.*

Example 6. We now use the equations (♣♣) to solve for geodesics analytically in a few examples.

(a) Let $\mathbf{x}(u,v) = (u,v)$ be the obvious parametrization of the plane. Then all the Christoffel symbols vanish and the geodesics are the solutions of

$$u''(t) = v''(t) = 0,$$

so we get the lines $\boldsymbol{\alpha}(t) = (u(t), v(t)) = (a_1 t + b_1, a_2 t + b_2)$, as expected. Note that $\boldsymbol{\alpha}$ does in fact have constant speed.

(b) Using the standard spherical coordinate parametrization of the sphere, we obtain (see Example 1 or 2 of Section 3) the equations

$$(*) \qquad u''(t) - \sin u(t) \cos u(t) v'(t)^2 = 0 = v''(t) + 2\cot u(t) u'(t) v'(t).$$

Well, one obvious set of solutions is to take $u(t) = t$, $v(t) = v_0$ (and these, indeed, give the great circles through the north pole). Integrating the second equation in $(*)$ we obtain $\ln v'(t) = -2\ln\sin u(t) + \text{const}$, so

$$v'(t) = \frac{c}{\sin^2 u(t)}$$

for some constant c. Substituting this in the first equation in $(*)$ we find that

$$u''(t) - \frac{c^2 \cos u(t)}{\sin^3 u(t)} = 0;$$

multiplying both sides by $u'(t)$ (the "energy trick" from physics) and integrating, we get

$$u'(t)^2 = C^2 - \frac{c^2}{\sin^2 u(t)}, \quad \text{and so} \quad u'(t) = \pm\sqrt{C^2 - \frac{c^2}{\sin^2 u(t)}}$$

for some constant C. Switching to Leibniz notation for obvious reasons, we obtain

$$\frac{dv}{du} = \frac{v'(t)}{u'(t)} = \pm\frac{c\csc^2 u}{\sqrt{C^2 - c^2\csc^2 u}}; \quad \text{thus, separating variables gives}$$

$$dv = \pm\frac{c\csc^2 u\,du}{\sqrt{C^2 - c^2\csc^2 u}} = \pm\frac{c\csc^2 u\,du}{\sqrt{(C^2 - c^2) - c^2\cot^2 u}}.$$

Now we make the substitution $c\cot u = \sqrt{C^2 - c^2}\sin w$; then we have

$$dv = \pm\frac{c\csc^2 u\,du}{\sqrt{(C^2 - c^2) - c^2\cot^2 u}} = \mp dw,$$

and so, at long last, we have $w = \pm v + a$ for some constant a. Thus,

$$c\cot u = \sqrt{C^2 - c^2}\sin w = \sqrt{C^2 - c^2}\sin(\pm v + a) = \sqrt{C^2 - c^2}(\sin a\cos v \pm \cos a\sin v),$$

and so, finally, we have the equation

$$c\cos u + \sqrt{C^2 - c^2}\sin u(A\cos v + B\sin v) = 0,$$

which we should recognize as the equation of a great circle! (Here's a hint: This curve lies on the plane $\sqrt{C^2 - c^2}(Ax + By) + cz = 0$.) $\qquad \nabla$

We can now give a beautiful geometric description of the geodesics on a surface of revolution.

Proposition 4.4 (Clairaut's relation). *The geodesics on a surface of revolution satisfy the equation*

$$r\cos\phi = \text{const}, \tag{$\diamond$}$$

where r is the distance from the axis of revolution and ϕ is the angle between the geodesic and the parallel. Conversely, any (constant speed) curve satisfying ($\diamond$) *that is* not *a parallel is a geodesic.*

Proof. For the surface of revolution parametrized as in Example 9 of Section 2, we have $E = 1$, $F = 0$, $G = f(u)^2$, $\Gamma^v_{uv} = \Gamma^v_{vu} = f'(u)/f(u)$, $\Gamma^u_{vv} = -f(u)f'(u)$, and all other Christoffel symbols are 0 (see Exercise 2.3.2d.). Then the system (♣♣) of differential equations becomes

$$u'' - ff'(v')^2 = 0 \tag{$\dagger_1$}$$

$$v'' + \frac{2f'}{f}u'v' = 0. \tag{$\dagger_2$}$$

Rewriting the equation ($\dagger_2$) and integrating, we obtain

$$\begin{aligned} \frac{v''(t)}{v'(t)} &= -\frac{2f'(u(t))u'(t)}{f(u(t))} \\ \ln v'(t) &= -2\ln f(u(t)) + \text{const} \\ v'(t) &= \frac{c}{f(u(t))^2}, \end{aligned}$$

so along a geodesic the quantity $f(u)^2 v' = Gv'$ is constant. We recognize this as the dot product of the tangent vector of our geodesic with the vector $\mathbf{x}_v$, and so we infer that $\|\mathbf{x}_v\|\cos\phi = r\cos\phi$ is constant. (Recall that, by Proposition 4.2, the tangent vector of the geodesic has constant length.)

To this point we have seen that the equation ($\dagger_2$) is equivalent to the condition $r\cos\phi = \text{const}$, provided we assume $\|\boldsymbol{\alpha}'\|^2 = u'^2 + Gv'^2$ is constant as well. But if

$$u'(t)^2 + Gv'(t)^2 = u'(t)^2 + f(u(t))^2 v'(t)^2 = \text{const},$$

we differentiate and obtain

$$u'(t)u''(t) + f(u(t))^2 v'(t)v''(t) + f(u(t))f'(u(t))u'(t)v'(t)^2 = 0;$$

substituting for $v''(t)$ using ($\dagger_2$), we find

$$u'(t)\big(u''(t) - f(u(t))f'(u(t))v'(t)^2\big) = 0.$$

In other words, *provided* $u'(t) \neq 0$, a constant-speed curve satisfying ($\dagger_2$) satisfies ($\dagger_1$) as well. (See Exercise 6 for the case of the parallels.) $\square$

Remark. We can give a simple physical interpretation of Clairaut's relation. Imagine a particle with mass 1 constrained to move along a surface. If no external forces are acting, then the particle moves along a geodesic and, moreover, angular momentum is conserved (because there are no torques). In the case of our surface of revolution, the vertical component of the angular momentum $\mathbf{L} = \boldsymbol{\alpha} \times \boldsymbol{\alpha}'$ is—surprise, surprise!—$f^2 v'$, which we've shown is constant. Perhaps some forces normal to the surface are required to keep the particle on the surface; then the particle still moves along a geodesic (why?). Moreover, since $(\boldsymbol{\alpha} \times \mathbf{n}) \cdot (0, 0, 1) = 0$, the resulting torques *still* have no vertical component.

Returning to our original motivation for geodesics, we now consider the following scenario. Choose $P \in M$ arbitrary and a geodesic $\boldsymbol{\gamma}$ through P, and draw a curve C_0 through P orthogonal to $\boldsymbol{\gamma}$. We now choose a parametrization $\mathbf{x}(u, v)$ so that $\mathbf{x}(0,0) = P$, the u-curves are geodesics orthogonal to C_0, and the v-curves are the orthogonal trajectories of the u-curves, as pictured in Figure 4.5. (It follows from Theorem

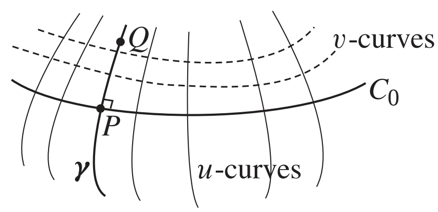


FIGURE 4.5

3.3 of the Appendix that we can do this on some neighborhood of P.)

In this parametrization we have $F = 0$ and $E = E(u)$ (see Exercise 13). Now, if $\boldsymbol{\alpha}(t) = \mathbf{x}(u(t), v(t))$, $a \leq t \leq b$, is any path from $P = \mathbf{x}(0,0)$ to $Q = \mathbf{x}(u_0, 0)$, we have

$$\begin{aligned}\text{length}(\boldsymbol{\alpha}) &= \int_a^b \sqrt{E(u(t))u'(t)^2 + G(u(t), v(t))v'(t)^2}\,dt \geq \int_a^b \sqrt{E(u(t))}|u'(t)|\,dt \\ &\geq \int_0^{u_0} \sqrt{E(u)}\,du,\end{aligned}$$

which is the length of the geodesic arc $\boldsymbol{\gamma}$ from P to Q. Thus, we have deduced the following.

Proposition 4.5. *For any point Q on $\boldsymbol{\gamma}$ contained in this parametrization, any path from P to Q* contained in this parametrization *is at least as long as the length of the geodesic segment. More colloquially, geodesics are* locally *distance-minimizing.*

Example 7. Why is Proposition 4.5 a local statement? Well, consider a great circle on a sphere, as shown in Figure 4.6. If we go more than halfway around, we obviously have not taken the shortest path. ∇

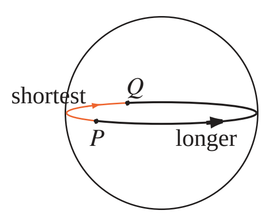


FIGURE 4.6

Remark. It turns out that any surface can be endowed with a *metric* (or *distance measure*) by defining the distance between any two points to be the infimum (usually, the minimum) of the lengths of all piecewise-$\mathcal{C}^1$ paths joining them. (Although the distance measure is different from the Euclidean distance as the surface sits in $\mathbb{R}^3$, the topology—notion of "neighborhood"—induced by this metric structure is the induced topology that the surface inherits as a subspace of $\mathbb{R}^3$.) It is a consequence of the Hopf-Rinow Theorem (see M. doCarmo, *Differential Geometry of Curves and Surfaces*, Prentice Hall, 1976, p. 333, or M. Spivak, *A*

Comprehensive Introduction to Differential Geometry, third edition, volume 1, Publish or Perish, Inc., 1999, p. 342) that in a surface in which every parametrized geodesic is defined for all time (a "complete" surface), every two points are in fact joined by a geodesic of least length. The proof of this result is quite tantalizing: To find the shortest path from P to Q, one walks around the "geodesic circle" of points a small distance from P and finds the point R on it closest to Q; one then proves that the unique geodesic emanating from P that passes through R must eventually pass through Q, and there can be no shorter path.

We referred earlier to two surfaces M and M^* as being globally isometric (e.g., in Example 6 in Section 1). We can now give the official definition: There should be a function $f: M \to M^*$ that establishes a one-to-one correspondence and preserves distance—for any $P, Q \in M$, the distance between P and Q in M should be equal to the distance between $f(P)$ and $f(Q)$ in M^*.

EXERCISES 2.4

1. Determine the result of parallel translating the vector $(0,0,1)$ once around the circle $x^2 + y^2 = a^2$, $z = 0$, on the right circular cylinder $x^2 + y^2 = a^2$.

2. Prove that $\kappa^2 = \kappa_g^2 + \kappa_n^2$.

3. Suppose $\boldsymbol{\alpha}$ is a non-arclength-parametrized curve. Using the formula $(**)$ on p. 14, prove that the velocity vector of $\boldsymbol{\alpha}$ is parallel along $\boldsymbol{\alpha}$ if and only if $\kappa_g = 0$ and $\upsilon' = 0$.

*4. Find the geodesic curvature κ_g of a latitude circle $u = u_0$ on the unit sphere (see Example 1(d) on p. 37)
   a. directly
   b. by applying the result of Exercise 2

5. Consider the right circular cone with vertex angle 2ϕ parametrized by

$$\mathbf{x}(u, v) = (u \tan\phi \cos v, u \tan\phi \sin v, u), \quad 0 < u \le u_0,\ 0 \le v \le 2\pi.$$

   Find the geodesic curvature κ_g of the circle $u = u_0$ by using trigonometric considerations. Check that your answer agrees with the curvature of the circle you get by unrolling the cone to form a "pacman" figure, as shown on the left in Figure 4.7. (For a proof that these curvatures should agree, see Exercise 2.1.10 and Exercise 3.1.7.)

6. Check that the parallel $u = u_0$ is a geodesic on the surface of revolution parametrized as in Proposition 4.4 if and only if $f'(u_0) = 0$. Give a geometric interpretation of and explanation for this result.

7. Use the equations (♣), as in Example 3, to determine through what angle a vector turns when it is parallel-translated once around the circle $u = u_0$ on the cone $\mathbf{x}(u, v) = (u\cos v, u \sin v, cu)$, $c \neq 0$. (See Exercise 2.3.2c.)

8. a. Prove that if the surfaces M and M^* are tangent along the curve C, parallel translation along C is the same in both surfaces.

b. Use the result of part a to determine the effect of parallel translation around the latitude circle $u = u_0$ on the unit sphere (once again, see Example 1(d) on p. 37), using only geometry, trigonometry, and Figure 4.7. (Note the Remark on p. 68.)

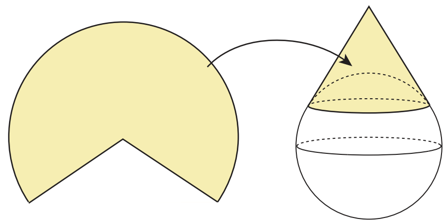


FIGURE 4.7

*9. What curves lying on a sphere have constant geodesic curvature?

10. Use the equations (♣♣) to find the geodesics on parametrized surface $\mathbf{x}(u, v) = (e^u \cos v, e^u \sin v, 0)$. (Hint: Aim for dv/du. Use the second equation in (♣♣) and the fact that geodesics must have constant speed.)

11. Use the equations (♣♣) to find the geodesics on the plane parametrized by polar coordinates. (Hint: Examine Example 6(b).)

12. Prove or give a counterexample:
    a. A curve is both an asymptotic curve and a geodesic if and only if it is a line.
    b. If a curve is both a geodesic and a line of curvature, then it must be planar.

♯13. a. Suppose $F = 0$ and the u-curves are geodesics. Use the equations (♣♣) to prove that E is a function of u only.
    b. Suppose $F = 0$ and the u- and v-curves are geodesics. Prove that the surface is flat.

14. Suppose $F = 0$ and the u-curves are geodesics. Prove that the length of the u-curve from $u = u_0$ to $u = u_1$ is independent of v. (See Figure 4.8.)

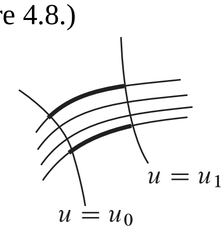


FIGURE 4.8

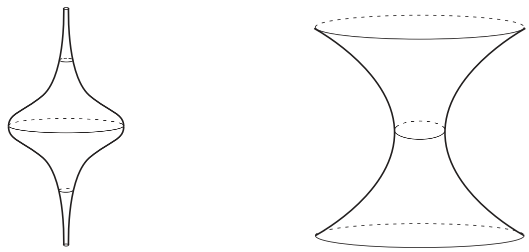
15. a. Prove that an arclength-parametrized curve $\boldsymbol{\alpha}$ on a surface M with $\kappa \neq 0$ is a geodesic if and only if $\mathbf{n} = \pm\mathbf{N}$.
    b. Let $\boldsymbol{\alpha}$ be a space curve, and let M be the ruled surface generated by its binormals. Prove that the curve is a geodesic on M.

16. a. Suppose a geodesic is planar and has $\kappa \neq 0$ at P. Prove that its tangent vector at P must be a principal direction. (Hint: Use Exercise 15.)

b. Prove that if every geodesic of a (connected) surface is planar, then the surface is contained in a plane or a sphere.

17. Show that the geodesic curvature at P of a curve C in M is equal (in absolute value) to the curvature at P of the projection of C into $T_P M$.

*18. Use Clairaut's relation, Proposition 4.4, to analyze the geodesics on each of the surfaces pictured in Figure 4.9. In particular, other than the meridians, in each case which geodesics are unbounded (i.e., go off to infinity)?



(a) (b)

FIGURE 4.9

19. Check using Clairaut's relation, Proposition 4.4, that great circles are geodesics on a sphere. (Hint: The result of Exercise A.1.3 may be useful.)

20. Let M be a surface and $P \in M$. We say $\mathbf{U}, \mathbf{V} \in T_P M$ are *conjugate* if $\mathrm{II}_P(\mathbf{U}, \mathbf{V}) = 0$.
   a. Let $C \subset M$ be a curve (with the property that its tangent vector is never a principal direction with principal curvature 0). Define the *envelope* M^* of the tangent planes to M along C to be the ruled surface whose generator at $P \in C$ is the limiting position as $Q \to P$ of the intersection line of the tangent planes to M at P and Q. Prove that the generator at P is conjugate to the tangent line to C at P.
   b. Prove that if C is nowhere tangent to an asymptotic direction, then M^* is smooth (at least near C). Prove, moreover, that M^* is tangent to M along C and is a developable (flat ruled) surface.
   c. Apply part b to give a geometric way of computing parallel translation. In particular, do this for a latitude circle on the sphere. (Cf. Exercise 8.)

21. Suppose that on a surface M the parallel translation of a vector from one point to another is independent of the path chosen. Prove that M must be flat. (Hint: Fix an orthonormal basis $\mathbf{e}_1^o, \mathbf{e}_2^o$ for $T_P M$ and define vector fields $\mathbf{e}_1, \mathbf{e}_2$ by parallel translating. Choose coordinates so that the u-curves are always tangent to $\mathbf{e}_1$ and the v-curves are always tangent to $\mathbf{e}_2$. See Exercise 13.)

22. Use the Clairaut relation, Proposition 4.4, to describe the geodesics on the torus as parametrized in Example 1(c) of Section 1. (Start with a geodesic starting at and making angle ϕ_0 with the outer parallel. Your description should distinguish between the cases $0 < \cos\phi_0 \le \frac{a-b}{a+b}$ and $\cos\phi_0 > \frac{a-b}{a+b}$. Which geodesics never cross the outer parallel at all? Also, remember that through each point there is a *unique* geodesic in each direction.)

23. Use the proof of the Clairaut relation, Proposition 4.4, to show that a unit-speed geodesic on a surface of revolution is given in terms of the standard parametrization in Example 9 of Section 2 by

$$v = c\int \frac{du}{f(u)\sqrt{f(u)^2 - c^2}} + \text{const}.$$

Now deduce that in the case of a non-arclength parametrization we obtain

$$v = c\int \frac{\sqrt{f'(u)^2 + g'(u)^2}}{f(u)\sqrt{f(u)^2 - c^2}}\,du + \text{const}.$$

*24. Use Exercise 23 to give equations of the geodesics on the pseudosphere (see Example 8 of Section 2). Deduce, in particular, that the only geodesics that are unbounded are the meridians.

25. Use Exercise 23 to show that any geodesic on the paraboloid $z = x^2 + y^2$ that is not a meridian intersects every meridian. (Hint: Show that it cannot approach a meridian asymptotically.)

26. Let M be the hyperboloid $x^2 + y^2 - z^2 = 1$, and let C be the circle $x^2 + y^2 = 1$, $z = 0$.
   a. Use Clairaut's relation, Proposition 4.4, to show that, with the exception of the circle C, every geodesic on M is unbounded.
   b. Show that there are geodesics that approach the circle C asymptotically. (Hint: Use Exercise 23.)

27. Let C be a parallel (with $u = u_0$) in a surface of revolution M. Suppose a geodesic $\boldsymbol{\gamma}$ approaches C asymptotically.
   a. Use Clairaut's relation, Proposition 4.4, to show that $\boldsymbol{\gamma}$ must approach "from above" (i.e., with $r > r_0 = f(u_0)$).
   b. Use Exercise 23 to show that C must itself be a geodesic. (Hint: Consider the Taylor expansion $f(u) = f(u_0) + f'(u_0)(u - u_0) + \frac{1}{2}f''(u_0)(u - u_0)^2 + \ldots$.)
   c. Give an alternative argument for the result of part b by using the fact that the metric discussed in the Remark on p. 74 is a continuous function of the pair of points. You will also need to use the fact that when points are sufficiently close, there is a unique shortest geodesic joining them.

28. Consider the surface $z = f(u, v)$. A curve $\boldsymbol{\alpha}$ whose tangent vector at each point $P = (u, v, f(u, v))$ projects to a scalar multiple of $\nabla f(u, v)$ is a curve of steepest ascent (why?). Suppose such a curve $\boldsymbol{\alpha}$ is also a geodesic.
   a. Prove that the projection of $\boldsymbol{\alpha}$ into the uv-plane is, suitably reparametrized, a geodesic in the uv-plane. (Hint: What is the projection of $\boldsymbol{\alpha}''$?)
   b. Deduce that $\boldsymbol{\alpha}$ is also a line of curvature. (Hint: See Exercise 16 when $\boldsymbol{\alpha}$ is not a line. The case of a line can be deduced from the computation in part c.)
   c. Show that if all the curves of steepest ascent are geodesics, then f satisfies the partial differential equation $f_u f_v (f_{vv} - f_{uu}) + f_{uv}(f_u^2 - f_v^2) = 0$. (Hint: When are the integral curves of ∇f lines?)
   d. Show that if all the curves of steepest ascent are geodesics, the level curves of f are parallel (see Exercise 1.2.24). (Hint: Show that $\|\nabla f\|$ is constant along level curves.)
   e. Give a characterization of the surfaces with the property that all curves of steepest ascent are geodesics.

CHAPTER 3

Surfaces: Further Topics

The first section is required reading, but the remaining sections of this chapter are independent of one another.

1. Holonomy and the Gauss-Bonnet Theorem

Let's now pursue the discussion of parallel translation that we began in Chapter 2. Let M be a surface and $\boldsymbol{\alpha}$ a closed curve in M. We begin by fixing a smoothly-varying orthonormal basis $\mathbf{e}_1, \mathbf{e}_2$ (a so-called *framing*) for the tangent planes of M in an open set of M containing $\boldsymbol{\alpha}$, as shown in Figure 1.1 below. Now

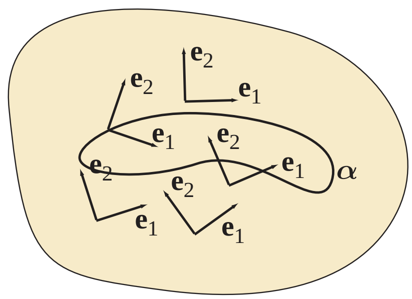


FIGURE 1.1

we make the following

Definition. Let $\boldsymbol{\alpha}$ be a *closed* curve in a surface M. The angle through which a vector turns *relative to the given framing* as we parallel translate it once around the curve $\boldsymbol{\alpha}$ is called the *holonomy*[1] around $\boldsymbol{\alpha}$.

For example, if we take a framing around $\boldsymbol{\alpha}$ by using the unit tangent vectors to $\boldsymbol{\alpha}$ as our vectors $\mathbf{e}_1$, then, by the definition of a geodesic, there there will be zero holonomy around a closed geodesic (why?). For another example, if we use the framing on (most of) the sphere given by the tangents to the lines of longitude and lines of latitude, the computation in Example 3 of Section 4 of Chapter 2 shows that the holonomy around a latitude circle $u = u_0$ of the unit sphere is $-2\pi \cos u_0$.

To make this more precise, for ease of understanding, let's work in an orthogonal parametrization[2] and define a framing by setting

$$\mathbf{e}_1 = \frac{\mathbf{x}_u}{\sqrt{E}} \qquad \text{and} \qquad \mathbf{e}_2 = \frac{\mathbf{x}_v}{\sqrt{G}}.$$

Since (much as in the case of curves) $\mathbf{e}_1$ and $\mathbf{e}_2$ give an orthonormal basis for the tangent space of our surface at each point, all the intrinsic curvature information (such as given by the Christoffel symbols)

[1] from *holo-+-nomy*, the study of the whole

[2] As usual, away from umbilic points, we can apply Theorem 3.3 of the Appendix to obtain a parametrization where the u- and v-curves are lines of curvature.

is encapsulated in knowing how $\mathbf{e}_1$ twists towards $\mathbf{e}_2$ as we move around the surface. In particular, if $\boldsymbol{\alpha}(t) = \mathbf{x}(u(t), v(t))$, $a \le t \le b$, is a parametrized curve, we can set

$$\phi_{12}(t) = \frac{d}{dt}\big(\mathbf{e}_1(u(t), v(t))\big) \cdot \mathbf{e}_2(u(t), v(t)),$$

which we may write more casually as $\mathbf{e}_1'(t) \cdot \mathbf{e}_2(t)$, with the understanding that everything must be done in terms of the parametrization. We emphasize that ϕ_{12} depends in an essential way on the *parametrized* curve $\boldsymbol{\alpha}$. Perhaps it's better, then, to write

$$\phi_{12} = \nabla_{\boldsymbol{\alpha}'}\mathbf{e}_1 \cdot \mathbf{e}_2.$$

Note, moreover, that the proof of Proposition 4.2 of Chapter 2 shows that $\nabla_{\boldsymbol{\alpha}'}\mathbf{e}_2 \cdot \mathbf{e}_1 = -\phi_{12}$ and $\nabla_{\boldsymbol{\alpha}'}\mathbf{e}_1 \cdot \mathbf{e}_1 = \nabla_{\boldsymbol{\alpha}'}\mathbf{e}_2 \cdot \mathbf{e}_2 = 0$. (Why?)

Remark. Although the notation seems cumbersome, it reminds us that ϕ_{12} is measuring how $\mathbf{e}_1$ *twists* towards $\mathbf{e}_2$ as we move along the curve $\boldsymbol{\alpha}$. This notation will fit in a more general context in Section 3.

Let's now derive an explicit formula for the function ϕ_{12}.

Proposition 1.1. *In an orthogonal parametrization with* $\mathbf{e}_1 = \mathbf{x}_u/\sqrt{E}$ *and* $\mathbf{e}_2 = \mathbf{x}_v/\sqrt{G}$, *we have* $\phi_{12} = \dfrac{1}{2\sqrt{EG}}\big(-E_v u' + G_u v'\big)$.

Proof. The key point is to take full advantage of the orthogonality of $\mathbf{x}_u$ and $\mathbf{x}_v$.

$$\begin{aligned}\phi_{12} &= \frac{d}{dt}\left(\frac{\mathbf{x}_u}{\sqrt{E}}\right) \cdot \frac{\mathbf{x}_v}{\sqrt{G}} \\ &= \frac{1}{\sqrt{EG}}\big(\mathbf{x}_{uu}u' + \mathbf{x}_{uv}v'\big) \cdot \mathbf{x}_v\end{aligned}$$

(since the term that would arise from differentiating $\sqrt{E}$ will involve $\mathbf{x}_u \cdot \mathbf{x}_v = 0$)

$$= \frac{1}{2\sqrt{EG}}\big(-E_v u' + G_u v'\big),$$

by the formulas (♠) on p. 58. $\square$

Suppose now that $\boldsymbol{\alpha}$ is a *closed* curve and we are interested in the holonomy around $\boldsymbol{\alpha}$. If $\mathbf{e}_1$ happens to be parallel along $\boldsymbol{\alpha}$, then the holonomy will, of course, be 0. If not, let's consider $\mathbf{X}(t)$ to be the parallel translation of $\mathbf{e}_1$ along $\boldsymbol{\alpha}(t)$ and write $\mathbf{X}(t) = \cos\psi(t)\mathbf{e}_1 + \sin\psi(t)\mathbf{e}_2$, taking $\psi(0) = 0$. Then $\mathbf{X}$ is parallel along $\boldsymbol{\alpha}$ if and only if

$$\begin{aligned}\mathbf{0} = \nabla_{\boldsymbol{\alpha}'}\mathbf{X} &= \nabla_{\boldsymbol{\alpha}'}(\cos\psi\,\mathbf{e}_1 + \sin\psi\,\mathbf{e}_2) \\ &= \cos\psi\,\nabla_{\boldsymbol{\alpha}'}\mathbf{e}_1 + \sin\psi\,\nabla_{\boldsymbol{\alpha}'}\mathbf{e}_2 + (-\sin\psi\,\mathbf{e}_1 + \cos\psi\,\mathbf{e}_2)\psi' \\ &= \cos\psi\,\phi_{12}\mathbf{e}_2 - \sin\psi\,\phi_{12}\mathbf{e}_1 + (-\sin\psi\,\mathbf{e}_1 + \cos\psi\,\mathbf{e}_2)\psi' \\ &= (\phi_{12} + \psi')(-\sin\psi\,\mathbf{e}_1 + \cos\psi\,\mathbf{e}_2).\end{aligned}$$

Thus, $\mathbf{X}$ is parallel along $\boldsymbol{\alpha}$ if and only if $\psi'(t) = -\phi_{12}(t)$. We therefore conclude:

Proposition 1.2. *The holonomy around the closed curve* C *equals* $\Delta\psi = -\displaystyle\int_a^b \phi_{12}(t)\,dt$.

Remark. Note that the angle ψ is measured from $\mathbf{e}_1$ in the direction of $\mathbf{e}_2$. Whether the vector turns counterclockwise or clockwise from our external viewpoint depends on the orientation of the framing.

Example 1. Back to our example of the latitude circle $u = u_0$ on the unit sphere. Then $\mathbf{e}_1 = \mathbf{x}_u$ and $\mathbf{e}_2 = (1/\sin u)\mathbf{x}_v$. If we parametrize the curve by taking $v = t$, $0 \le t \le 2\pi$, then we have (see Example 1 of Chapter 2, Section 3)

$$\nabla_{\boldsymbol{\alpha}'}\mathbf{e}_1 = \nabla_{\boldsymbol{\alpha}'}\mathbf{x}_u = (\mathbf{x}_{uv})^{\parallel} = \cot u_0 \mathbf{x}_v = \cos u_0 \mathbf{e}_2,$$

and so $\phi_{12} = \cos u_0$. Therefore, the holonomy around the latitude circle (oriented counterclockwise) is $\Delta\psi = -\int_0^{2\pi} \cos u_0\, dt = -2\pi\cos u_0$, confirming our previous results.

Note that if we wish to parametrize the curve by arclength (as will be important shortly), we take $s = (\sin u_0)v$, $0 \le s \le 2\pi \sin u_0$. Then, with respect to this parametrization, we have $\phi_{12}(s) = \cot u_0$. (Why?)

For completeness, we can use Proposition 1.1 to calculate ϕ_{12} as well: With $E = 1$, $G = \sin^2 u$, $u = u_0$, and $v(s) = s/\sin u_0$, we have $\phi_{12} = \dfrac{1}{2\sin u_0}\left(2\sin u_0 \cos u_0 \cdot \dfrac{1}{\sin u_0}\right) = \cot u_0$, as before. ∇

Suppose now that $\boldsymbol{\alpha}$ is an arclength-parametrized curve and let's write $\boldsymbol{\alpha}(s) = \mathbf{x}(u(s), v(s))$ and $\mathbf{T}(s) = \boldsymbol{\alpha}'(s) = \cos\theta(s)\mathbf{e}_1 + \sin\theta(s)\mathbf{e}_2$, $s \in [0, L]$, for a $\mathcal{C}^1$ function $\theta(s)$ (cf. Lemma 3.6 of Chapter 1), as indicated in Figure 1.2. A formula fundamental for the rest of our work is the following:

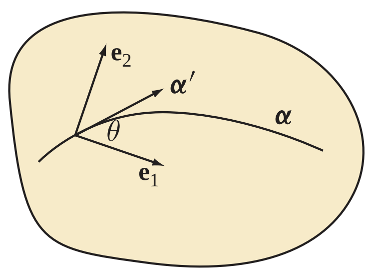


FIGURE 1.2

Proposition 1.3. *When $\boldsymbol{\alpha}$ is an arclength-parametrized curve, the geodesic curvature of $\boldsymbol{\alpha}$ is given by*

$$\kappa_g(s) = \phi_{12}(s) + \theta'(s) = \frac{1}{2\sqrt{EG}}\big(-E_v u'(s) + G_u v'(s)\big) + \theta'(s).$$

Proof. Recall that $\kappa_g = \kappa\mathbf{N}\cdot(\mathbf{n}\times\mathbf{T}) = \mathbf{T}'\cdot(\mathbf{n}\times\mathbf{T})$. Now, since $\mathbf{T} = \cos\theta\mathbf{e}_1 + \sin\theta\mathbf{e}_2$, $\mathbf{n}\times\mathbf{T} = -\sin\theta\mathbf{e}_1 + \cos\theta\mathbf{e}_2$ (why?), and so

$$\begin{aligned}
\kappa_g &= \nabla_{\mathbf{T}}\mathbf{T}\cdot(-\sin\theta\mathbf{e}_1 + \cos\theta\mathbf{e}_2)\\
&= \nabla_{\mathbf{T}}(\cos\theta\mathbf{e}_1 + \sin\theta\mathbf{e}_2)\cdot(-\sin\theta\mathbf{e}_1 + \cos\theta\mathbf{e}_2)\\
&= \big(\cos\theta\nabla_{\mathbf{T}}\mathbf{e}_1 + \sin\theta\nabla_{\mathbf{T}}\mathbf{e}_2\big)\cdot(-\sin\theta\mathbf{e}_1 + \cos\theta\mathbf{e}_2) + \big((-\sin\theta)\theta'(-\sin\theta) + (\cos\theta)\theta'(\cos\theta)\big)\\
&= (\cos^2\theta + \sin^2\theta)(\phi_{12} + \theta') = \phi_{12} + \theta',
\end{aligned}$$

as required. Now the result follows by applying Proposition 1.1 when $\boldsymbol{\alpha}$ is arclength-parametrized. $\square$

Remark. The first equality in Proposition 1.3 should not be surprising in the least. Curvature of a plane curve measures the rate at which its unit tangent vector turns relative to a fixed reference direction. Similarly, the geodesic curvature of a curve in a surface measures the rate at which its unit tangent vector turns relative to a parallel vector field along the curve; θ' measures its turning relative to $\mathbf{e}_1$, which is itself turning at a rate given by ϕ_{12}, so the geodesic curvature is the sum of those two rates.

Now suppose that $\boldsymbol{\alpha}$ is a *closed* curve bounding a region $R \subset M$. We denote the boundary of R by ∂R. Then by Green's Theorem (see Theorem 2.6 of the Appendix), we have

$$\begin{aligned}
\int_0^L \phi_{12}(s)ds &= \int_0^L \frac{1}{2\sqrt{EG}}\big(-E_v u'(s) + G_u v'(s)\big)ds = \int_{\partial R} \frac{1}{2\sqrt{EG}}\big(-E_v du + G_u dv\big) \\
&= \iint_R \left(\left(\frac{E_v}{2\sqrt{EG}}\right)_v + \left(\frac{G_u}{2\sqrt{EG}}\right)_u\right) du dv \\
(\dagger) \qquad &= \iint_R \frac{1}{2\sqrt{EG}}\left(\left(\frac{E_v}{\sqrt{EG}}\right)_v + \left(\frac{G_u}{\sqrt{EG}}\right)_u\right) \underbrace{\sqrt{EG}du dv}_{dA} \\
&= -\iint_R K dA
\end{aligned}$$

by the formula $(*)$ for Gaussian curvature on p. 60. (Recall from the end of Section 1 of Chapter 2 that the element of surface area on a parametrized surface is given by $dA = \|\mathbf{x}_u \times \mathbf{x}_v\| du dv = \sqrt{EG - F^2} du dv$.)

We now see that Gaussian curvature and holonomy are intimately related:

Corollary 1.4. *When R is a region with smooth boundary and lying in an orthogonal parametrization, the holonomy around ∂R is $\Delta\psi = \iint_R K dA$.*

Proof. This follows immediately from Proposition 1.2 and the formula $(\dagger)$ above. $\square$

We conclude further from Proposition 1.3 that

$$\int_{\partial R} \kappa_g ds = \int_{\partial R} \phi_{12} ds + \underbrace{\theta(L) - \theta(0)}_{\Delta\theta},$$

so the total angle through which the tangent vector to ∂R turns is given by

$$\Delta\theta = \int_{\partial R} \kappa_g ds + \iint_R K dA.$$

Now, when R is simply connected (i.e., can be continuously deformed to a point), it is not too surprising that $\Delta\theta = 2\pi$. Intuitively, as we shrink the curve to a point, $\mathbf{e}_1$ becomes almost constant along the curve, but the tangent vector must make one full rotation (as a consequence of the Hopf Umlaufsatz, Theorem 3.5 of Chapter 1). Since $\Delta\theta$ is an integral multiple of 2π that varies continuously as we deform the curve, it must stay equal to 2π throughout.

Corollary 1.5. *If R is a simply connected region lying in an orthogonal parametrization and whose boundary curve is a geodesic, then $\iint_R K dA = \Delta\theta = 2\pi$.*

Example 2. We take R to be the upper hemisphere and use the usual spherical coordinates parametrization. Then the unit tangent vector along ∂R is $\mathbf{e}_2$ everywhere, so $\Delta\theta = 0$, in contradiction with Corollary

1.5. Alternatively, $C = \partial R$ is a geodesic, so there should be zero holonomy around C (computed with respect to this framing).

How do we resolve this paradox? Well, although we've been sloppy about this point, the spherical coordinates parametrization actually fails at the north pole (since $\mathbf{x}_v = \mathbf{0}$). Indeed, there is no framing of the upper hemisphere with $\mathbf{e}_2$ everywhere tangent to the equator. However, the reader can rest assured that there *is* some orthogonal parametrization of the upper hemisphere, e.g., by stereographic projection from the south pole (cf. Example 1(e) in Section 1 of Chapter 2). ∇

Remark. In more advanced courses, the holonomy around the closed curve $\boldsymbol{\alpha}$ is interpreted as a rotation of the tangent plane of M at $\boldsymbol{\alpha}(0)$. That is, what matters is $\Delta\psi \pmod{2\pi}$, i.e., the change in angle disregarding multiples of 2π. This quantity does not depend on the choice of framing $\mathbf{e}_1, \mathbf{e}_2$.

We now set to work on one of the crowning results of surface theory.

Theorem 1.6 (Local Gauss-Bonnet)**.** *Suppose R is a simply connected region with piecewise smooth boundary and lying in an orthogonal parametrization. If $C = \partial R$ has exterior angles ϵ_j, $j = 1, \dots, \ell$, then*

$$\int_{\partial R} \kappa_g \, ds + \iint_R K \, dA + \sum_{j=1}^{\ell} \epsilon_j = 2\pi .$$

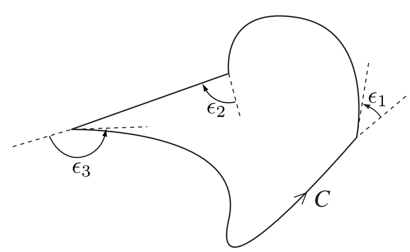


FIGURE 1.3

Note, as we indicate in Figure 1.3, that we measure exterior angles so that $|\epsilon_j| \leq \pi$ for all j.

Proof. If ∂R is smooth, then from our earlier discussion we infer that

$$\int_{\partial R} \kappa_g \, ds + \iint_R K \, dA = \Delta\theta = 2\pi .$$

But when ∂R has corners, the unit tangent vector turns *less* by the amount $\sum_{j=1}^{\ell} \epsilon_j$, so the result follows. (Technically, what we need is the correction of the Hopf Umlaufsatz when the curve has corners. See Exercise 1.3.12.) $\square$

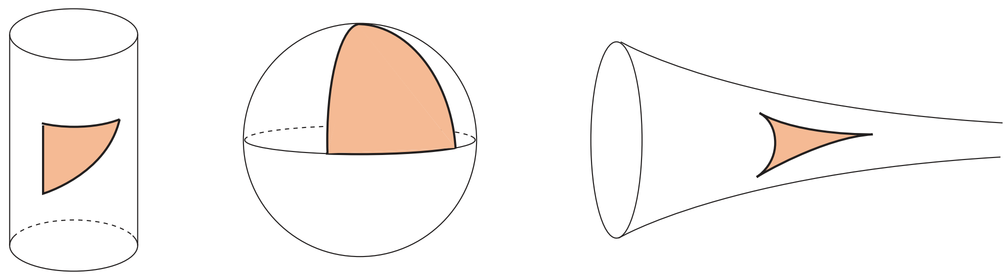
Corollary 1.7. *For a geodesic triangle (i.e., a region whose boundary consists of three geodesic segments) R with* interior *angles $\iota_1, \iota_2, \iota_3$, we have $\iint_R K \, dA = (\iota_1 + \iota_2 + \iota_3) - \pi$, the* angle excess.

Proof. Since the boundary consists of geodesic segments, the geodesic curvature integral drops out, and we are left with

$$\iint_R K \, dA = 2\pi - \sum_{j=1}^{3} \epsilon_j = 2\pi - \sum_{j=1}^{3} (\pi - \iota_j) = \sum_{j=1}^{3} \iota_j - \pi ,$$

as required. $\square$

Remark. It is worthwhile to consider the three special cases $K = 0$, $K = 1$, $K = -1$, as pictured in Figure 1.4. When M is flat, the sum of the angles of a triangle is π, as in the Euclidean case. When M



$K = 0$ $K = 1$ $K = -1$

FIGURE 1.4

is positively curved, it takes more than π for the triangle to close up, and when M is negatively curved, it takes less. Intuitively, this is because geodesics seem to "bow out" when $K > 0$ and "bow in" when $K < 0$ (cf. Exercise 3.2.17).

Example 3. Let's consider Theorem 1.6 in the case of a spherical cap, as shown in Figure 1.5. Using the usual spherical coordinates parametrization, we have $0 \le u \le u_0$. By Proposition 1.3 and Example 1,

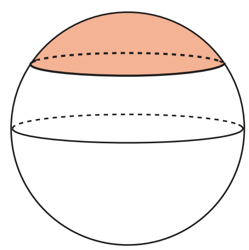


FIGURE 1.5

since $\theta = \pi/2$ along the v-curve, we have $\kappa_g = \phi_{12}(s) = \cot u_0$ (cf. also Exercise 2.4.4). Therefore, we have

$$\iint_R K\,dA = 2\pi - \int_{\partial R} \kappa_g\,ds = 2\pi(1 - \cos u_0),$$

which checks, of course, since $K = 1$ and the area of this cap is indeed

$$\int_0^{2\pi}\int_0^{u_0} \sin u\,du\,dv = 2\pi(1 - \cos u_0). \qquad \nabla$$

Remark. Notice that the sign of κ_g depends on both the orientation of $\boldsymbol{\alpha}$ and the orientation of the surface. If we rescale the surface by a factor of c, then the integral $\int_{\partial R} \kappa_g\,ds$ does not change, as the arclength changes by a factor of c and the geodesic curvature by a factor of $1/c$. Similarly, the integral $\iint_R K\,dA$ does not change when we rescale the surface: Area changes by a factor of c^2 and Gaussian curvature changes by a factor of $1/c^2$.

We now come to one of the crowning results of modern-day mathematics, one which has led to much subsequent research and generalization. We say a surface $M \subset \mathbb{R}^3$ is *oriented* if we have chosen a continuous unit normal field defined everywhere on M. We now consider a compact, oriented surface with

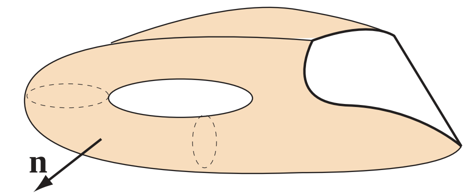


FIGURE 1.6

piecewise-smooth boundary, as pictured in Figure 1.6. T. Radó proved in 1925 that any such surface M can be triangulated. That is, we may write $M = \bigcup_{\lambda=1}^{m} \Delta_\lambda$ where

(i) Δ_λ is the image of a triangle under an (orientation-preserving) orthogonal parametrization;
(ii) $\Delta_\lambda \cap \Delta_\mu$ $(\lambda \neq \mu)$ is either empty, a single vertex, or a single edge;
(iii) when $\Delta_\lambda \cap \Delta_\mu$ consists of a single edge, the orientations of the edge are opposite in Δ_λ and Δ_μ; and
(iv) at most one edge of Δ_λ is contained in the boundary of M.

We now make a standard

Definition. Given a triangulation $\mathcal{T}$ of a surface M with V vertices, E edges, and F faces, we define the *Euler characteristic* $\chi(M, \mathcal{T}) = V - E + F$.

Example 4. We can triangulate a disk as shown in Figure 1.7, obtaining $\chi = 1$. Without being so

V–E+F = 5–8+4 = 1 $\qquad$ V–E+F = 9–18+10 = 1

FIGURE 1.7

pedantic as to require that each Δ_λ be the image of a triangle under an orthogonal parametrization, we might just think of the disk as a single triangle with its edges puffed out; then we would have $\chi = V - E + F = 3 - 3 + 1 = 1$, as well. We leave it to the reader to triangulate a sphere and check that $\chi(\Sigma, \mathcal{T}) = 2$. ∇

Remark. It's important to note that by choosing the orientations on the "triangles" Δ_λ compatibly, we get an orientation on the boundary of M. That is, a choice of **n** on M determines which direction we proceed on ∂M. This is precisely the case any time one deals with Green's Theorem (or its generalization to oriented surfaces, Stokes's Theorem). Nevertheless, following up on the Remark on p. 85, the sign of κ_g on ∂M is independent of the choice of orientation on M, for, if we change **n** to $-$**n**, the orientation on ∂M switches and $\mathbf{n} \times \mathbf{T}$ stays the same.

The beautiful result to which we've been headed is now the following

Theorem 1.8 (Global Gauss-Bonnet)**.** *Let M be a compact, oriented surface with piecewise-smooth boundary, equipped with a triangulation $\mathcal{T}$ as above. If ϵ_k, $k = 1, \dots, \ell$, are the exterior angles of ∂M, then*

$$\int_{\partial M} \kappa_g \, ds + \iint_M K \, dA + \sum_{k=1}^{\ell} \epsilon_k = 2\pi \chi(M, \mathcal{T}).$$

Proof. As we illustrate in Figure 1.8, we will distinguish vertices on the boundary and in the interior, denoting the respective total numbers by V_b and V_i. Similarly, we distinguish among edges on the boundary, edges in the interior, and edges that join a boundary vertex to an interior vertex; we denote the respective

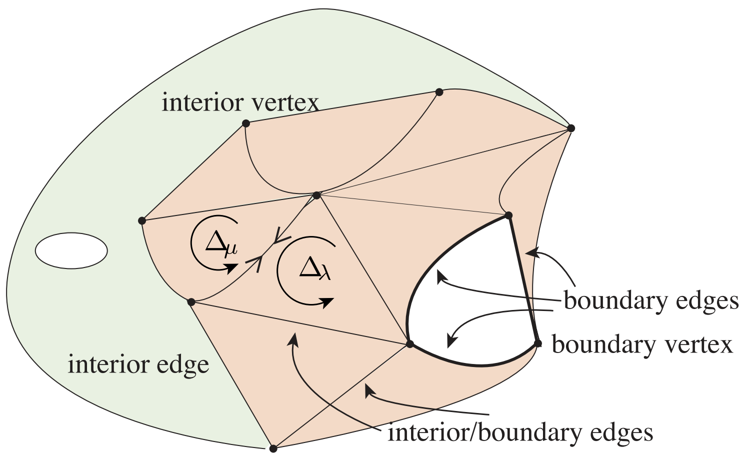


FIGURE 1.8

numbers of these by E_b, E_i, and E_{ib}. Now observe that

$$\iint_M K \, dA = \sum_{\lambda=1}^{m} \iint_{\Delta_\lambda} K \, dA$$

since all the orientations are compatible, and

$$\int_{\partial R} \kappa_g \, ds = \sum_{\lambda=1}^{m} \int_{\partial \Delta_\lambda} \kappa_g \, ds$$

because the line integrals over interior and interior/boundary edges cancel in pairs (recall that κ_g changes sign when we reverse the orientation of the curve). Let $\epsilon_{\lambda j}$, $j = 1, 2, 3$, denote the exterior angles of the "triangle" Δ_λ. Then, applying Theorem 1.6 to Δ_λ, we have

$$\int_{\partial \Delta_\lambda} \kappa_g \, ds + \iint_{\Delta_\lambda} K \, dA + \sum_{j=1}^{3} \epsilon_{\lambda j} = 2\pi,$$

and now, summing over the triangles, we obtain

$$\int_{\partial M} \kappa_g \, ds + \iint_M K \, dA + \sum_{\lambda=1}^{m} \sum_{j=1}^{3} \epsilon_{\lambda j} = 2\pi m = 2\pi F.$$

Now we must do some careful accounting: Letting $\iota_{\lambda j}$ denote the respective interior angles of triangle Δ_λ, we have

$$(*) \qquad \sum_{\substack{\text{interior}\\ \text{vertices}}} \epsilon_{\lambda j} = \sum_{\substack{\text{interior}\\ \text{vertices}}} (\pi - \iota_{\lambda j}) = \pi(2E_i + E_{ib}) - 2\pi V_i$$

inasmuch as each interior edge contributes two interior vertices, whereas each interior/boundary edge contributes just one, and the interior angles at each interior vertex sum to 2π. Next,

$$(**) \qquad \sum_{\substack{\text{boundary}\\ \text{vertices}}} \epsilon_{\lambda j} = \pi E_{ib} + \sum_{k=1}^{\ell} \epsilon_k .$$

To see this, we reason as follows. Given a boundary vertex v, denote by a superscript (v) the relevant angle or number for which the vertex v is involved. Note first of all that any boundary vertex v is contained in $E_{ib}^{(v)} + 1$ faces. Moreover, for a fixed boundary vertex v,

$$\sum \iota_{\lambda j}^{(v)} = \begin{cases} \pi , & v \text{ a smooth boundary vertex} \\ \pi - \epsilon_k , & v \text{ a corner of } \partial M \text{ with exterior angle } \epsilon_k \end{cases} .$$

Thus,

$$\begin{aligned} \sum_{\substack{\text{boundary}\\ \text{vertices}}} \epsilon_{\lambda j} = \sum_{\substack{\text{boundary}\\ \text{vertices } v}} (\pi - \iota_{\lambda j}) &= \sum_{\substack{\text{boundary}\\ \text{vertices } v}} \pi(E_{ib}^{(v)} + 1) - \Big(\sum_{v \text{ smooth}} \iota_{\lambda j} + \sum_{v \text{ corner}} \iota_{\lambda j} \Big) \\ &= \pi E_{ib} + \sum_{k=1}^{\ell} \epsilon_k . \end{aligned}$$

Adding equations $(*)$ and $(**)$ yields

$$\sum_{\lambda, j} \epsilon_{\lambda j} = \sum_{\substack{\text{interior}\\ \text{vertices}}} \epsilon_{\lambda j} + \sum_{\substack{\text{boundary}\\ \text{vertices}}} \epsilon_{\lambda j} = 2\pi(E_i + E_{ib} - V_i) + \sum_{k=1}^{\ell} \epsilon_k .$$

At long last, therefore, our reckoning concludes:

$$\begin{aligned} \int_{\partial M} \kappa_g \, ds + \iint_M K \, dA + \sum_{k=1}^{\ell} \epsilon_k &= 2\pi\big(F - (E_i + E_{ib}) + V_i\big) \\ &= 2\pi\big(F - (E_i + E_{ib} + E_b) + (V_i + V_b)\big) = 2\pi(V - E + F) \\ &= 2\pi\chi(M, \mathcal{T}). \end{aligned}$$

(Note that because the boundary curve ∂M is closed, we have $V_b = E_b$.) □

We now derive some interesting conclusions:

Corollary 1.9. *The Euler characteristic $\chi(M, \mathcal{T})$ does not depend on the triangulation $\mathcal{T}$ of M.*

Proof. The left-hand side of the equality in Theorem 1.8 has nothing whatsoever to do with the triangulation. □

It is therefore legitimate to denote the Euler characteristic by $\chi(M)$, with no reference to the triangulation. It is proved in a course in algebraic topology that the Euler characteristic is a "topological invariant"; i.e., if we deform the surface M in a bijective, continuous manner (so as to obtain a *homeomorphic* surface), the Euler characteristic does not change. We therefore deduce:

Corollary 1.10. *The quantity*

$$\int_{\partial M} \kappa_g\, ds + \iint_M K\, dA + \sum_{k=1}^{\ell} \epsilon_k$$

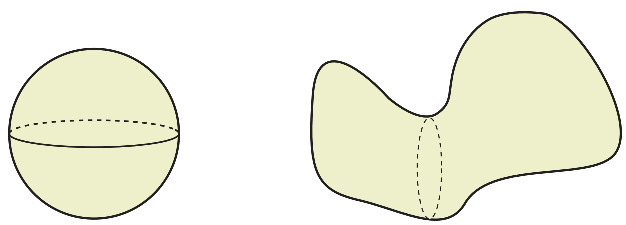
is a topological invariant, i.e., does not change as we deform the surface M.

In particular, in the event that $\partial M = \emptyset$ (so many people refer to the surface M as a *closed* surface), we have

Corollary 1.11. *When M is a compact, oriented surface without boundary, we have*

$$\iint_M K\, dA = 2\pi\chi(M).$$

It is very interesting that the *total curvature* does not change as we deform the surface, for example, as shown in Figure 1.9. In a topology course, one proves that any compact, oriented surface without boundary must



$\iint_M K\, dA = 4\pi$

FIGURE 1.9

have the topological type of a sphere or of a g-holed torus for some positive integer g. Thus (cf. Exercise 4), the possible Euler characteristics of such a surface are 2, 0, -2, -4, ...; moreover, the integral $\iint_M K\, dA$ determines the topological type of the surface.

We conclude this section with a few applications of the Gauss-Bonnet Theorem.

Example 5. Suppose M is a surface of nonpositive Gaussian curvature. Then there cannot be a geodesic 2-gon R on M that bounds a simply connected region. For if there were, by Theorem 1.6 we would have

$$0 \geq \iint_R K\, dA = 2\pi - (\epsilon_1 + \epsilon_2) > 0,$$

which is a contradiction. (Note that the exterior angles must be strictly less than π because there is a unique (smooth) geodesic with a given tangent direction.) ∇

Example 6. Suppose M is topologically equivalent to a cylinder and its Gaussian curvature is negative. Then there is at most one simple closed geodesic in M. Note, first, as indicated in Figure 1.10, that if there is a simple closed geodesic $\boldsymbol{\alpha}$, either it must separate M into two unbounded pieces or else it bounds

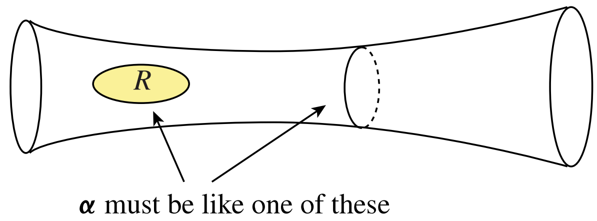


FIGURE 1.10

a disk R, in which case we would have $0 > \iint_R K\,dA = 2\pi\chi(R) = 2\pi$, which is a contradiction. On the other hand, suppose there were two. If they don't intersect, then they bound a cylinder R and we get $0 > \iint_R K\,dA = 2\pi\chi(R) = 0$, which is a contradiction. If they do intersect, then we we have a geodesic 2-gon bounding a simply connected region, which cannot happen by Example 5. $\triangledown$

EXERCISES 3.1

1. Compute the holonomy around the parallel $u = u_0$ (and indicate which direction the rotation occurs from the viewpoint of an observer away from the surface down the x-axis) on
   *a. the torus $\mathbf{x}(u, v) = \big((a + b\cos u)\cos v, (a + b\cos u)\sin v, b\sin u\big)$
   b. the paraboloid $\mathbf{x}(u, v) = (u\cos v, u\sin v, u^2)$
   c. the catenoid $\mathbf{x}(u, v) = (\cosh u\cos v, \cosh u\sin v, u)$

*2. Determine whether there can be a (smooth) closed geodesic on a surface when
   a. $K > 0$
   b. $K = 0$
   c. $K < 0$

   If the closed geodesic can bound a simply connected region, give an example.

3. Calculate the Gaussian curvature of a torus (as parametrized in Example 1(c) of Section 1 of Chapter 2) and verify Corollary 1.11.

4. a. Triangulate a cylinder, a sphere, a torus, and a two-holed torus; verify that $\chi = 0$, 2, 0, and -2, respectively. Pay particular attention to condition (ii) in the definition of triangulation.
   b. Prove by induction that a g-holed torus has $\chi = 2 - 2g$.

5. Suppose M is a compact, oriented surface without boundary that is *not* of the topological type of a sphere. Prove that there are points in M where Gaussian curvature is positive, zero, and negative.

6. Consider a surface with $K > 0$ that is topologically a cylinder. Prove that there cannot be two disjoint simple closed geodesics both going around the neck of the surface.

7. Suppose M and M^* are locally isometric and compatibly oriented. Use Proposition 1.3 to prove that if $\boldsymbol{\alpha}$ and $\boldsymbol{\alpha}^*$ are corresponding arclength-parametrized curves, then their geodesic curvatures are equal at corresponding points.

8. Consider the paraboloid M parametrized by $\mathbf{x}(u,v) = (u\cos v, u\sin v, u^2)$, $0 \le u$, $0 \le v \le 2\pi$. Denote by M_r that portion of the paraboloid defined by $0 \le u \le r$.
   a. Calculate the geodesic curvature of the boundary circle and compute $\displaystyle\int_{\partial M_r} \kappa_g \, ds$.
   b. Calculate $\chi(M_r)$.
   c. Use the Gauss-Bonnet Theorem to compute $\displaystyle\iint_{M_r} K \, dA$. Find the limit as $r \to \infty$. (This is the total curvature of the paraboloid.)
   d. Calculate K directly (however you wish) and compute $\displaystyle\iint_M K \, dA$ explicitly.
   e. Explain the relation between the total curvature and the image of the Gauss map of M.

9. Consider the pseudosphere (with boundary) M parametrized as in Example 8 of Chapter 2, Section 2, but here we take $u \ge 0$. Denote by M_r that portion defined by $0 \le u \le r$. (Note that we are including the boundary circle $u = 0$.)
   a. Calculate the geodesic curvature of the circle $u = u_0$ and compute $\displaystyle\int_{\partial M_r} \kappa_g \, ds$. Watch out for the orientations of the two circles.
   b. Calculate $\chi(M_r)$.
   c. Use the Gauss-Bonnet Theorem to compute $\displaystyle\iint_{M_r} K \, dA$. Find the limit as $r \to \infty$. (This is the total curvature of the pseudosphere.)
   d. Calculate the area of M_r directly, and use this to deduce the value of $\displaystyle\iint_M K \, dA$.
   e. Explain the relation between the total curvature and the image of the Gauss map of M.

10. Give a different version of the accounting to prove Theorem 1.8 as follows.
    a. Show that $3F = 2(E_i + E_{ib}) + E_b$, and conclude that $3F = 2E - V_b$.
    b. Show that $\sum_{\text{interior vertices}} \iota_{\lambda j} = 2\pi V_i$ and $\sum_{\text{boundary vertices}} \iota_{\lambda j} = \pi V_b - \sum \epsilon_k$.
    c. Conclude that $\sum_{\lambda,j} \epsilon_{\lambda j} = 3\pi F - \sum_{\lambda,j} \iota_{\lambda j} = 2\pi(E - V) + \sum \epsilon_k$ and complete the proof of the theorem.

11. a. Use Corollary 1.4 to prove that M is flat if and only if the holonomy around all ("small") closed curves that bound a region in M is zero.
    b. Show that even on a flat surface, holonomy can be nontrivial around certain curves.

12. Reprove the result of part a of Exercise 2.3.14 by considering the holonomy around a (sufficiently small) quadrilateral formed by four of the lines. Does the result hold if there are two families of *geodesics* in M always intersecting at right angles?

13. In this exercise we explore what happens when we try to apply the Gauss-Bonnet Theorem to the simplest non-smooth surface, a right circular cone. Let R denote the surface given in Exercise 2.4.5 and ∂R its boundary curve.
    a. Show that if we make R by gluing the edges of a circular sector ("pacman") of central angle β, as indicated in Figure 1.11, then $\displaystyle\int_{\partial R} \kappa_g \, ds = 2\pi \sin\phi = \beta$. We call β the *cone angle* of R at its vertex.
    b. Show that Theorem 1.6 holds for R if we add $2\pi - \beta$ to $\iint_R K \, dA$.

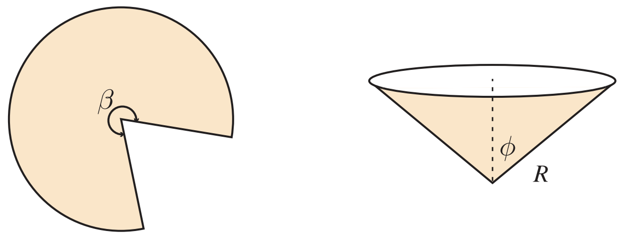


FIGURE 1.11

c. Show that we obtain the same result by "smoothing" the cone point, as pictured in Figure 1.12. (Hint: Interpret $\iint_R K\,dA$ as the area of the image of the Gauss map.)

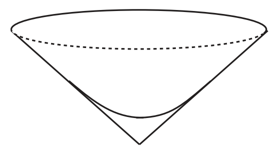


FIGURE 1.12

Remark. It is not hard to give an explicit $\mathcal{C}^2$ such smoothing. For example, construct a $\mathcal{C}^2$ convex function f on $[0, 1]$ with $f(0) = f'(0) = 0$, $f(1) = f'(1) = 1$, and $f''(1) = 0$.

14. Suppose $\boldsymbol{\alpha}$ is a closed space curve with $\kappa \neq 0$. Assume that the *normal indicatrix* (i.e., the curve traced out on the unit sphere by the principal normal) is a simple closed curve in the unit sphere. Prove then that it divides the unit sphere into two regions of equal area. (Hint: Apply the Gauss-Bonnet Theorem to one of those regions.)

15. Suppose $M \subset \mathbb{R}^3$ is a compact, oriented surface with no boundary with $K > 0$. It follows that M is topologically a sphere (why?). Prove that M is convex; i.e., for each $P \in M$, M lies on only one side of the tangent plane $T_P M$. (Hint: Use the Gauss-Bonnet Theorem and Gauss's original interpretation of curvature indicated in the remark on p. 51 to show the Gauss map must be one-to-one (except perhaps on a subset with no area). Then look at the end of the proof of Theorem 3.4 of Chapter 1.)

2. An Introduction to Hyperbolic Geometry

Hilbert proved in 1901 that there is no surface (without boundary) in $\mathbb{R}^3$ with constant negative curvature with the property that it is a closed subset of $\mathbb{R}^3$ (i.e., every Cauchy sequence of points in the surface converges to a point of the surface). The pseudosphere fails the latter condition. Nevertheless, it is possible to give a definition of an "abstract surface" (not sitting inside $\mathbb{R}^3$) together with a first fundamental form. As we know, this will be all we need to calculate Christoffel symbols, curvature (Theorem 3.1 of Chapter 2), geodesics, and so on.

Definition. The *hyperbolic plane* $\mathbb{H}$ is defined to be the half-plane $\{(u, v) \in \mathbb{R}^2 : v > 0\}$, equipped with the first fundamental form I given by $E = G = 1/v^2$, $F = 0$.

Now, using the formulas (‡) on p. 58, we find that

$$\Gamma^u_{uu} = \frac{E_u}{2E} = 0 \qquad \Gamma^v_{uu} = -\frac{E_v}{2G} = \frac{1}{v}$$
$$\Gamma^u_{uv} = \frac{E_v}{2E} = -\frac{1}{v} \qquad \Gamma^v_{uv} = \frac{G_u}{2G} = 0$$
$$\Gamma^u_{vv} = -\frac{G_u}{2E} = 0 \qquad \Gamma^v_{vv} = \frac{G_v}{2G} = -\frac{1}{v}.$$

Using the formula ($*$) for Gaussian curvature on p. 60, we find

$$K = -\frac{1}{2\sqrt{EG}}\left(\left(\frac{E_v}{\sqrt{EG}}\right)_v + \left(\frac{G_u}{\sqrt{EG}}\right)_u\right) = -\frac{v^2}{2}\left(-\frac{2}{v^3}\cdot v^2\right)_v = -\frac{v^2}{2}\cdot\frac{2}{v^2} = -1.$$

Thus, the hyperbolic plane has constant curvature -1. Note that it is a consequence of Corollary 1.7 that the area of a geodesic triangle in $\mathbb{H}$ is equal to $\pi - (\iota_1 + \iota_2 + \iota_3)$.

What are the geodesics in this surface? Using the equations (♣♣) on p. 71, we obtain the equations

$$u'' - \frac{2}{v}u'v' = v'' + \frac{1}{v}(u'^2 - v'^2) = 0.$$

Obviously, the vertical rays $u = \text{const}$ give us solutions (with $v(t) = c_1 e^{c_2 t}$). Next we seek geodesics with $u' \neq 0$, so we start with $\dfrac{dv}{du} = \dfrac{v'}{u'}$ and apply the chain rule judiciously:

$$\begin{aligned}\frac{d^2v}{du^2} &= \frac{d}{du}\left(\frac{v'}{u'}\right) = \frac{u'v'' - u''v'}{u'^2}\cdot\frac{1}{u'} \\ &= \frac{1}{u'^3}\left(u'\left(\frac{1}{v}\right)(v'^2 - u'^2) - v'\left(\frac{2}{v}u'v'\right)\right) \\ &= -\frac{1}{v}\left(1 + \left(\frac{v'}{u'}\right)^2\right) = -\frac{1}{v}\left(1 + \left(\frac{dv}{du}\right)^2\right).\end{aligned}$$

This means we are left with the differential equation

$$v\frac{d^2v}{du^2} + \left(\frac{dv}{du}\right)^2 = \frac{d}{du}\left(v\frac{dv}{du}\right) = -1,$$

and integrating this twice gives us the solutions

$$u^2 + v^2 = au + b.$$

That is, the geodesics in $\mathbb{H}$ are the vertical rays and the semicircles centered on the u-axis, as pictured in Figure 2.1. Note that any semicircle centered on the u-axis intersects each vertical line at most one time. It now follows that any two points $P, Q \in \mathbb{H}$ are joined by a unique geodesic. If P and Q lie on a vertical line, then the vertical ray through them is the unique geodesic joining them. If P and Q do not lie on a vertical line, let C be the intersection of the perpendicular bisector of $\overline{PQ}$ and the u-axis; then the semicircle centered at C is the unique geodesic joining P and Q.

Example 1. Given $P, Q \in \mathbb{H}$, we would like to find a formula for the (geodesic) distance $d(P, Q)$ between them. Let's start with $P = (u_0, a)$ and $Q = (u_0, b)$, with $0 < a < b$. Parametrizing the line segment from P to Q by $u = u_0$, $v = t$, $a \le t \le b$, we have

$$d(P, Q) = \int_a^b \sqrt{Eu'(t)^2 + Gv'(t)^2}\,dt = \int_a^b \frac{dt}{t} = \ln\frac{b}{a}.$$

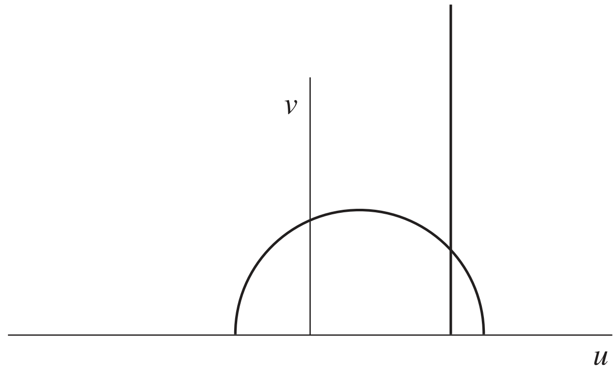


FIGURE 2.1

Note that, fixing Q and letting P approach the u-axis, $d(P,Q) \to \infty$; thus, it is reasonable to think of points on the u-axis as "virtual" points at infinity.

In general, we parametrize the arc of a semicircle $(u_0 + r\cos t, r\sin t)$, $\theta_1 \le t \le \theta_2$, going from P to

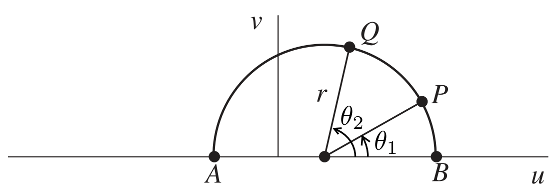


FIGURE 2.2

Q, as shown in Figure 2.2. Then we have

$$\begin{aligned} d(P,Q) &= \left| \int_{\theta_1}^{\theta_2} \sqrt{Eu'(t)^2 + Gv'(t)^2}\,dt \right| = \left| \int_{\theta_1}^{\theta_2} \frac{r\,dt}{r\sin t} \right| = \left| \int_{\theta_1}^{\theta_2} \frac{dt}{\sin t} \right| \\ &= \left| \ln\left(\frac{1+\cos\theta_1}{\sin\theta_1} \Big/ \frac{1+\cos\theta_2}{\sin\theta_2} \right) \right| = \left| \ln\left(\frac{2\cos(\theta_1/2)}{2\sin(\theta_1/2)} \Big/ \frac{2\cos(\theta_2/2)}{2\sin(\theta_2/2)} \right) \right| \\ &= \left| \ln\left(\frac{AP}{BP} \Big/ \frac{AQ}{BQ} \right) \right|, \end{aligned}$$

where the lengths in the final formula are Euclidean. (See Exercise 12 for the connection with cross ratio.) ∇

It follows from the first part of Example 1 that the curves $v = a$ and $v = b$ are a constant distance apart (measured along geodesics orthogonal to both), like parallel lines in Euclidean geometry. These curves are classically called *horocycles*. As we see in Figure 2.3, these curves are the curves orthogonal to the family of the "vertical geodesics." If, instead, we consider all the geodesics passing through a given point Q "at infinity" on $v = 0$, as we ask the reader to check in Exercise 5, the orthogonal trajectories will be curves in $\mathbb{H}$ represented by circles tangent to the u-axis at Q.

Example 2. Let's calculate the geodesic curvature of the horocycle $v = a$, oriented to the right. We start by parametrizing the curve by $\boldsymbol{\alpha}(t) = (t, a)$. Then $\boldsymbol{\alpha}'(t) = (1, 0)$. Note that $\upsilon(t) = \|\boldsymbol{\alpha}'(t)\| =$

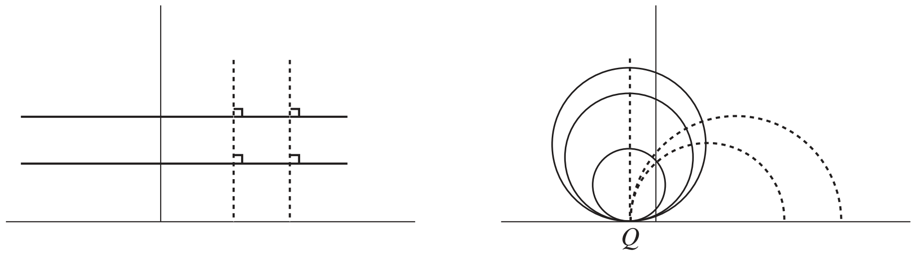


FIGURE 2.3

$\sqrt{E(1)^2 + G(0)^2} = 1/a$. By Proposition 1.1,

$$\phi_{12} = \frac{1}{2\sqrt{\frac{1}{a^4}}}(2a^{-3} \cdot 1) = \frac{1}{a}.$$

(Here $\mathbf{e}_1 = v(1,0)$ and $\mathbf{e}_2 = v(0,1)$ at the point $(u,v) \in \mathbb{H}$. Why?) To calculate the geodesic curvature, we wish to apply Proposition 1.3, which requires differentiation with respect to arclength, so we'll use the chain rule as in Chapter 1, multiplying the t-derivative by $1/v(t) = a$. Note, also, that $\boldsymbol{\alpha}'$ makes the constant angle $\theta = 0$ with $\mathbf{e}_1$, so $\theta' = 0$. Thus,

$$\kappa_g = \frac{1}{v(t)}\phi_{12} = a \cdot \frac{1}{a} = 1,$$

as required. (Note that if we move to the left, the sign changes and $\kappa_g = -1$.) ∇

We ask the reader to do the analogous calculations for the circles tangent to the u-axis in Exercise 6. Moreover, as we ask the reader to check in Exercise 7, every curve in $\mathbb{H}$ of constant geodesic curvature $\kappa_g = \pm 1$ is a horocycle.

Remark. It seems somewhat surprising to find in Example 2 that $\phi_{12} = 1/a$, as $\mathbf{e}_1$ certainly doesn't appear to be turning as we move along the path. However, as we discussed in the Remark on p. 71, at any point of $v = a$ the geodesic with the same tangent vector is a semicircle heading "to the right," and so this means that $\mathbf{e}_1$ is turning to the left, i.e., towards $\mathbf{e}_2$.

The isometries of the Euclidean plane form a group, the Euclidean group $E(2)$; the isometries of the sphere likewise form a group, the orthogonal group $O(3)$. Each of these is a 3-dimensional Lie group. Intuitively, there are three degrees of freedom because we must specify where a point P goes (two degrees of freedom) and where a single unit tangent vector at that point P goes (one more degree of freedom). We might likewise expect the isometries of $\mathbb{H}$ to form a 3-dimensional group. And indeed it is. We deal with just the orientation-preserving isometries here.

We consider $\mathbb{H} \subset \mathbb{C}$ by letting (u,v) correspond to $z = u + iv$, and we consider the collection of *linear fractional transformations*

$$T(z) = \frac{az+b}{cz+d}, \qquad a,b,c,d \in \mathbb{R}, \quad ad - bc = 1.$$

We must now check several things:

(i) Composition of functions corresponds to multiplication of the 2×2 matrices $\begin{bmatrix} a & b \\ c & d \end{bmatrix}$ with determinant 1, so we obtain a group.

(ii) T maps $\mathbb{H}$ bijectively to $\mathbb{H}$.

(iii) T is an isometry of $\mathbb{H}$.

We leave it to the reader to check the first two in Exercise 8, and we check the third here. Given the point $z = u + iv$, we want to compute the lengths of the vectors T_u and T_v at the image point $T(z) = x + iy$ and see that the two vectors are orthogonal. Note that

$$\frac{az+b}{cz+d} = \frac{(az+b)(c\overline{z}+d)}{|cz+d|^2} = \frac{\big(a(u+iv)+b\big)\big(c(u-iv)+d\big)}{|cz+d|^2}$$
$$= \frac{\big(ac(u^2+v^2)+(ad+bc)u+bd\big)+i\big((ad-bc)v\big)}{|cz+d|^2},$$

so $y = \dfrac{v}{|cz+d|^2}$. Now we have[3]

$$x_u + iy_u = -ix_v + y_v = T'(z) = \frac{(cz+d)a-(az+b)c}{(cz+d)^2} = \frac{1}{(cz+d)^2},$$

so we have

$$\tilde{E} = \frac{x_u^2+y_u^2}{y^2} = \frac{1}{y^2}|T'(z)|^2 = \frac{1}{y^2}\cdot\frac{1}{|cz+d|^4} = \frac{1}{v^2} = E,$$

and, similarly, $\tilde{G} = \dfrac{x_v^2+y_v^2}{y^2} = G$. On the other hand,

$$\tilde{F} = \frac{x_u y_u + x_v y_v}{y^2} = \frac{x_u(-x_v)+x_v(x_u)}{y^2} = 0 = F,$$

as desired.

Now, as we verify in Exercise 12 or in Exercise 14, linear fractional transformations carry lines and circles in $\mathbb{C}$ to either lines or circles. Since our particular linear fractional transformations preserve the real axis ($\cup\{\infty\}$) and preserve angles as well, it follows that vertical lines and semicircles centered on the real axis map to one another. Thus, our isometries do in fact map geodesics to geodesics (how comforting!).

If we think of $\mathbb{H}$ as modeling non-Euclidean geometry, with lines in our geometry being the geodesics, note that given any line ℓ and point $P \notin \ell$, there are *infinitely many* lines passing through P "parallel" to (i.e., not intersecting) ℓ. As we see in Figure 2.4, there are two special lines through P that "meet ℓ at

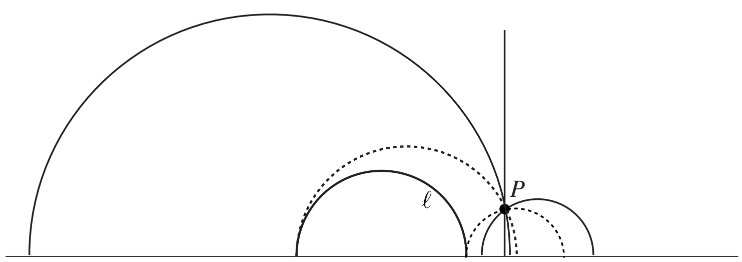


FIGURE 2.4

infinity"; the rest are often called *ultraparallels*.

We conclude with an interesting application. As we saw in the previous section, the Gauss-Bonnet Theorem gives a deep relation between the total curvature of a surface and its topological structure (Euler

[3]These are the Cauchy-Riemann equations from basic complex analysis.

characteristic). We know that if a compact surface M is topologically equivalent to a sphere, then its total curvature must be that of a round sphere, namely 4π. If M is topologically equivalent to a torus, then (as the reader checked in Exercise 3.1.3) its total curvature must be 0. We know that there is no way of making

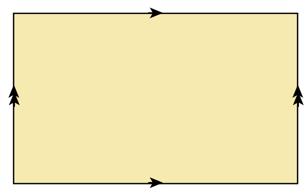


FIGURE 2.5

the torus in $\mathbb{R}^3$ in such a way that it has constant Gaussian curvature $K = 0$ (why?), but we *can* construct a flat torus in $\mathbb{R}^4$ by taking

$$\mathbf{x}(u, v) = (\cos u, \sin u, \cos v, \sin v), \qquad 0 \le u, v \le 2\pi.$$

(We take a piece of paper and identify opposite edges, as indicated in Figure 2.5; this can be rolled into a cylinder in $\mathbb{R}^3$ but into a torus only in $\mathbb{R}^4$.) So what happens with a 2-holed torus? In that case, $\chi(M) = -2$, so the total curvature should be -4π, and we can reasonably ask if there's a 2-holed torus with *constant* negative curvature. Note that we can obtain a 2-holed torus by identifying pairs of edges on an octagon, as

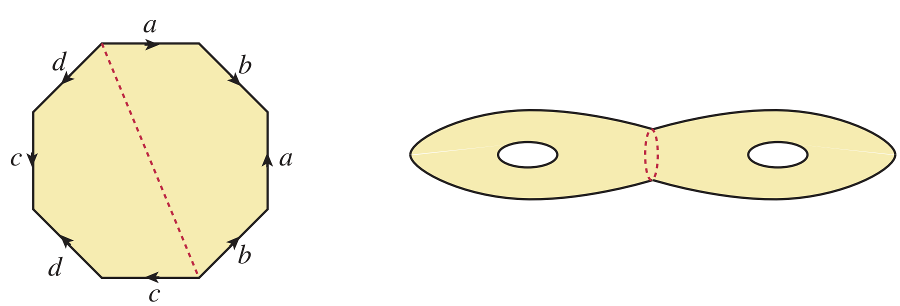


FIGURE 2.6

shown in Figure 2.6.

This leads us to wonder whether we might have regular n-gons R in $\mathbb{H}$. By the Gauss-Bonnet formula, we would have $\text{area}(R) = (n-2)\pi - \sum \iota_j$, so it's obviously necessary that $\sum \iota_j < (n-2)\pi$. This shouldn't be difficult so long as $n \ge 3$. First, let's convince ourselves that, given any point $P \in \mathbb{H}$, $0 < \alpha < \pi$, and $0 < \beta < (\pi - \alpha)/2$, we can construct an isosceles triangle with vertex angle α at P and base angle β. We draw two geodesics emanating from P with angle α between them, as shown in Figure 2.7. Proceeding a geodesic distance r on each of them to points Q and R, we then obtain an isosceles triangle $\triangle PQR$ with vertex angle α. Now, the base angle of that triangle approaches $(\pi - \alpha)/2$ as $r \to 0^+$ and approaches 0 as $r \to \infty$. It follows (presuming that the angle varies continuously with r) that for some r, we obtain the desired base angle β. Let's now apply this construction with $\alpha = 2\pi/n$ and $\beta = \pi/n$, $n \ge 5$. Repeating the construction n times (dividing the angle at P into n angles of $2\pi/n$ each), we obtain a regular n-gon with the property that $\sum \iota_j = 2\pi$, as shown (approximately?) in Figure 2.8 for the case $n = 8$. The point is that because the interior angles add up to 2π, when we identify edges as in Figure 2.6, we will obtain a

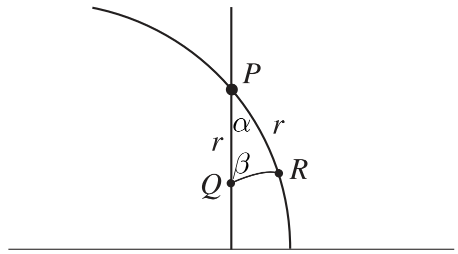


FIGURE 2.7

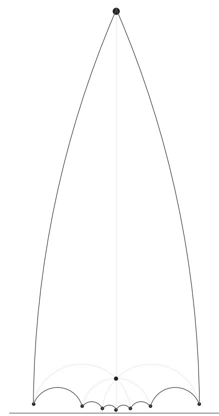


FIGURE 2.8

smooth 2-holed torus with constant curvature $K = -1$. The analogous construction works for the g-holed torus, constructing a regular $4g$-gon whose interior angles sum to 2π.

EXERCISES 3.2

1. Find the geodesic joining P and Q in $\mathbb{H}$ and calculate $d(P, Q)$.
   a. $P = (4, 3)$, $Q = (-3, 4)$
   *b. $P = (1, 2)$, $Q = (0, 1)$
   c. $P = (20, 7)$, $Q = (16, 15)$

2. Suppose there is a geodesic perpendicular to two geodesics in $\mathbb{H}$. What can you prove about the latter two?

3. Prove the angle-angle-angle *congruence* theorem for hyperbolic (geodesic) triangles: If $\angle A \cong \angle A'$, $\angle B \cong \angle B'$, and $\angle C \cong \angle C'$, then $\triangle ABC \cong \triangle A'B'C'$. (Hint: Use an isometry to move A' to A, B' along the geodesic from A to B, and C' along the geodesic from A to C.)

4. a. Verify Local Gauss-Bonnet, Theorem 1.6, for the region R bounded by $u = A$, $u = B$, $v = a$, and $v = b$.
   b. Verify Local Gauss-Bonnet for the region R bounded by the segment $v = a$, $A \leq u \leq B$, and the geodesic joining the two endpoints.

c. Use Local Gauss-Bonnet (and the analysis of part b) to deduce the result of Example 2.

5. Show that the circles tangent to the u-axis at the origin are the orthogonal trajectories of the family of geodesics $u^2 - 2cu + v^2 = 0$, $c \in \mathbb{R}$ (together with the positive v-axis). (Hint: Remember that orthogonal lines have slopes that are negative reciprocals. Eliminate c to obtain the differential equation $\dfrac{dv}{du} = \dfrac{2uv}{u^2 - v^2}$, and solve this "homogeneous" differential equation by substituting $v = uz$ and getting a separable differential equation for u and z.)

6. a. Prove that circles tangent to the u-axis have $\kappa_g = 1$.
   b. Prove that the horocycles $u^2 + v^2 - 2av = 0$ and $u^2 + v^2 - 2bv = 0$ are a constant geodesic distance apart. (Hint: Consider the intersections of the two horocycles with a geodesic $u^2 - 2cu + v^2 = 0$ orthogonal to them both.)

7. Prove that every curve in $\mathbb{H}$ of constant geodesic curvature $\kappa_g = 1$ is either a horizontal line (as in Example 2) or a circle tangent to the u-axis. (Hints: Assume we start with an arclength parametrization $(u(s), v(s))$, and use Proposition 1.3 to show that we have $1 = \dfrac{u'}{v} + \theta'$ and $u'^2 + v'^2 = v^2$. Obtain the differential equation

$$v\frac{d^2v}{du^2} = \left(1 + \left(\frac{dv}{du}\right)^2\right)^{3/2} - \left(\frac{dv}{du}\right)^2 - 1,$$

and solve this by substituting $z = dv/du$ and getting a separable differential equation for dz/dv.)

8. Let $T_{a,b,c,d}(z) = \dfrac{az+b}{cz+d}$, $a, b, c, d \in \mathbb{R}$, with $ad - bc = 1$.
   a. Suppose $a', b', c', d' \in \mathbb{R}$ and $a'd' - b'c' = 1$. Check that

$$T_{a',b',c',d'} \circ T_{a,b,c,d} = T_{a'a+b'c,a'b+b'd,c'a+d'c,c'b+d'd} \quad \text{and}$$
$$(a'a + b'c)(c'b + d'd) - (a'b + b'd)(c'a + d'c) = 1.$$

   Show, moreover, that $T_{d,-b,-c,a} = T^{-1}_{a,b,c,d}$. (Note that $T_{a,b,c,d} = T_{-a,-b,-c,-d}$. The reader who's taken group theory will recognize that we're defining an isomorphism between the group of linear fractional transformations and the group $SL(2,\mathbb{R})/\{\pm I\}$ of 2×2 matrices with determinant 1, identifying a matrix and its additive inverse.)
   b. Let $T = T_{a,b,c,d}$. Prove that if $z = u + iv$ and $v > 0$, then $T(z) = x + iy$ with $y > 0$. Deduce that T maps $\mathbb{H}$ to itself bijectively.

9. Show that *reflection* across the geodesic $u = 0$ is given by $r(z) = -\bar{z}$. Use this to determine the form of the reflection across a general geodesic.

10. The geodesic circle of radius R centered at P is the set of points Q so that $d(P, Q) = R$. Prove that geodesic circles in $\mathbb{H}$ are Euclidean circles. One way to proceed is as follows: The geodesic circle centered at $P = (0, 1)$ with radius $R = \ln a$ must pass through $(0, a)$ and $(0, 1/a)$, and hence ought to be a Euclidean circle centered at $(0, \frac{1}{2}(a + 1/a))$. Check that all the points on this circle are in fact a hyperbolic distance R away from P. (Hint: It is probably easiest to work with the cartesian equation of the circle. Find the equation of the geodesic through P and an arbitrary point of the circle.)

*11. What is the geodesic curvature of a geodesic circle of radius R in $\mathbb{H}$? (See Exercise 10.)

12. Recall (see, for example, p. 298 and pp. 350–1 of Shifrin's *Abstract Algebra: A Geometric Approach*) that the *cross ratio* of four numbers $A, B, P, Q \in \mathbb{C} \cup \{\infty\}$ is defined to be

$$[A : B : P : Q] = \frac{Q - A}{P - A} \Big/ \frac{Q - B}{P - B}.$$

   a. Show that A, B, P, and Q lie on a line or circle if and only if their cross ratio is a real number.
   b. Prove that if S is a linear fractional transformation with $S(A) = 0$, $S(B) = \infty$, and $S(P) = 1$, then $S(Q) = [A : B : P : Q]$. Use this to deduce that for any linear fractional transformation T, we have $[T(A) : T(B) : T(P) : T(Q)] = [A : B : P : Q]$.
   c. Prove that linear fractional transformations map lines and circles to either lines or circles. (For which such transformations do lines necessarily map to lines?)
   d. Show that if A, B, P, and Q lie on a line or circle, then

$$|[A : B : P : Q]| = \frac{AQ}{AP} \Big/ \frac{BQ}{BP}.$$

   Conclude that $d(P, Q) = |\ln[A : B : P : Q]|$, where A, B, P, and Q are as illustrated in Figure 2.2.
   e. Check that if T is a linear fractional transformation carrying A to 0, B to ∞, P to P', and Q to Q', then $d(P, Q) = d(P', Q')$.

13. a. Let O be any point not lying on a circle $\mathcal{C}$ and let P and Q be points on the circle $\mathcal{C}$ so that O, P, and Q are collinear. Let T be the point on $\mathcal{C}$ so that $\overline{OT}$ is tangent to $\mathcal{C}$. Prove that $(OP)(OQ) = (OT)^2$.
   b. Define *inversion* in the circle of radius R centered at O by sending a point P to the point P' on the ray OP with $(OP)(OP') = R^2$. Show that an inversion in a circle centered at the origin maps a circle $\mathcal{C}$ centered on the u-axis and *not* passing through O to another circle $\mathcal{C}'$ centered on the u-axis. (Hint: For any $P \in \mathcal{C}$, let Q be the other point on $\mathcal{C}$ collinear with O and P, and let Q' be the image of Q under inversion. Use the result of part a to show that OP/OQ' is constant. If C is the center of $\mathcal{C}$, let C' be the point on the u-axis so that $\overline{C'Q'} \| \overline{CP}$. Show that Q' traces out a circle $\mathcal{C}'$ centered at C'.)
   c. Show that inversion in the circle of radius R centered at O maps vertical lines to circles centered on the u-axis and passing through O and vice-versa.

14. a. Prove that every (orientation-preserving) isometry of $\mathbb{H}$ can be written as the composition of linear fractional transformations of the form

$$T_1(z) = z + b \quad \text{for some } b \in \mathbb{R}, \quad T_2(z) = -\frac{1}{z}, \quad \text{and} \quad T_3(z) = cz \quad \text{for some } c > 0.$$

   (Hint: It's probably easiest to work with matrices. Show that you have matrices of the form $\begin{bmatrix} a & 0 \\ 0 & 1/a \end{bmatrix}$, $\begin{bmatrix} 1 & b \\ 0 & 1 \end{bmatrix}$, $\begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$, and therefore $\begin{bmatrix} 1 & 0 \\ b & 1 \end{bmatrix}$, and that any matrix of determinant 1 can be obtained as a product of such.)
   b. Prove that T_2 maps circles centered on the u-axis and vertical lines to circles centered on the u-axis and vertical lines (not necessarily respectively). Either do this algebraically or use Exercise 13.
   c. Use the results of parts a and b to prove that isometries of $\mathbb{H}$ map geodesics to geodesics.

15. We say a linear fractional transformation $T = T_{a,b,c,d}$ is elliptic if it has one fixed point, parabolic if it has one fixed point at infinity, and hyperbolic if it has two fixed points at infinity.
 a. Show that T is elliptic if $|a+d| < 2$, parabolic if $|a+d| = 2$, and hyperbolic if $|a+d| > 2$.
 b. Describe the three types of isometries geometrically. (Hint: In particular, what is the relation between horocycles and parabolic linear fractional transformations?)

16. Suppose $\triangle ABC$ is a hyperbolic right triangle with "hypotenuse" c. Use Figure 2.9 to prove the following:
$$\sin\angle A = \frac{\sinh a}{\sinh c}, \qquad \cos\angle A = \frac{\tanh b}{\tanh c}, \qquad \cosh c = \cosh a \cosh b.$$
(The last is the hyperbolic Pythagorean Theorem.) (Hint: Start by showing, for example, that $\cosh b =$

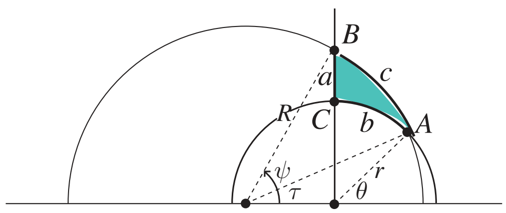


FIGURE 2.9

$\csc\theta$, $\cosh c = (1-\cos\psi\cos\tau)/(\sin\psi\sin\tau)$, and $\cos\tau - \cos\psi = \sin\tau\cot\theta$. You will need two equations trigonometrically relating R and r.)

17. Given a point P on a surface M, we define the geodesic circle of radius R centered at P to be the locus of points whose (geodesic) distance from P is R. Let $C(R)$ denote its circumference.
 a. Show that on the unit sphere
$$\lim_{R\to 0^+} \frac{2\pi R - C(R)}{\pi R^3} = \frac{1}{3}.$$
 b. Show that the geodesic curvature κ_g of a spherical geodesic circle of radius R is $\cot R \approx \frac{1}{R}(1 - \frac{R^2}{3} + \dots)$.

The *Poincaré disk* is defined to be the "abstract surface" $\mathbb{D} = \{(u,v) : u^2+v^2 < 1\}$ with the first fundamental form given, in polar coordinates (r,θ), by $E = \dfrac{4}{(1-r^2)^2}$, $F = 0$, $G = \dfrac{4r^2}{(1-r^2)^2}$. This is called the *hyperbolic metric* on $\mathbb{D}$.
 c. Check that in $\mathbb{D}$ the geodesics through the origin are Euclidean line segments; conclude that the Euclidean circle of radius r centered at the origin is a hyperbolic circle of radius $R = \ln\left(\dfrac{1+r}{1-r}\right)$, and so $r = \tanh\dfrac{R}{2}$. (Remark: Other geodesics are semicircles orthogonal to the unit circle, the "virtual boundary" of $\mathbb{D}$. This should make sense since there is a linear fractional transformation mapping $\mathbb{H}$ to $\mathbb{D}$; by Exercise 12c, it will map semicircles orthogonal to the u-axis to semicircles orthogonal to the unit circle.)
 d. Check that the circumference of the hyperbolic circle is $2\pi\sinh R \approx 2\pi(R + \frac{R^3}{6} + \dots)$, and so
$$\lim_{R\to 0^+} \frac{2\pi R - C(R)}{\pi R^3} = -\frac{1}{3}.$$

e. Compute (using a double integral) that the area of a disk of hyperbolic radius R is $4\pi\sinh^2\frac{R}{2}\approx \pi R^2(1+\frac{R^2}{12}+\dots)$. Use the Gauss-Bonnet Theorem to deduce that the geodesic curvature κ_g of the hyperbolic circle of radius R is $\coth R\approx \frac{1}{R}(1+\frac{R^2}{3}+\dots)$.

18. Here we give another model for hyperbolic geometry, called the Klein-Beltrami model. Consider the following parametrization of the hyperbolic disk: Start with the open unit disk, $\left\{x_1^2+x_2^2<1,\ x_3=0\right\}$, vertically project to the southern hemisphere of the unit sphere, and then stereographically project (from the north pole) back to the unit disk.

a. Show that this mapping is given in polar coordinates by

$$\mathbf{x}(R,\theta)=(r,\theta)=\left(\frac{R}{1+\sqrt{1-R^2}},\theta\right).$$

Compute that the first fundamental form of the Poincaré metric on $\mathbb{D}$ (see Exercise 17) is given in (R,θ) coordinates by $\tilde{E}=\dfrac{1}{(1-R^2)^2}$, $\tilde{F}=0$, $\tilde{G}=\dfrac{R^2}{1-R^2}$. (Hint: Compute carefully and economically!)

b. Compute the distance from $(0,0)$ to $(a,0)$; compare with the formula for distance in the Poincaré model.

c. Changing now to Euclidean coordinates (u,v), show that

$$\hat{E}=\frac{1-v^2}{(1-u^2-v^2)^2},\quad \hat{F}=\frac{uv}{(1-u^2-v^2)^2},\quad \hat{G}=\frac{1-u^2}{(1-u^2-v^2)^2},$$

whence you derive

$$\begin{aligned}
\Gamma_{uu}^{u}&=\frac{2u}{1-u^2-v^2}, & \Gamma_{uu}^{v}&=0,\\
\Gamma_{uv}^{u}&=\frac{v}{1-u^2-v^2}, & \Gamma_{uv}^{v}&=\frac{u}{1-u^2-v^2},\\
\Gamma_{vv}^{u}&=0, & \Gamma_{vv}^{v}&=\frac{2v}{1-u^2-v^2}.
\end{aligned}$$

d. Use part b to show that the geodesics of the disk using the first fundamental form $\hat{\mathrm{I}}$ are chords of the circle $u^2+v^2=1$. (Hint: Show (by using the chain rule) that the equations for a geodesic give $d^2v/du^2=0$.) Discuss the advantages and disadvantages of this model (compared to Poincaré's).

e. Check your answer in part c by proving (geometrically?) that chords of the circle map by $\mathbf{x}$ to geodesics in the hyperbolic disk. (See Exercise 2.1.8.)

3. Surface Theory with Differential Forms

We've seen that it can be quite awkward to work with coordinates to study surfaces. (For example, the Codazzi and Gauss Equations in Section 3 of Chapter 2 are far from beautiful.) For those who've learned about differential forms, we can given a quick and elegant treatment that is conceptually quite clean.

We start (much like the situation with curves) with a *moving frame* $\mathbf{e}_1,\mathbf{e}_2,\mathbf{e}_3$ on (an open subset of) our (oriented) surface M. Here $\mathbf{e}_i$ are vector fields defined on M with the properties that

(i) $\{\mathbf{e}_1,\mathbf{e}_2,\mathbf{e}_3\}$ gives an orthonormal basis for $\mathbb{R}^3$ at each point (so the matrix with those respective column vectors is an orthogonal matrix);

(ii) $\{\mathbf{e}_1, \mathbf{e}_2\}$ is a basis for the tangent space of M and $\mathbf{e}_3 = \mathbf{n}$.

How do we know such a moving frame exists? If $\mathbf{x}: U \to M$ is a parametrized surface, we can start with our usual vectors $\mathbf{x}_u$, $\mathbf{x}_v$ and apply the Gram-Schmidt process to obtain an orthonormal basis. Or, if M is a surface containing no umbilic points, then we can choose $\mathbf{e}_1$ and $\mathbf{e}_2$ to be unit vectors pointing in the principal directions; this approach was tacit in many of our proofs earlier.

If $\mathbf{x}: M \to \mathbb{R}^3$ is the inclusion map (which we may choose, in a computational setting, to consider as the parametrization mapping $U \to \mathbb{R}^3$), then we define 1-forms ω_1, ω_2 on M by

$$d\mathbf{x} = \omega_1\mathbf{e}_1 + \omega_2\mathbf{e}_2;$$

i.e., for any $\mathbf{V} \in T_P M$, we have $\mathbf{V} = \omega_1(\mathbf{V})\mathbf{e}_1 + \omega_2(\mathbf{V})\mathbf{e}_2$, so $\omega_\alpha(\mathbf{V}) = \mathrm{I}(\mathbf{V}, \mathbf{e}_\alpha)$ for $\alpha = 1, 2$. So far, ω_1 and ω_2 keep track of how our point moves around on M. Next we want to see how the frame itself twists, so we define 1-forms ω_{ij}, $i, j = 1, 2, 3$, by

$$d\mathbf{e}_i = \sum_{j=1}^{3} \omega_{ij}\mathbf{e}_j .$$

Note that since $\mathbf{e}_i \cdot \mathbf{e}_j = \text{const}$ for any $i, j = 1, 2, 3$, we have

$$\begin{aligned} 0 = d(\mathbf{e}_i \cdot \mathbf{e}_j) &= d\mathbf{e}_i \cdot \mathbf{e}_j + \mathbf{e}_i \cdot d\mathbf{e}_j = \Big(\sum_{k=1}^{3} \omega_{ik}\mathbf{e}_k\Big)\cdot \mathbf{e}_j + \Big(\sum_{k=1}^{3} \omega_{jk}\mathbf{e}_k\Big)\cdot \mathbf{e}_i \\ &= \omega_{ij} + \omega_{ji}, \end{aligned}$$

so $\omega_{ji} = -\omega_{ij}$ for all $i, j = 1, 2, 3$. (In particular, since $\mathbf{e}_i$ is always a unit vector, $\omega_{ii} = 0$ for all i.) If $\mathbf{V} \in T_P M$, $\omega_{ij}(\mathbf{V})$ tells us how fast $\mathbf{e}_i$ is twisting towards $\mathbf{e}_j$ at P as we move with velocity $\mathbf{V}$.

Note, in particular, that the shape operator is embodied in the equation

$$d\mathbf{e}_3 = \omega_{31}\mathbf{e}_1 + \omega_{32}\mathbf{e}_2 = -\big(\omega_{13}\mathbf{e}_1 + \omega_{23}\mathbf{e}_2\big).$$

Then for any $\mathbf{V} \in T_P M$ we have $\omega_{13}(\mathbf{V}) = \mathrm{II}(\mathbf{V}, \mathbf{e}_1)$ and $\omega_{23}(\mathbf{V}) = \mathrm{II}(\mathbf{V}, \mathbf{e}_2)$. Indeed, when we write

$$\begin{aligned} \omega_{13} &= h_{11}\omega_1 + h_{12}\omega_2 \\ \omega_{23} &= h_{21}\omega_1 + h_{22}\omega_2 \end{aligned}$$

for appropriate coefficient functions $h_{\alpha\beta}$, we see that the matrix of the shape operator S_P with respect to the basis $\{\mathbf{e}_1, \mathbf{e}_2\}$ for $T_P M$ is nothing but $\big[h_{\alpha\beta}\big]$.

Most of our results will come from the following

Theorem 3.1 (Structure Equations)**.**

$$d\omega_1 = \omega_{12} \wedge \omega_2 \quad \textit{and} \quad d\omega_2 = \omega_1 \wedge \omega_{12}, \qquad \textit{and}$$

$$d\omega_{ij} = \sum_{k=1}^{3} \omega_{ik} \wedge \omega_{kj} \quad \textit{for all } i, j = 1, 2, 3.$$

Proof. From the properties of the exterior derivative, we have

$$\mathbf{0} = d(d\mathbf{x}) = d\omega_1\mathbf{e}_1 + d\omega_2\mathbf{e}_2 - \omega_1 \wedge \Big(\sum_{j=1}^{3} \omega_{1j}\mathbf{e}_j\Big) - \omega_2 \wedge \Big(\sum_{j=1}^{3} \omega_{2j}\mathbf{e}_j\Big)$$

$$= \big(d\omega_1 - \omega_2 \wedge \omega_{21}\big)\mathbf{e}_1 + \big(d\omega_2 - \omega_1 \wedge \omega_{12}\big)\mathbf{e}_2 - \big(\omega_1 \wedge \omega_{13} + \omega_2 \wedge \omega_{23}\big)\mathbf{e}_3,$$

so from the fact that $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$ is a basis for $\mathbb{R}^3$ we infer that

$$d\omega_1 = \omega_2 \wedge \omega_{21} = -\omega_2 \wedge \omega_{12} = \omega_{12} \wedge \omega_2 \qquad \text{and} \qquad d\omega_2 = \omega_1 \wedge \omega_{12}.$$

Similarly, we obtain

$$\begin{aligned}\mathbf{0} = d(d\mathbf{e}_i) &= d\Big(\sum_{k=1}^{3} \omega_{ik}\mathbf{e}_k\Big) = \sum_{k=1}^{3} \Big(d\omega_{ik}\mathbf{e}_k - \omega_{ik} \wedge \sum_{j=1}^{3} \omega_{kj}\mathbf{e}_j\Big) \\ &= \sum_{j=1}^{3} d\omega_{ij}\mathbf{e}_j - \sum_{j=1}^{3}\Big(\sum_{k=1}^{3} \omega_{ik} \wedge \omega_{kj}\Big)\mathbf{e}_j = \sum_{j=1}^{3}\Big(d\omega_{ij} - \sum_{k=1}^{3} \omega_{ik} \wedge \omega_{kj}\Big)\mathbf{e}_j,\end{aligned}$$

so $d\omega_{ij} - \sum_{k=1}^{3} \omega_{ik} \wedge \omega_{kj} = 0$ for all i, j. $\square$

We also have the following additional consequence of the proof:

Proposition 3.2. *The shape operator is symmetric, i.e.,* $h_{12} = h_{21}$.

Proof. From the $\mathbf{e}_3$ component of the equation $d(d\mathbf{x}) = \mathbf{0}$ in the proof of Theorem 3.1 we have

$$0 = \omega_1 \wedge \omega_{13} + \omega_2 \wedge \omega_{23} = \omega_1 \wedge (h_{11}\omega_1 + h_{12}\omega_2) + \omega_2 \wedge (h_{21}\omega_1 + h_{22}\omega_2) = (h_{12} - h_{21})\omega_1 \wedge \omega_2,$$

so $h_{12} - h_{21} = 0$. $\square$

Recall that $\mathbf{V}$ is a principal direction if $d\mathbf{e}_3(\mathbf{V})$ is a scalar multiple of $\mathbf{V}$. So $\mathbf{e}_1$ and $\mathbf{e}_2$ are principal directions if and only if $h_{12} = 0$ and we have $\omega_{13} = k_1\omega_1$ and $\omega_{23} = k_2\omega_2$, where k_1 and k_2 are, as usual, the principal curvatures.

It is important to understand how our battery of forms changes if we change our moving frame by rotating $\mathbf{e}_1, \mathbf{e}_2$ through some angle θ (which may be a function).

Lemma 3.3. *Suppose* $\bar{\mathbf{e}}_1 = \cos\theta\mathbf{e}_1 + \sin\theta\mathbf{e}_2$ *and* $\bar{\mathbf{e}}_2 = -\sin\theta\mathbf{e}_1 + \cos\theta\mathbf{e}_2$ *for some function* θ*. Then we have*

$$\begin{aligned}\overline{\omega}_1 &= \cos\theta\omega_1 + \sin\theta\omega_2 \\ \overline{\omega}_2 &= -\sin\theta\omega_1 + \cos\theta\omega_2 \\ \overline{\omega}_{12} &= \omega_{12} + d\theta \\ \overline{\omega}_{13} &= \cos\theta\omega_{13} + \sin\theta\omega_{23} \\ \overline{\omega}_{23} &= -\sin\theta\omega_{13} + \cos\theta\omega_{23}\end{aligned}$$

Note, in particular, that $\overline{\omega}_1 \wedge \overline{\omega}_2 = \omega_1 \wedge \omega_2$ *and* $\overline{\omega}_{13} \wedge \overline{\omega}_{23} = \omega_{13} \wedge \omega_{23}$.

Proof. We leave this to the reader in Exercise 1. $\square$

It is often convenient when we study curves in surfaces (as we did in Sections 3 and 4 of Chapter 2) to use the Darboux frame, a moving frame for the surface adapted so that $\mathbf{e}_1$ is tangent to the curve. (See Exercise 3.) For example, $\boldsymbol{\alpha}$ is a geodesic if and only if in terms of the Darboux frame we have $\omega_{12} = 0$ as a 1-form on $\boldsymbol{\alpha}$.

Let's now examine the structure equations more carefully.

Gauss equation:	$d\omega_{12} = -\omega_{13} \wedge \omega_{23}$
Codazzi equations:	$d\omega_{13} = \omega_{12} \wedge \omega_{23}$
	$d\omega_{23} = -\omega_{12} \wedge \omega_{13}$

Example 1. To illustrate the power of the moving frame approach, we reprove Proposition 3.4 of Chapter 2: Suppose $K = 0$ and M has no planar points. Then we claim that M is ruled and the tangent plane of M is constant along the rulings. We work in a principal moving frame with $k_1 = 0$, so $\omega_{13} = 0$. Therefore, by the first Codazzi equation, $d\omega_{13} = 0 = \omega_{12} \wedge \omega_{23} = \omega_{12} \wedge k_2\omega_2$. Since $k_2 \neq 0$, we must have $\omega_{12} \wedge \omega_2 = 0$, and so $\omega_{12} = f\omega_2$ for some function f. Therefore, $\omega_{12}(\mathbf{e}_1) = 0$, and so $d\mathbf{e}_1(\mathbf{e}_1) = \omega_{12}(\mathbf{e}_1)\mathbf{e}_2 + \omega_{13}(\mathbf{e}_1)\mathbf{e}_3 = \mathbf{0}$. It follows that $\mathbf{e}_1$ stays constant as we move in the $\mathbf{e}_1$ direction, so following the $\mathbf{e}_1$ direction gives us a line. Moreover, $d\mathbf{e}_3(\mathbf{e}_1) = \mathbf{0}$ (since $k_1 = 0$), so the tangent plane to M is constant along that line. ∇

The Gauss equation is particularly interesting. First, note that

$$\omega_{13} \wedge \omega_{23} = (h_{11}\omega_1 + h_{12}\omega_2) \wedge (h_{12}\omega_1 + h_{22}\omega_2) = (h_{11}h_{22} - h_{12}^2)\omega_1 \wedge \omega_2 = K\,dA,$$

where $K = \det\left[h_{\alpha\beta}\right] = \det S_P$ is the Gaussian curvature. So, the Gauss equation really reads:

$$(\star) \qquad d\omega_{12} = -K\,dA.$$

(How elegant!) Note, moreover, that, by Lemma 3.3, for any two moving frames $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$ and $\bar{\mathbf{e}}_1, \bar{\mathbf{e}}_2, \bar{\mathbf{e}}_3$, we have $d\bar{\omega}_{12} = d\omega_{12}$ (which is good, since the right-hand side of $(\star)$ doesn't depend on the frame field). Next, we observe that, because of the first equations in Theorem 3.1, ω_{12} can be computed just from knowing ω_1 and ω_2, hence depends just on the first fundamental form of the surface. (If we write $\omega_{12} = P\omega_1 + Q\omega_2$, then the first equation determines P and the second determines Q.) We therefore arrive at a new proof of Gauss's Theorema Egregium, Theorem 3.1 of Chapter 2.

The 1-form ω_{12} is called the *connection form* and measures the tangential twist of $\mathbf{e}_1$. Just as we saw in Section 1, then, $\nabla_{\mathbf{V}}\mathbf{e}_1$ is the tangential component of $D_{\mathbf{V}}\mathbf{e}_1 = d\mathbf{e}_1(\mathbf{V}) = \omega_{12}(\mathbf{V})\mathbf{e}_2 + \omega_{13}(\mathbf{V})\mathbf{e}_3$, which is, of course, $\omega_{12}(\mathbf{V})\mathbf{e}_2$. In particular, $\omega_{12}(\mathbf{e}_1)$ recovers the geodesic curvature of the $\mathbf{e}_1$-curve.

Example 2. Let's go back to our usual parametrization of the unit sphere,

$$\mathbf{x}(u, v) = (\sin u \cos v, \sin u \sin v, \cos u), \qquad 0 < u < \pi, \quad 0 < v < 2\pi.$$

Then we have

$$d\mathbf{x} = \mathbf{x}_u\,du + \mathbf{x}_v\,dv = \underbrace{(\cos u \cos v, \cos u \sin v, -\sin u)}_{\mathbf{e}_1}\,du + \underbrace{(-\sin v, \cos v, 0)}_{\mathbf{e}_2}(\sin u\,dv).$$

Note that $\mathbf{e}_1 = \mathbf{x}_u$ and $\mathbf{e}_2 = \mathbf{x}_v/\sqrt{G}$, as we might expect. So this gives us

$$\omega_1 = du \qquad \text{and} \qquad \omega_2 = \sin u\,dv.$$

Next, $d\omega_1 = 0$ and $d\omega_2 = \cos u\,du \wedge dv = du \wedge (\cos u\,dv)$, so we see from the first structure equations that $\omega_{12} = \cos u\,dv$. It is hard to miss the similarity this bears to the discussion of ϕ_{12} and Example 1 in Section 1. Now we have $d\omega_{12} = -\sin u\,du \wedge dv = -\omega_1 \wedge \omega_2$, so, indeed, the sphere has Gaussian curvature $K = 1$.

Let's now compute the geodesic curvature κ_g of the latitude circle $u = u_0$. We obtain a Darboux frame by taking $\bar{\mathbf{e}}_1 = \mathbf{e}_2$ and $\bar{\mathbf{e}}_2 = -\mathbf{e}_1$. Now, $\overline{\omega}_{12} = -\omega_{21} = \omega_{12}$ (this also follows from Lemma 3.3). Then $\kappa_g = \overline{\omega}_{12}(\bar{\mathbf{e}}_1) = \omega_{12}(\mathbf{e}_2)$. Now note that $\omega_{12} = \cos u\, dv = \cot u\, \omega_2$, so $\kappa_g = \cot u_0$. ∇

To illustrate the power of the differential forms approach, we give a proof of the following result (see Exercise 2.3.16).

Proposition 3.4. *Suppose M has no umbilic points and k_1 is constant. Then M is (a subset of) a tube of radius $r = 1/|k_1|$ about a regular curve $\boldsymbol{\alpha}$.*

Proof. Choose a principal moving frame $\mathbf{e}_1, \mathbf{e}_2$. We have $\omega_{13} = k_1\omega_1$ and $\omega_{23} = k_2\omega_2$. Differentiating the first, since k_1 is constant, we get $\omega_{12}\wedge\omega_{23} = k_1\omega_{12}\wedge\omega_2$, so $\omega_{12}\wedge(k_2-k_1)\omega_2 = 0$. Since $k_2-k_1 \neq 0$, we infer that $\omega_{12} = \lambda\omega_2$ for some scalar function λ. Now let $\bar{\mathbf{e}}_1 = \mathbf{e}_1$, $\bar{\mathbf{e}}_2$, $\bar{\mathbf{e}}_3$ be the Frenet frame of the $\mathbf{e}_1$-curve and apply Exercise 3. Since both $\omega_{12} = 0$ and $\omega_{13} \neq 0$ when restricted to (pulled back to) an $\mathbf{e}_1$-curve, we infer that $\cos\theta = 0$ and $\theta = \pm\pi/2$ all along the curve. Then $\overline{\omega}_{23} = \tau\omega_1 = 0$ on the $\mathbf{e}_1$-curve, so $\tau = 0$ and the curve is planar. But then $\kappa\omega_1 = \overline{\omega}_{12} = \pm\omega_{13} = \pm k_1\omega_1$, so $\kappa = |k_1|$ is constant and the $\mathbf{e}_1$-curves are circles.

Now consider $\boldsymbol{\alpha} = \mathbf{x} + \dfrac{1}{k_1}\mathbf{e}_3$. Then
$$d\boldsymbol{\alpha} = d\mathbf{x} + \frac{1}{k_1}d\mathbf{e}_3 = \omega_1\mathbf{e}_1 + \omega_2\mathbf{e}_2 + \frac{1}{k_1}(-k_1\omega_1\mathbf{e}_1 - k_2\omega_2\mathbf{e}_2) = \Big(1 - \frac{k_2}{k_1}\Big)\omega_2\mathbf{e}_2,$$
so $\boldsymbol{\alpha}$ is constant along the $\mathbf{e}_1$-curves and $d\boldsymbol{\alpha} \neq \mathbf{0}$, which means that the image of $\boldsymbol{\alpha}$ is a regular curve, the center of the tube, as desired. $\square$

From the Gauss equation and Stokes's Theorem, the Gauss-Bonnet formula follows immediately for an oriented surface M with (piecewise smooth) boundary ∂M *on which we can globally define a moving frame*. That is, we can reprove the Local Gauss-Bonnet formula, Theorem 1.6, quite effortlessly.

Proof. We start with an arbitrary moving frame $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$ and take a Darboux frame $\bar{\mathbf{e}}_1, \bar{\mathbf{e}}_2, \bar{\mathbf{e}}_3$ *along* ∂M. We write $\bar{\mathbf{e}}_1 = \cos\theta\mathbf{e}_1 + \sin\theta\mathbf{e}_2$ and $\bar{\mathbf{e}}_2 = -\sin\theta\mathbf{e}_1 + \cos\theta\mathbf{e}_2$ (where θ is smoothly chosen along the smooth pieces of ∂M and the exterior angle ϵ_j at P_j gives the "jump" of θ as we cross P_j). Then, by Stokes's Theorem and Lemma 3.3, we have
$$\iint_M K\,dA = -\iint_M d\omega_{12} = -\int_{\partial M}\omega_{12} = -\int_{\partial M}\big(\overline{\omega}_{12} - d\theta\big) = -\int_{\partial M}\kappa_g\,ds + (2\pi - \sum\epsilon_j).$$
(See Exercise 2.) $\square$

EXERCISES 3.3

1. Prove Lemma 3.3.

2. Let $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$ be the Darboux frame along a curve $\boldsymbol{\alpha}$. Show that as a 1-form on $\boldsymbol{\alpha}$, $\omega_{12} = \kappa_g\omega_1$. Use this result to reprove the result of Exercise 3.1.7.

3. Suppose $\boldsymbol{\alpha}$ is a curve lying in the surface M. Let $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$ be the Darboux frame along $\boldsymbol{\alpha}$ (i.e., a moving frame for the surface with $\mathbf{e}_1$ tangent to $\boldsymbol{\alpha}$), and let $\bar{\mathbf{e}}_1 = \mathbf{e}_1, \bar{\mathbf{e}}_2, \bar{\mathbf{e}}_3$ be the Frenet frame. Then,

by analogy with Lemma 3.3, $\bar{\mathbf{e}}_2, \bar{\mathbf{e}}_3$ are obtained from $\mathbf{e}_2, \mathbf{e}_3$ by rotating through some angle θ. Show that, as 1-forms on $\boldsymbol{\alpha}$, we have:

$$\begin{aligned} \overline{\omega}_{12} &= \kappa\omega_1 = \cos\theta\omega_{12} + \sin\theta\omega_{13} \\ \overline{\omega}_{13} &= \ \ 0 \ \ \ = -\sin\theta\omega_{12} + \cos\theta\omega_{13} \\ \overline{\omega}_{23} &= \tau\omega_1 = \omega_{23} + d\theta. \end{aligned}$$

*4. Use Exercise 3 to prove Meusnier's Theorem (Proposition 2.5 of Chapter 2).

5. Use Exercise 3 to prove that if $C \subset M$ is a line of curvature and the osculating plane of C makes a constant angle with the tangent plane of M, then C is planar.

6. Use moving frames to redo Exercise 2.2.14. (Hint: Use the Codazzi equations to show that $dk \wedge \omega_1 = dk \wedge \omega_2 = 0$.)

7. Use moving frames to redo Exercise 2.2.15.

*8. Use moving frames to compute the Gaussian curvature of the torus, parametrized as in Example 1(c) of Chapter 2.

9. The vectors $\mathbf{e}_1 = v(1,0)$ and $\mathbf{e}_2 = v(0,1)$ give a moving frame at $(u,v) \in \mathbb{H}$. Set $\omega_1 = du/v$ and $\omega_2 = dv/v$.
   a. Check that for any $\mathbf{V} \in T_{(u,v)}\mathbb{H}$, $\omega_1(\mathbf{V}) = \mathrm{I}(\mathbf{V}, \mathbf{e}_1)$ and $\omega_2(\mathbf{V}) = \mathrm{I}(\mathbf{V}, \mathbf{e}_2)$.
   b. Compute ω_{12} and $d\omega_{12}$ and verify that $K = -1$.

10. Use moving frames to redo
    a. Exercise 3.1.8
    b. Exercise 3.1.9

11. a. Use moving frames to reprove the result of Exercise 2.3.14.
    b. Use moving frames to reprove the result of Exercise 2.4.13. That is, prove that if there are two families of geodesics in M that are everywhere orthogonal, then M is flat.
    c. Suppose there are two families of geodesics in M making a constant angle θ. Prove or disprove: M is flat.

12. Use moving frames to redo Exercise 2.3.17. (See Proposition 3.4.)

13. Recall that locally any 1-form ϕ with $d\phi = 0$ can be written in the form $\phi = df$ for some function f.
    a. Prove that if a surface M is flat, then locally we can find a moving frame $\mathbf{e}_1, \mathbf{e}_2$ on M so that $\omega_{12} = 0$. (Hint: Start with an arbitrary moving frame.)
    b. Deduce that if M is flat, locally we can find a parametrization $\mathbf{x}$ of M with $E = G = 1$ and $F = 0$. (That is, locally M is isometric to a plane.)

14. *(The Bäcklund transform)* Suppose M and $\overline{M}$ are two surfaces in $\mathbb{R}^3$ and $f\colon M \to \overline{M}$ is a smooth bijective function with the properties that
    (i) the line from P to $f(P)$ is tangent to M at P and tangent to $\overline{M}$ at $f(P)$;
    (ii) the distance between P and $f(P)$ is a constant r, independent of P;

(iii) the angle between $\mathbf{n}(P)$ and $\overline{\mathbf{n}}(f(P))$ is a constant θ, independent of P.

Prove that both M and $\overline{M}$ have constant curvature $K = -(\sin^2\theta)/r^2$. (Hints: Write $\overline{P} = f(P)$, and let $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$ (resp. $\overline{\mathbf{e}}_1, \overline{\mathbf{e}}_2, \overline{\mathbf{e}}_3$) be moving frames at P (resp. $\overline{P}$) with $\overline{\mathbf{e}}_1 = \mathbf{e}_1$ in the direction of $\overrightarrow{P\overline{P}}$. Let $\mathbf{x}$ and $\overline{\mathbf{x}} = f\circ\mathbf{x}$ be local parametrizations. How else are $\mathbf{x}$ and $\overline{\mathbf{x}}$ related?)

4. Calculus of Variations and Surfaces of Constant Mean Curvature

Every student of calculus is familiar with the necessary condition for a differentiable function $f:\mathbb{R}^n \to \mathbb{R}$ to have a local extreme point (minimum or maximum) at P: We must have $\nabla f(P) = \mathbf{0}$. Phrased slightly differently, for every vector $\mathbf{V}$, the directional derivative

$$D_{\mathbf{V}} f(P) = \lim_{\varepsilon\to 0} \frac{f(P+\varepsilon\mathbf{V}) - f(P)}{\varepsilon}$$

should vanish. Moreover, if we are given a constraint set $M = \{\mathbf{x} \in \mathbb{R}^n : g_1(\mathbf{x}) = 0, g_2(\mathbf{x}) = 0, \dots, g_k(\mathbf{x}) = 0\}$, the method of Lagrange multipliers tells us that at a constrained extreme point P we must have

$$\nabla f(P) = \sum_{i=1}^{k} \lambda_i \nabla g_i(P)$$

for some scalars $\lambda_1, \dots, \lambda_k$. (There is also a nondegeneracy hypothesis here that $\nabla g_1(P), \dots, \nabla g_k(P)$ be linearly independent.)

Suppose we are given a regular parametrized surface $\mathbf{x}: U \to \mathbb{R}^3$ and want to find—without the benefit of the analysis of Section 4 of Chapter 2—a geodesic from $P = \mathbf{x}(u_0, v_0)$ to $Q = \mathbf{x}(u_1, v_1)$. Among all paths $\boldsymbol{\alpha}: [0,1] \to M$ with $\boldsymbol{\alpha}(0) = P$ and $\boldsymbol{\alpha}(1) = Q$, we wish to find the shortest. That is, we want to choose the path $\boldsymbol{\alpha}(t) = \mathbf{x}(u(t), v(t))$ so as to *minimize* the integral

$$\int_0^1 \|\boldsymbol{\alpha}'(t)\|dt = \int_0^1 \sqrt{E(u(t),v(t))(u'(t))^2 + 2F(u(t),v(t))u'(t)v'(t) + G(u(t),v(t))(v'(t))^2}\,dt$$

subject to the constraints that $(u(0), v(0)) = (u_0, v_0)$ and $(u(1), v(1)) = (u_1, v_1)$, as indicated in Figure 4.1. Now we're doing a minimization problem in the space of all ($\mathcal{C}^1$) curves $(u(t), v(t))$ with $(u(0), v(0)) =$

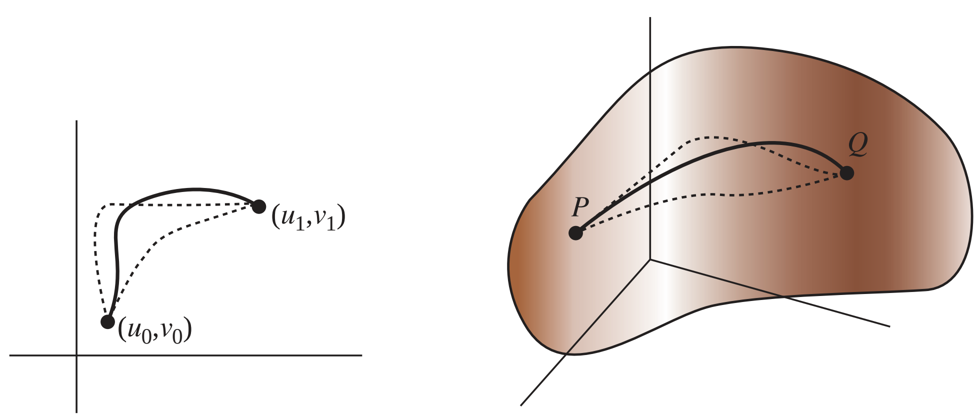


FIGURE 4.1

(u_0, v_0) and $(u(1), v(1)) = (u_1, v_1)$. Even though we're now working in an infinite-dimensional setting, we should not panic. In classical terminology, we have a *functional* F defined on the space X of $\mathcal{C}^1$ curves $\mathbf{u}\colon [0,1] \to \mathbb{R}^3$, i.e.,

$$(*) \qquad F(\mathbf{u}) = \int_0^1 f(t, \mathbf{u}(t), \mathbf{u}'(t))\,dt.$$

For example, in the case of the arclength problem, we have

$$f\big(t, (u(t), v(t)), (u'(t), v'(t))\big) = \\ \sqrt{E(u(t), v(t))(u'(t))^2 + 2F(u(t), v(t))u'(t)v'(t) + G(u(t), v(t))(v'(t))^2}.$$

To say that a particular curve $\mathbf{u}^*$ is a local extreme point (with fixed endpoints) of the functional F given in $(*)$ is to say that for any *variation* $\boldsymbol{\xi}\colon [0,1] \to \mathbb{R}^2$ with $\boldsymbol{\xi}(0) = \boldsymbol{\xi}(1) = \mathbf{0}$, the directional derivative

$$D_{\boldsymbol{\xi}} F(\mathbf{u}^*) = \lim_{\varepsilon\to 0} \frac{F(\mathbf{u}^* + \varepsilon\boldsymbol{\xi}) - F(\mathbf{u}^*)}{\varepsilon} = \frac{d}{d\varepsilon}\bigg|_{\varepsilon=0} F(\mathbf{u}^* + \varepsilon\boldsymbol{\xi})$$

should vanish. This leads us to the

Theorem 4.1 (Euler-Lagrange Equations)**.** *If* $\mathbf{u}^*$ *is a local extreme point of the functional* F *given above in* $(*)$*, then at* $\mathbf{u}^*$ *we have*

$$\frac{\partial f}{\partial \mathbf{u}} = \frac{d}{dt}\left(\frac{\partial f}{\partial \mathbf{u}'}\right),$$

evaluating these both at $(t, \mathbf{u}^*(t), \mathbf{u}^{*\prime}(t))$*, for all* $0 \le t \le 1$.

Proof. Let $\boldsymbol{\xi}\colon [0,1] \to \mathbb{R}^2$ be a $\mathcal{C}^1$ curve with $\boldsymbol{\xi}(0) = \boldsymbol{\xi}(1) = \mathbf{0}$. Then, using the fact that we can pull the derivative under the integral sign (see Exercise 1) and then the chain rule, we have

$$\begin{aligned}
\frac{d}{d\varepsilon}\bigg|_{\varepsilon=0} F(\mathbf{u}^* + \varepsilon\boldsymbol{\xi}) &= \frac{d}{d\varepsilon}\bigg|_{\varepsilon=0} \int_0^1 f(t, \mathbf{u}^*(t) + \varepsilon\boldsymbol{\xi}(t), \mathbf{u}^{*\prime}(t) + \varepsilon\boldsymbol{\xi}'(t))\,dt \\
&= \int_0^1 \frac{\partial}{\partial\varepsilon}\bigg|_{\varepsilon=0} f(t, \mathbf{u}^*(t) + \varepsilon\boldsymbol{\xi}(t), \mathbf{u}^{*\prime}(t) + \varepsilon\boldsymbol{\xi}'(t))\,dt \\
&= \int_0^1 \Big(\frac{\partial f}{\partial \mathbf{u}}(t, \mathbf{u}^*(t), \mathbf{u}^{*\prime}(t)) \cdot \boldsymbol{\xi}(t) + \frac{\partial f}{\partial \mathbf{u}'}(t, \mathbf{u}^*(t), \mathbf{u}^{*\prime}(t)) \cdot \boldsymbol{\xi}'(t)\Big)dt
\end{aligned}$$

and so, integrating by parts, we have

$$\begin{aligned}
&= \int_0^1 \left(\frac{\partial f}{\partial \mathbf{u}} \cdot \boldsymbol{\xi}(t) - \frac{d}{dt}\left(\frac{\partial f}{\partial \mathbf{u}'}\right) \cdot \boldsymbol{\xi}(t)\right)dt + \frac{\partial f}{\partial \mathbf{u}'} \cdot \boldsymbol{\xi}(t)\bigg]_0^1 \\
&= \int_0^1 \left(\frac{\partial f}{\partial \mathbf{u}} - \frac{d}{dt}\left(\frac{\partial f}{\partial \mathbf{u}'}\right)\right) \cdot \boldsymbol{\xi}(t)\,dt.
\end{aligned}$$

Now, applying Exercise 2, since this holds for *all* $\mathcal{C}^1$ $\boldsymbol{\xi}$ with $\boldsymbol{\xi}(0) = \boldsymbol{\xi}(1) = \mathbf{0}$, we infer that

$$\frac{\partial f}{\partial \mathbf{u}} - \frac{d}{dt}\left(\frac{\partial f}{\partial \mathbf{u}'}\right) = \mathbf{0},$$

as desired. $\square$

Of course, the Euler-Lagrange equations really give a system of differential equations:

$$(\clubsuit)\qquad \begin{aligned}\frac{\partial f}{\partial u} &= \frac{d}{dt}\left(\frac{\partial f}{\partial u'}\right)\\ \frac{\partial f}{\partial v} &= \frac{d}{dt}\left(\frac{\partial f}{\partial v'}\right).\end{aligned}$$

Example 1. Recall that for the unit sphere in the usual parametrization we have $E = 1$, $F = 0$, and $G = \sin^2 u$. To find the shortest path from $(u_0, v_0) = (u_0, v_0)$ to the point $(u_1, v_1) = (u_1, v_0)$, we want to minimize the functional

$$F(u,v) = \int_0^1 \sqrt{(u'(t))^2 + \sin^2 u(t)(v'(t))^2}\,dt.$$

Assuming our critical path $\mathbf{u}^*$ is parametrized at constant speed, the equations ($\clubsuit$) give us $v'(t) = \text{const}$ and $u''(t) = \sin u(t)\cos u(t) v'(t)^2$. (Cf. Example 6(b) in Section 4 of Chapter 2.) ∇

We now come to two problems that interest us here: What is the surface of least area with a given boundary curve? And what is the surface of least area containing a given volume? For this we must consider parametrized surfaces and hence functionals defined on functions of two variables. In particular, for functions $\mathbf{x}\colon D \to \mathbb{R}^3$ defined on a given domain $D \subset \mathbb{R}^2$, we consider

$$F(\mathbf{x}) = \iint_D \|\mathbf{x}_u \times \mathbf{x}_v\|\,du\,dv.$$

We seek a function $\mathbf{x}^*$ so that, for all variations $\boldsymbol{\xi}\colon D \to \mathbb{R}^3$ with $\boldsymbol{\xi} = \mathbf{0}$ on ∂D,

$$D_{\boldsymbol{\xi}} F(\mathbf{u}^*) = \lim_{\varepsilon\to 0} \frac{F(\mathbf{u}^* + \varepsilon\boldsymbol{\xi}) - F(\mathbf{u}^*)}{\varepsilon} = \left.\frac{d}{d\varepsilon}\right|_{\varepsilon=0} F(\mathbf{u}^* + \varepsilon\boldsymbol{\xi}) = 0.$$

Now we compute: Recalling that $\dfrac{d}{dt}\|\mathbf{f}(t)\| = \dfrac{\mathbf{f}(t)\cdot\mathbf{f}'(t)}{\|\mathbf{f}(t)\|}$ and setting $\mathbf{x} = \mathbf{x}^* + \varepsilon\boldsymbol{\xi}$, we have

$$\begin{aligned}\left.\frac{d}{d\varepsilon}\right|_{\varepsilon=0} \|\mathbf{x}_u \times \mathbf{x}_v\| &= \frac{1}{\|\mathbf{x}_u^* \times \mathbf{x}_v^*\|}\big((\boldsymbol{\xi}_u \times \mathbf{x}_v^* + \mathbf{x}_u^* \times \boldsymbol{\xi}_v)\cdot(\mathbf{x}_u^* \times \mathbf{x}_v^*)\big)\\ &= (\boldsymbol{\xi}_u \times \mathbf{x}_v^* + \mathbf{x}_u^* \times \boldsymbol{\xi}_v)\cdot\mathbf{n}.\end{aligned}$$

Next we observe that

$$\begin{aligned}(\boldsymbol{\xi}_u \times \mathbf{x}_v^*)\cdot\mathbf{n} &= \big((\boldsymbol{\xi} \times \mathbf{x}_v^*)\cdot\mathbf{n}\big)_u - (\boldsymbol{\xi} \times \mathbf{x}_{uv}^*)\cdot\mathbf{n} - (\boldsymbol{\xi} \times \mathbf{x}_v^*)\cdot\mathbf{n}_u\\ (\mathbf{x}_u^* \times \boldsymbol{\xi}_v)\cdot\mathbf{n} &= \big((\mathbf{x}_u^* \times \boldsymbol{\xi})\cdot\mathbf{n}\big)_v - (\mathbf{x}_{uv}^* \times \boldsymbol{\xi})\cdot\mathbf{n} - (\mathbf{x}_u^* \times \boldsymbol{\xi})\cdot\mathbf{n}_v,\end{aligned}$$

and so, adding these equations, we obtain

$$\begin{aligned}(\boldsymbol{\xi}_u \times \mathbf{x}_v^* + \mathbf{x}_u^* \times \boldsymbol{\xi}_v)\cdot\mathbf{n} &= \big((\boldsymbol{\xi} \times \mathbf{x}_v^*)\cdot\mathbf{n}\big)_u + \big((\mathbf{x}_u^* \times \boldsymbol{\xi})\cdot\mathbf{n}\big)_v - \big((\boldsymbol{\xi} \times \mathbf{x}_v^*)\cdot\mathbf{n}_u + (\mathbf{x}_u^* \times \boldsymbol{\xi})\cdot\mathbf{n}_v\big)\\ &= \big((\boldsymbol{\xi} \times \mathbf{x}_v^*)\cdot\mathbf{n}\big)_u - \big((\boldsymbol{\xi} \times \mathbf{x}_u^*)\cdot\mathbf{n}\big)_v - \big((\boldsymbol{\xi} \times \mathbf{x}_v^*)\cdot\mathbf{n}_u + (\mathbf{x}_u^* \times \boldsymbol{\xi})\cdot\mathbf{n}_v\big)\\ &= \big((\boldsymbol{\xi} \times \mathbf{x}_v^*)\cdot\mathbf{n}\big)_u - \big((\boldsymbol{\xi} \times \mathbf{x}_u^*)\cdot\mathbf{n}\big)_v - \boldsymbol{\xi}\cdot\big(\mathbf{x}_v^* \times \mathbf{n}_u + \mathbf{n}_v \times \mathbf{x}_u^*\big).\end{aligned}$$

At the last step, we've used the identity $(\mathbf{U} \times \mathbf{V})\cdot\mathbf{W} = (\mathbf{W} \times \mathbf{U})\cdot\mathbf{V} = (\mathbf{V} \times \mathbf{W})\cdot\mathbf{U}$. The appropriate way to integrate by parts in the two-dimensional setting is to apply Green's Theorem, Theorem 2.6 of the Appendix, and so we let $P = (\boldsymbol{\xi} \times \mathbf{x}_u^*)\cdot\mathbf{n}$ and $Q = (\boldsymbol{\xi} \times \mathbf{x}_v^*)\cdot\mathbf{n}$ and obtain

$$\iint_D (\boldsymbol{\xi}_u \times \mathbf{x}_v^* + \mathbf{x}_u^* \times \boldsymbol{\xi}_v)\cdot\mathbf{n}\,du\,dv$$

$$= \iint_D \Big(\underbrace{\big((\boldsymbol{\xi} \times \mathbf{x}_v^*) \cdot \mathbf{n}\big)_u}_{Q_u} - \underbrace{\big((\boldsymbol{\xi} \times \mathbf{x}_u^*) \cdot \mathbf{n}\big)_v}_{P_v} \Big) du\,dv - \iint_D \boldsymbol{\xi} \cdot \big(\mathbf{x}_v^* \times \mathbf{n}_u + \mathbf{n}_v \times \mathbf{x}_u^*\big) du\,dv$$

$$= \int_{\partial D} \underbrace{(\boldsymbol{\xi} \times \mathbf{x}_u^*) \cdot \mathbf{n}}_{P}\, du + \underbrace{(\boldsymbol{\xi} \times \mathbf{x}_v^*) \cdot \mathbf{n}}_{Q}\, dv - \iint_D \boldsymbol{\xi} \cdot \big(\mathbf{x}_v^* \times \mathbf{n}_u + \mathbf{n}_v \times \mathbf{x}_u^*\big) du\,dv.$$

Since $\boldsymbol{\xi} = \mathbf{0}$ on ∂D, the line integral vanishes. Using the equations (††) on p. 59, we find that $\mathbf{x}_v^* \times \mathbf{n}_u = a(\mathbf{x}_u^* \times \mathbf{x}_v^*)$ and $\mathbf{n}_v \times \mathbf{x}_u^* = d(\mathbf{x}_u^* \times \mathbf{x}_v^*)$, so, at long last, we obtain

$$\begin{aligned} \frac{d}{d\varepsilon}\bigg|_{\varepsilon=0} \iint_D \|\mathbf{x}_u \times \mathbf{x}_v\| du\,dv &= \iint_D (\boldsymbol{\xi}_u \times \mathbf{x}_v^* + \mathbf{x}_u^* \times \boldsymbol{\xi}_v) \cdot \mathbf{n}\, du\,dv \\ &= -\iint_D \boldsymbol{\xi} \cdot \big(\mathbf{x}_v^* \times \mathbf{n}_u + \mathbf{n}_v \times \mathbf{x}_u^*\big) du\,dv \\ &= -\iint_D (a+d)\boldsymbol{\xi} \cdot (\mathbf{x}_u^* \times \mathbf{x}_v^*) du\,dv = -\iint_D 2H\boldsymbol{\xi} \cdot \mathbf{n}\, dA, \end{aligned}$$

since $H = \frac{1}{2}\operatorname{tr} S_P$.

We conclude from this, using a two-dimensional analogue of Exercise 2, the following

Theorem 4.2. *Among all (parametrized) surfaces with a given boundary curve, the one of least area is minimal, i.e., has $H = 0$.*

This result, indeed, is the origin of the terminology.

Next, suppose we wish to characterize those closed surfaces (compact surfaces with no boundary) of least area containing a given volume V. To make a parametrized surface closed, we require that $\mathbf{x}(u,v) = \mathbf{x}_0$ for all $(u,v) \in \partial D$. But how do we express the volume constraint in terms of $\mathbf{x}$? The answer comes from the Divergence Theorem and is the three-dimensional analogue of the result of Exercise A.2.5: The volume enclosed by the parametrized surface $\mathbf{x}$ is given by

$$\mathrm{vol}(V) = \frac{1}{3} \iint_D \mathbf{x} \cdot \mathbf{n}\, dA.$$

Thus, the method of Lagrange multipliers suggests that for a surface of least area there must be a constant λ so that $\iint_D (2H - \lambda)\boldsymbol{\xi} \cdot \mathbf{n}\, dA = 0$ for all variations $\boldsymbol{\xi}$ with $\boldsymbol{\xi} = \mathbf{0}$ on ∂D. Once again, using a two-dimensional analogue of Exercise 2, we see that $2H - \lambda = 0$ and hence H must be constant. (Also see Exercise 6.) We conclude:

Theorem 4.3. *Among all (parametrized) surfaces containing a fixed volume, the one of least area has constant mean curvature.*

In particular, a soap bubble should have constant mean curvature. A nontrivial theorem of Alexandrov, analogous to Theorem 3.6 of Chapter 2, states that a smooth, compact surface of constant mean curvature must be a sphere. So soap bubbles should be spheres. How do you explain "double bubbles"?

Example 2. If we ask which surfaces of revolution have constant mean curvature H_0, the statement of Exercise 2.2.20a. leads us to the differential equation

$$\frac{h''}{(1+h'^2)^{3/2}} - \frac{1}{h(1+h'^2)^{1/2}} = 2H_0.$$

(Here the surface is obtained by rotating the graph of h about the coordinate axis.) We can rewrite this equation as follows:

$$\frac{-hh'' + (1+h'^2)}{(1+h'^2)^{3/2}} + 2H_0 h = 0$$

and, multiplying through by h',

$$h'\frac{-hh'' + (1+h'^2)}{(1+h'^2)^{3/2}} + 2H_0 hh' = 0$$
$$\left(\frac{h}{\sqrt{1+h'^2}}\right)' + 2H_0\left(\frac{1}{2}h^2\right)' = 0$$

$$(\dagger) \qquad \frac{h}{\sqrt{1+h'^2}} + H_0 h^2 = \text{const.}$$

We now show that such functions have a wonderful geometric characterization, as suggested in Figure 4.2. Starting with an ellipse with semimajor axis a and semiminor axis b, we consider the locus of one

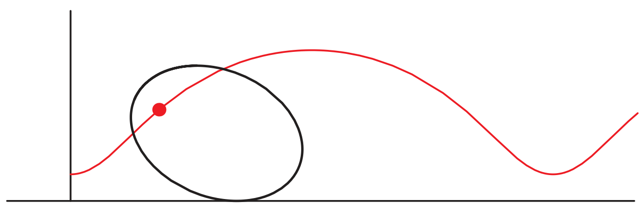


FIGURE 4.2

focus as we roll the ellipse along the x-axis. By definition of an ellipse, we have $\|\overrightarrow{F_1Q}\| + \|\overrightarrow{F_2Q}\| = 2a$, and by Exercise 7, we have $yy_2 = b^2$ (see Figure 4.3). On the other hand, we deduce from Exercise 8 that $\overrightarrow{F_1Q}$ is normal to the curve, and that, therefore, $y = \|\overrightarrow{F_1Q}\|\cos\phi$. Since the “reflectivity” property of the ellipse tells us that $\angle F_1QP_1 \cong \angle F_2QP_2$, we have $y_2 = \|\overrightarrow{F_2Q}\|\cos\phi$. Since $\cos\phi = dx/ds$ and

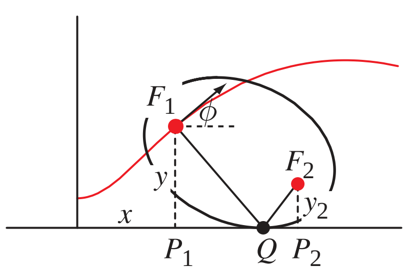


FIGURE 4.3

$ds/dx = \sqrt{1+(dy/dx)^2}$, we have

$$y + \frac{b^2}{y} = y + y_2 = 2a\cos\phi = 2a\frac{dx}{ds}$$

and so

$$0 = y^2 - 2ay\frac{dx}{ds} + b^2 = y^2 - \frac{2ay}{\sqrt{1+y'^2}} + b^2 = 0.$$

Setting $H_0 = -1/2a$, we see that this matches the equation (†) above. ∇

EXERCISES 3.4

♯1. Suppose $g\colon [0,1]\times(-1,1) \to \mathbb{R}$ is continuous and let $G(\varepsilon) = \displaystyle\int_0^1 g(t,\varepsilon)\,dt$. Prove that if $\dfrac{\partial g}{\partial \varepsilon}$ is continuous, then $G'(0) = \displaystyle\int_0^1 \frac{\partial g}{\partial \varepsilon}(t,0)\,dt$. (Hint: Consider $h(\varepsilon) = \displaystyle\int_0^\varepsilon \int_0^1 \frac{\partial g}{\partial \varepsilon}(t,u)\,dt\,du$.)

♯2. *a. Suppose f is a continuous function on $[0,1]$ and $\displaystyle\int_0^1 f(t)\xi(t)\,dt = 0$ for all continuous functions ξ on $[0,1]$. Prove that $f = 0$. (Hint: Take $\xi = f$.)

b. Suppose f is a continuous function on $[0,1]$ and $\displaystyle\int_0^1 f(t)\xi(t)\,dt = 0$ for all continuous functions ξ on $[0,1]$ with $\xi(0) = \xi(1) = 0$. Prove that $f = 0$. (Hint: Take $\xi = \psi f$ for an appropriate continuous function ψ.)

c. Deduce the same result for $\mathcal{C}^1$ functions ξ.

d. Deduce the same result for vector-valued functions $\mathbf{f}$ and $\boldsymbol{\xi}$.

3. Use the Euler-Lagrange equations to show that the shortest path joining two points in the Euclidean plane is a line segment.

4. Use the functional $F(u) = \displaystyle\int_a^b 2\pi u(t)\sqrt{1+(u'(t))^2}\,dt$ to determine the surface of revolution of least area with two parallel circles (perhaps of different radii) as boundary. (Hint: You should end up with the same differential equation as in Exercise 2.2.20.)

5. Prove the analogue of Theorem 4.3 for curves. That is, show that of all closed plane curves enclosing a given area, the circle has the least perimeter. (Cf. Theorem 3.10 of Chapter 1. Hint: Start with Exercise A.2.5. Show that the constrained Euler-Lagrange equations imply that the extremizing curve has constant curvature. Proposition 2.2 of Chapter 1 will help.)

6. Interpreting the integral $\displaystyle\int_0^1 f(t)g(t)\,dt$ as an inner product (dot product) $\langle f, g\rangle$ on the vector space of continuous functions on $[0,1]$, prove that if $\displaystyle\int_0^1 f(t)g(t)\,dt = 0$ for all continuous functions g with $\displaystyle\int_0^1 g(t)\,dt = 0$, then f must be constant. (Hint: Write $f = \langle f, 1\rangle 1 + f^\perp$, where $\langle f^\perp, 1\rangle = 0$.)

7. Prove the *pedal property* of the ellipse: The product of the distances from the foci to the tangent line of the ellipse at any point is a constant (in fact, the square of the semiminor axis).

8. The arclength-parametrized curve $\boldsymbol{\alpha}(s)$ rolls without slipping along the x-axis, starting at the point $\boldsymbol{\alpha}(0) = \mathbf{0}$. A point F is fixed relative to the curve. Let $\boldsymbol{\beta}(s)$ be the curve that F traces out. As indicated in Figure 4.4, let $\theta(s)$ be the angle $\boldsymbol{\alpha}'(s)$ makes with the positive x-axis. Denote by $R_\theta =$

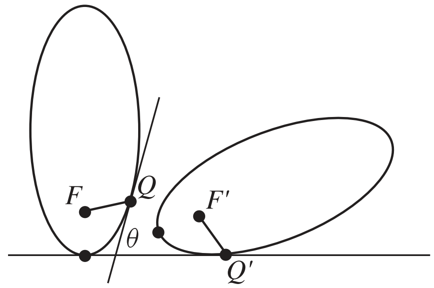


FIGURE 4.4

$\begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$ the matrix that gives rotation of the plane through angle θ.

a. Show that $\boldsymbol{\beta}(s) = (s, 0) + R_{-\theta(s)}(F - \boldsymbol{\alpha}(s))$.

b. Show that $\boldsymbol{\beta}'(s) \cdot R_{-\theta(s)}(F - \boldsymbol{\alpha}(s)) = 0$. That is, as F moves, instantaneously it rotates about the contact point on the x-axis. (Cf. Exercise A.1.4.)

9. Find the path followed by the focus of the parabola $y = x^2/2$ as the parabola rolls along the x-axis. The focus is originally at $(0, 1/2)$. (Hint: See Example 2.)

10. Generalizing Exercise 8, prove that the result remains true if $\boldsymbol{\alpha}$ rolls without slipping along another smooth curve. (Hint: Parametrize the other curve by $\boldsymbol{\gamma}(s)$, where s is arclength of $\boldsymbol{\alpha}$. Note that if the rolling starts at $\boldsymbol{\alpha}(0) = \boldsymbol{\gamma}(0)$, then the fact that the curve rolls without slipping tells us that s is likewise the arclength of $\boldsymbol{\gamma}$.)

APPENDIX

Review of Linear Algebra and Calculus

1. Linear Algebra Review

Recall that the set $\{\mathbf{v}_1, \dots, \mathbf{v}_k\}$ of vectors in $\mathbb{R}^n$ gives a basis for a subspace V of $\mathbb{R}^n$ if and only if every vector $\mathbf{v} \in V$ can be written *uniquely* as a linear combination $\mathbf{v} = c_1\mathbf{v}_1 + \dots + c_k\mathbf{v}_k$. In particular, $\mathbf{v}_1, \dots, \mathbf{v}_n$ will form a basis for $\mathbb{R}^n$ if and only if the $n \times n$ matrix

$$A = \begin{bmatrix} | & | & & | \\ \mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_n \\ | & | & & | \end{bmatrix}$$

is invertible, and are said to be *positively oriented* if the determinant $\det A$ is positive. In particular, given two linearly independent vectors $\mathbf{v}, \mathbf{w} \in \mathbb{R}^3$, the set $\{\mathbf{v}, \mathbf{w}, \mathbf{v}\times\mathbf{w}\}$ always gives a positively oriented basis for $\mathbb{R}^3$.

We say $\mathbf{e}_1, \dots, \mathbf{e}_k \in \mathbb{R}^n$ form an *orthonormal* set in $\mathbb{R}^n$ if $\mathbf{e}_i \cdot \mathbf{e}_j = 0$ for all $i \neq j$ and $\|\mathbf{e}_i\| = 1$ for all $i = 1, \dots, k$. Then we have the following

Proposition 1.1. *If $\{\mathbf{e}_1, \dots, \mathbf{e}_n\}$ is an orthonormal set of vectors in $\mathbb{R}^n$, then they form a basis for $\mathbb{R}^n$ and, given any $\mathbf{v} \in \mathbb{R}^n$, we have $\mathbf{v} = \sum_{i=1}^{n} (\mathbf{v}\cdot\mathbf{e}_i)\mathbf{e}_i$.*

We say an $n \times n$ matrix A is *orthogonal* if $A^{\mathsf{T}}A = I$. It is easy to check that the column vectors of A form an orthonormal basis for $\mathbb{R}^n$ (and the same for the row vectors). Moreover, from the basic formula $A\mathbf{x}\cdot\mathbf{y} = \mathbf{x}\cdot A^{\mathsf{T}}\mathbf{y}$ we deduce that if $\mathbf{e}_1, \dots, \mathbf{e}_k$ form an orthonormal set of vectors in $\mathbb{R}^n$ and A is an orthogonal $n \times n$ matrix, then $A\mathbf{e}_1, \dots, A\mathbf{e}_k$ are likewise an orthonormal set of vectors.

An important issue for differential geometry is to identify the isometries of $\mathbb{R}^3$ (although the same argument will work in any dimension). Recall that an *isometry* of $\mathbb{R}^3$ is a function $\mathbf{f}\colon \mathbb{R}^3 \to \mathbb{R}^3$ so that for any $\mathbf{x}, \mathbf{y} \in \mathbb{R}^3$, we have $\|\mathbf{f}(\mathbf{x}) - \mathbf{f}(\mathbf{y})\| = \|\mathbf{x} - \mathbf{y}\|$. We now prove the

Theorem 1.2. *Any isometry $\mathbf{f}$ of $\mathbb{R}^3$ can be written in the form $\mathbf{f}(\mathbf{x}) = A\mathbf{x} + \mathbf{c}$ for some orthogonal 3×3 matrix A and some vector $\mathbf{c} \in \mathbb{R}^3$.*

Proof. Let $\mathbf{f}(\mathbf{0}) = \mathbf{c}$, and replace $\mathbf{f}$ with the function $\mathbf{f} - \mathbf{c}$. It too is an isometry (why?) and fixes the origin. Then $\|\mathbf{f}(\mathbf{x})\| = \|\mathbf{f}(\mathbf{x}) - \mathbf{f}(\mathbf{0})\| = \|\mathbf{x} - \mathbf{0}\| = \|\mathbf{x}\|$, so that $\mathbf{f}$ preserves lengths of vectors. Using this fact, we prove that $\mathbf{f}(\mathbf{x})\cdot\mathbf{f}(\mathbf{y}) = \mathbf{x}\cdot\mathbf{y}$ for all $\mathbf{x}, \mathbf{y} \in \mathbb{R}^3$. We have

$$\|\mathbf{f}(\mathbf{x}) - \mathbf{f}(\mathbf{y})\|^2 = \|\mathbf{x} - \mathbf{y}\|^2 = (\mathbf{x} - \mathbf{y})\cdot(\mathbf{x} - \mathbf{y}) = \|\mathbf{x}\|^2 - 2\mathbf{x}\cdot\mathbf{y} + \|\mathbf{y}\|^2;$$

on the other hand, in a similar fashion,

$$\|\mathbf{f}(\mathbf{x}) - \mathbf{f}(\mathbf{y})\|^2 = \|\mathbf{f}(\mathbf{x})\|^2 - 2\mathbf{f}(\mathbf{x})\cdot\mathbf{f}(\mathbf{y}) + \|\mathbf{f}(\mathbf{y})\|^2 = \|\mathbf{x}\|^2 - 2\mathbf{f}(\mathbf{x})\cdot\mathbf{f}(\mathbf{y}) + \|\mathbf{y}\|^2.$$

We conclude that $\mathbf{f}(\mathbf{x})\cdot\mathbf{f}(\mathbf{y}) = \mathbf{x}\cdot\mathbf{y}$, as desired.

We next prove that $\mathbf{f}$ must be a linear function. Let $\{\mathbf{e}_1,\mathbf{e}_2,\mathbf{e}_3\}$ be the standard orthonormal basis for $\mathbb{R}^3$, and let $\mathbf{f}(\mathbf{e}_j) = \mathbf{v}_j$, $j = 1,2,3$. It follows from what we've already proved that $\{\mathbf{v}_1,\mathbf{v}_2,\mathbf{v}_3\}$ is also an orthonormal basis. Given an arbitrary vector $\mathbf{x}\in\mathbb{R}^3$, write $\mathbf{x} = \sum\limits_{i=1}^{3} x_i\mathbf{e}_i$ and $f(\mathbf{x}) = \sum\limits_{j=1}^{3} y_j\mathbf{v}_j$. Then it follows from Proposition 1.1 that

$$y_i = \mathbf{f}(\mathbf{x})\cdot\mathbf{v}_i = \mathbf{x}\cdot\mathbf{e}_i = x_i\,,$$

so $\mathbf{f}$ is in fact linear. The matrix A representing $\mathbf{f}$ with respect to the standard basis has as its j^{th} column the vector $\mathbf{v}_j$. Therefore, by our earlier remarks, A is an orthogonal matrix, as required. $\square$

Indeed, recall that if $T\colon\mathbb{R}^n\to\mathbb{R}^n$ is a linear map and $\mathcal{B} = \{\mathbf{v}_1,\dots,\mathbf{v}_k\}$ is a basis for $\mathbb{R}^n$, then the *matrix for T with respect to the basis* $\mathcal{B}$ is the matrix whose j^{th} column consists of the coefficients of $T(\mathbf{v}_j)$ with respect to the basis $\mathcal{B}$. That is, it is the matrix

$$A = \big[a_{ij}\big],\quad \text{where}\quad T(\mathbf{v}_j) = \sum_{i=1}^{n} a_{ij}\mathbf{v}_i\,.$$

Recall that if A is an $n\times n$ matrix (or $T\colon\mathbb{R}^n\to\mathbb{R}^n$ is a linear map), a nonzero vector $\mathbf{x}$ is called an *eigenvector* if $A\mathbf{x} = \lambda\mathbf{x}$ ($T(\mathbf{x}) = \lambda\mathbf{x}$, resp.) for some scalar λ, called the associated *eigenvalue*.

Theorem 1.3. *A symmetric 2×2 matrix $A = \begin{bmatrix} a & b\\ b & c\end{bmatrix}$ (or symmetric linear map $T\colon\mathbb{R}^2\to\mathbb{R}^2$) has two real eigenvalues λ_1 and λ_2, and, provided $\lambda_1\neq\lambda_2$, the corresponding eigenvectors $\mathbf{v}_1$ and $\mathbf{v}_2$ are orthogonal.*

Proof. Consider the function

$$f\colon\mathbb{R}^2\to\mathbb{R},\quad f(\mathbf{x}) = A\mathbf{x}\cdot\mathbf{x} = ax_1^2 + 2bx_1x_2 + cx_2^2.$$

By the maximum value theorem, f has a minimum and a maximum subject to the constraint $g(\mathbf{x}) = x_1^2 + x_2^2 = 1$. Say these occur, respectively, at $\mathbf{v}_1$ and $\mathbf{v}_2$. By the method of Lagrange multipliers, we infer that there are scalars λ_i so that $\nabla f(\mathbf{v}_i) = \lambda_i\nabla g(\mathbf{v}_i)$, $i = 1,2$. By Exercise 5, this means $A\mathbf{v}_i = \lambda_i\mathbf{v}_i$, and so the Lagrange multipliers are actually the associated eigenvalues. Now

$$\lambda_1(\mathbf{v}_1\cdot\mathbf{v}_2) = A\mathbf{v}_1\cdot\mathbf{v}_2 = \mathbf{v}_1\cdot A\mathbf{v}_2 = \lambda_2(\mathbf{v}_1\cdot\mathbf{v}_2).$$

It follows that if $\lambda_1\neq\lambda_2$, we must have $\mathbf{v}_1\cdot\mathbf{v}_2 = 0$, as desired. $\square$

We recall that, in practice, we find the eigenvalues by solving for the roots of the *characteristic polynomial* $p(t) = \det(A - tI)$. In the case of a symmetric 2×2 matrix $A = \begin{bmatrix} a & b\\ b & c\end{bmatrix}$, we obtain the polynomial $p(t) = t^2 - (a+c)t + (ac - b^2)$, whose roots are

$$\lambda_1 = \frac{1}{2}\Big((a+c) - \sqrt{(a-c)^2 + 4b^2}\Big)\qquad\text{and}\qquad \lambda_2 = \frac{1}{2}\Big((a+c) + \sqrt{(a-c)^2 + 4b^2}\Big).$$

EXERCISES A.1

♯*1. Suppose $\{\mathbf{v}_1, \mathbf{v}_2\}$ gives a basis for $\mathbb{R}^2$. Given vectors $\mathbf{x}, \mathbf{y} \in \mathbb{R}^2$, prove that $\mathbf{x} = \mathbf{y}$ if and only if $\mathbf{x} \cdot \mathbf{v}_i = \mathbf{y} \cdot \mathbf{v}_i$, $i = 1, 2$.

*2. The geometric-arithmetic mean inequality states that

$$\sqrt{ab} \le \frac{a+b}{2} \qquad \text{for positive numbers } a \text{ and } b,$$

with equality holding if and only if $a = b$. Give a one-line proof using the Cauchy-Schwarz inequality:

$$|\mathbf{u} \cdot \mathbf{v}| \le \|\mathbf{u}\| \|\mathbf{v}\| \qquad \text{for vectors } \mathbf{u} \text{ and } \mathbf{v} \in \mathbb{R}^n,$$

with equality holding if and only if one is a scalar multiple of the other.

3. Let $\mathbf{w}, \mathbf{x}, \mathbf{y}, \mathbf{z} \in \mathbb{R}^3$. Prove that

$$(\mathbf{w} \times \mathbf{x}) \cdot (\mathbf{y} \times \mathbf{z}) = (\mathbf{w} \cdot \mathbf{y})(\mathbf{x} \cdot \mathbf{z}) - (\mathbf{w} \cdot \mathbf{z})(\mathbf{x} \cdot \mathbf{y}).$$

(Hint: Both sides are linear in each of the four variables, so it suffices to check the result on basis vectors.)

♯4. Suppose $A(t)$ is a differentiable family of 3×3 orthogonal matrices. Prove that $A(t)^{-1}A'(t)$ is always skew-symmetric.

5. If $A = \begin{bmatrix} a & b \\ b & c \end{bmatrix}$ is a symmetric 2×2 matrix, set $f(\mathbf{x}) = A\mathbf{x} \cdot \mathbf{x}$ and check that $\nabla f(\mathbf{x}) = 2A\mathbf{x}$.

2. Calculus Review

Recall that a function $f\colon U \to \mathbb{R}$ defined on an open subset $U \subset \mathbb{R}^n$ is $\mathcal{C}^k$ $(k = 0, 1, 2, \dots, \infty)$ if all its partial derivatives of order $\le k$ exist and are continuous on U. We will use the notation $\dfrac{\partial f}{\partial u}$ and f_u interchangeably, and similarly with higher order derivatives: $\dfrac{\partial^2 f}{\partial v \partial u} = \dfrac{\partial}{\partial v}\Big(\dfrac{\partial f}{\partial u}\Big)$ is the same as f_{uv}, and so on.

One of the extremely important results for differential geometry is the following

Theorem 2.1. *If f is a $\mathcal{C}^2$ function, then $\dfrac{\partial^2 f}{\partial u \partial v} = \dfrac{\partial^2 f}{\partial v \partial u}$ (or $f_{uv} = f_{vu}$).*

The same results apply to vector-valued functions, working with component functions separately.

If $f\colon U \to \mathbb{R}$ is $\mathcal{C}^1$ we can form its *gradient* by taking the vector $\nabla f = \big(f_{x_1}, f_{x_2}, \dots, f_{x_n}\big)$ of its partial derivatives. One of the most fundamental formulas in differential calculus is the *chain rule*:

Theorem 2.2. *Suppose $f\colon \mathbb{R}^n \to \mathbb{R}$ and $\boldsymbol{\alpha}\colon \mathbb{R} \to \mathbb{R}^n$ are differentiable. Then $(f \circ \boldsymbol{\alpha})'(t) = \nabla f(\boldsymbol{\alpha}(t)) \cdot \boldsymbol{\alpha}'(t)$.*

In particular, if $\boldsymbol{\alpha}(0) = P$ and $\boldsymbol{\alpha}'(0) = \mathbf{V} \in \mathbb{R}^n$, then $(f \circ \boldsymbol{\alpha})'(0) = \nabla f(P) \cdot \mathbf{V}$. This is somewhat surprising, as the rate of change of f along $\boldsymbol{\alpha}$ at P depends only on the tangent vector and on nothing more subtle about the curve.

Proposition 2.3. $D_{\mathbf{V}} f(P) = \nabla f(P) \cdot \mathbf{V}$. *Thus, the directional derivative is a linear function of* $\mathbf{V}$.

Proof. If we take $\boldsymbol{\alpha}(t) = P + t\mathbf{V}$, then by definition of the directional derivative, $D_{\mathbf{V}} f(P) = (f \circ \boldsymbol{\alpha})'(0) = \nabla f(P) \cdot \mathbf{V}$. $\square$

Another important consequence of the chain rule, essential throughout differential geometry, is the following

Proposition 2.4. *Suppose* $S \subset \mathbb{R}^n$ *is a subset with the property that any pair of points of* S *can be joined by a* $\mathcal{C}^1$ *curve. Then a* $\mathcal{C}^1$ *function* $f\colon S \to \mathbb{R}$ *with* $\nabla f = \mathbf{0}$ *everywhere is a constant function.*

Proof. Fix $P \in S$ and let $Q \in S$ be arbitrary. Choose a $\mathcal{C}^1$ curve $\boldsymbol{\alpha}$ with $\boldsymbol{\alpha}(0) = P$ and $\boldsymbol{\alpha}(1) = Q$. Then $(f \circ \boldsymbol{\alpha})'(t) = \nabla f(\boldsymbol{\alpha}(t)) \cdot \boldsymbol{\alpha}'(t) = 0$ for all t. It is a consequence of the Mean Value Theorem in introductory calculus that a function $g\colon [0,1] \to \mathbb{R}$ that is continuous on $[0,1]$ and has zero derivative throughout the interval must be a constant. Therefore, $f(Q) = (f \circ \boldsymbol{\alpha})(1) = (f \circ \boldsymbol{\alpha})(0) = f(P)$. It follows that f must be constant on S. $\square$

We will also have plenty of occasion to use the vector versions of the product rule:

Proposition 2.5. *Suppose* $\mathbf{f}, \mathbf{g}\colon \mathbb{R} \to \mathbb{R}^3$ *are differentiable. Then we have*

$$(\mathbf{f} \cdot \mathbf{g})'(t) = \mathbf{f}'(t) \cdot \mathbf{g}(t) + \mathbf{f}(t) \cdot \mathbf{g}'(t) \qquad \textit{and}$$
$$(\mathbf{f} \times \mathbf{g})'(t) = \mathbf{f}'(t) \times \mathbf{g}(t) + \mathbf{f}(t) \times \mathbf{g}'(t).$$

Last, from vector integral calculus, we recall the analogue of the Fundamental Theorem of Calculus in $\mathbb{R}^2$:

Theorem 2.6 (Green's Theorem)**.** *Let* $R \subset \mathbb{R}^2$ *be a region, and let* ∂R *denote its boundary curve, oriented counterclockwise (i.e., so that the region is to its "left"). Suppose* P *and* Q *are* $\mathcal{C}^1$ *functions throughout* R*. Then*

$$\int_{\partial R} P(u,v)du + Q(u,v)dv = \iint_R \left(\frac{\partial Q}{\partial u} - \frac{\partial P}{\partial v}\right) du\,dv.$$

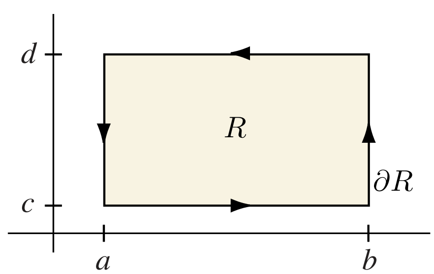


FIGURE 2.1

Proof. We give the proof here just for the case where R is a rectangle. Take $R = [a,b] \times [c,d]$, as shown in Figure 2.1. Now we merely calculate, using the Fundamental Theorem of Calculus appropriately:

$$\iint_R \left(\frac{\partial Q}{\partial u} - \frac{\partial P}{\partial v}\right) du\,dv = \int_c^d \left(\int_a^b \frac{\partial Q}{\partial u} du\right) dv - \int_a^b \left(\int_c^d \frac{\partial P}{\partial v} dv\right) du$$

$$
\begin{aligned}
&= \int_c^d \big(Q(b,v) - Q(a,v)\big)dv - \int_a^b \big(P(u,d) - P(u,c)\big)du \\
&= \int_a^b P(u,c)du + \int_c^d Q(b,v)dv - \int_a^b P(u,d)du - \int_c^d Q(a,v)dv \\
&= \int_{\partial R} P(u,v)du + Q(u,v)dv,
\end{aligned}
$$

as required. $\square$

EXERCISES A.2

$\sharp$1. Suppose $\mathbf{f}\colon (a,b) \to \mathbb{R}^n$ is $\mathcal{C}^1$ and nowhere zero. Prove that $\mathbf{f}/\|\mathbf{f}\|$ is constant if and only if $\mathbf{f}'(t) = \lambda(t)\mathbf{f}(t)$ for some continuous scalar function λ. (Hint: Set $\mathbf{g} = \mathbf{f}/\|\mathbf{f}\|$ and differentiate. Why must $\mathbf{g}' \cdot \mathbf{g} = 0$?)

2. Suppose $\boldsymbol{\alpha}\colon (a,b) \to \mathbb{R}^3$ is twice-differentiable and λ is a nowhere-zero twice differentiable scalar function. Prove that $\boldsymbol{\alpha}$, $\boldsymbol{\alpha}'$, and $\boldsymbol{\alpha}''$ are everywhere linearly independent if and only if $\lambda\boldsymbol{\alpha}$, $(\lambda\boldsymbol{\alpha})'$, and $(\lambda\boldsymbol{\alpha})''$ are everywhere linearly independent.

3. Let $\mathbf{f}, \mathbf{g}\colon \mathbb{R} \to \mathbb{R}^3$ be $\mathcal{C}^1$ vector functions with the property that $\mathbf{f}(0)$ and $\mathbf{g}(0)$ are linearly independent. Suppose

$$
\begin{aligned}
\mathbf{f}'(t) &= a(t)\mathbf{f}(t) + b(t)\mathbf{g}(t) \\
\mathbf{g}'(t) &= c(t)\mathbf{f}(t) - a(t)\mathbf{g}(t)
\end{aligned}
$$

for some continuous functions a, b, and c. Prove that the parallelogram spanned by $\mathbf{f}(t)$ and $\mathbf{g}(t)$ lies in a fixed plane and has constant area.

$\sharp$*4. Prove that for any continuous vector-valued function $\mathbf{f}\colon [a,b] \to \mathbb{R}^3$, we have

$$\left\| \int_a^b \mathbf{f}(t)dt \right\| \le \int_a^b \|\mathbf{f}(t)\| dt.$$

$\sharp$5. Let $R \subset \mathbb{R}^2$ be a region. Prove that

$$\operatorname{area}(R) = \int_{\partial R} u\,dv = -\int_{\partial R} v\,du = \frac{1}{2}\int_{\partial R} -v\,du + u\,dv.$$

3. Differential Equations

Theorem 3.1 (Fundamental Theorem of ODE's). *Suppose $U \subset \mathbb{R}^n$ is open and $I \subset \mathbb{R}$ is an open interval containing 0. Suppose $\mathbf{x}_0 \in U$. If $\mathbf{f}\colon U \times I \to \mathbb{R}^n$ is continuous and Lipschitz in $\mathbf{x}$ (this means that there is a constant C so that $\|\mathbf{f}(\mathbf{x},t) - \mathbf{f}(\mathbf{y},t)\| \le C\|\mathbf{x}-\mathbf{y}\|$ for all $\mathbf{x},\mathbf{y} \in U$ and all $t \in I$), then the* *differential equation*

$$\frac{d\mathbf{x}}{dt} = \mathbf{f}(\mathbf{x}, t), \quad \mathbf{x}(0) = \mathbf{x}_0$$

has a unique solution $\mathbf{x} = \mathbf{x}(t, \mathbf{x}_0)$ *defined for all* t *in some interval* $I' \subset I$*. Moreover, If* $\mathbf{f}$ *is* $\mathcal{C}^k$*, then* $\mathbf{x}$ *is* $\mathcal{C}^k$ *as a function of* both t *and the initial condition* $\mathbf{x}_0$ *(defined for* t *in some interval and* $\mathbf{x}_0$ *in some open set).*

Of special interest to us will be linear differential equations.

Theorem 3.2. *Suppose* $A(t)$ *is a continuous* $n \times n$ *matrix function on an interval* I*. Then the differential equation*

$$\frac{d\mathbf{x}}{dt} = A(t)\mathbf{x}(t), \quad \mathbf{x}_0 = \mathbf{x}_0,$$

has a unique solution on the entire original interval I*.*

For proofs of these, and related, theorems in differential equations, we refer the reader to any standard differential equations text (e.g., Hirsch-Smale or Birkhoff-Rota).

Theorem 3.3. *Let* $k \geq 1$*. Given two* $\mathcal{C}^k$ *vector fields* $\mathbf{X}$ *and* $\mathbf{Y}$ *that are linearly independent on a neighborhood* U *of* $\mathbf{0} \in \mathbb{R}^2$*,* locally *we can choose* $\mathcal{C}^k$ *coordinates* (u, v) *on* $U' \subset U$ *so that* $\mathbf{X}$ *is tangent to the* u*-curves (i.e., the curves* $v = \text{const}$*) and* $\mathbf{Y}$ *is tangent to the* v*-curves (i.e., the curves* $u = \text{const}$*).*

Proof. We make a linear change of coordinates so that $\mathbf{X}(\mathbf{0})$ and $\mathbf{Y}(\mathbf{0})$ are the unit standard basis vectors. Let $\mathbf{x}(t, \mathbf{x}_0)$ be the solution of the differential equation $d\mathbf{x}/dt = \mathbf{X}$, $\mathbf{x}(0) = \mathbf{x}_0$, given by Theorem 3.1. On a neighborhood of $\mathbf{0}$, each point (x, y) can be written as

$$(x, y) = \mathbf{x}(t, (0, v))$$

for some unique t and v, as illustrated in Figure 3.1. If we define the function $\mathbf{f}(t, v) = \mathbf{x}(t, (0, v)) =$

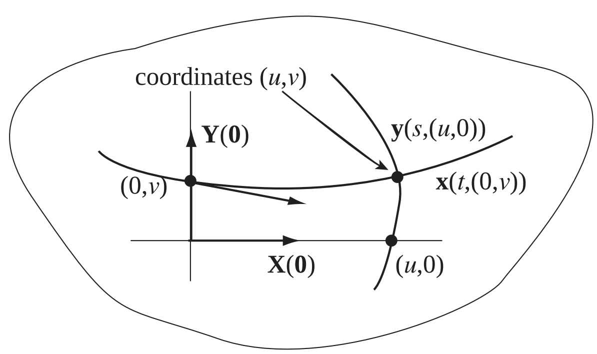


FIGURE 3.1

$(x(t, v), y(t, v))$, we note that $\mathbf{f}_t = \mathbf{X}(\mathbf{f}(t, v))$ and $\mathbf{f}_v(0, 0) = (0, 1)$, so the derivative matrix $D\mathbf{f}(0, 0)$ is the identity matrix. It follows from the Inverse Function Theorem that (locally) we can solve for (t, v) as a $\mathcal{C}^k$ function of (x, y). Note that the level curves of v have tangent vector $\mathbf{X}$, as desired.

Now we repeat this procedure with the vector field $\mathbf{Y}$. Let $\mathbf{y}(s, \mathbf{y}_0)$ be the solution of the differential equation $d\mathbf{y}/ds = \mathbf{Y}$ and write

$$(x, y) = \mathbf{y}(s, (u, 0))$$

for some unique s and u. We similarly obtain (s,u) locally as a $\mathcal{C}^k$ function of (x,y). We claim that (u,v) give the desired coordinates. We only need to check that on a suitable neighborhood of the origin they are independent; but from our earlier discussion we have $v_x = 0$, $v_y = 1$ at the origin, and, analogously, $u_x = 1$ and $u_y = 0$, as well. Thus, the derivative matrix of (u,v) is the identity at the origin and the functions therefore give a local parametrization. $\square$

EXERCISES A.3

1. Suppose $M(s)$ is a differentiable 3×3 matrix function of s, $K(s)$ is a skew-symmetric 3×3 matrix function of s, and
$$M'(s) = M(s)K(s), \quad M(0) = \mathrm{O}.$$
Show that $M(s) = \mathrm{O}$ for all s by showing that the trace of $(M^{\mathsf{T}}M)'(s)$ is identically 0.

2. (*Gronwall inequality* and consequences)
   a. Suppose $f\colon [a,b] \to \mathbb{R}$ is differentiable, nonnegative, and $f(a) = c > 0$. Suppose $g\colon [a,b] \to \mathbb{R}$ is continuous and $f'(t) \le g(t)f(t)$ for all t. Prove that
   $$f(t) \le c\exp\left(\int_a^t g(u)\,du\right) \quad \text{for all } t.$$
   b. Conclude that if $f(a) = 0$, then $f(t) = 0$ for all t.
   c. Suppose now $\mathbf{v}\colon [a,b] \to \mathbb{R}^n$ is a differentiable vector function, and $M(t)$ is a continuous $n \times n$ matrix function for $t \in [a,b)$, and $\mathbf{v}'(t) = M(t)\mathbf{v}(t)$. Apply the result of part b to conclude that if $\mathbf{v}(a) = \mathbf{0}$, then $\mathbf{v}(t) = \mathbf{0}$ for all t. Deduce uniqueness of solutions to *linear* first order differential equations for vector functions. (Hint: Let $f(t) = \|\mathbf{v}(t)\|^2$ and $g(t) = 2n\max\{|m_{ij}(t)|\}$.)
   d. Use part c to deduce uniqueness of solutions to linear n^{th} order differential equations. (Hint: Introduce new variables corresponding to higher derivatives.)

for some unique s and u. We similarly obtain (s,u) locally as a $\mathcal{C}^k$ function of (x,y). We claim that (u,v) give the desired coordinates. We only need to check that on a suitable neighborhood of the origin they are independent; but from our earlier discussion we have $v_x = 0$, $v_y = 1$ at the origin, and, analogously, $u_x = 1$ and $u_y = 0$, as well. Thus, the derivative matrix of (u,v) is the identity at the origin and the functions therefore give a local parametrization. $\square$

EXERCISES A.3

1. Suppose $M(s)$ is a differentiable 3×3 matrix function of s, $K(s)$ is a skew-symmetric 3×3 matrix function of s, and
$$M'(s) = M(s)K(s), \quad M(0) = \mathrm{O}.$$
Show that $M(s) = \mathrm{O}$ for all s by showing that the trace of $(M^{\mathsf{T}}M)'(s)$ is identically 0.

2. (*Gronwall inequality* and consequences)
   a. Suppose $f\colon[a,b]\to\mathbb{R}$ is differentiable, nonnegative, and $f(a)=c>0$. Suppose $g\colon[a,b]\to\mathbb{R}$ is continuous and $f'(t)\le g(t)f(t)$ for all t. Prove that
$$f(t)\le c\exp\left(\int_a^t g(u)du\right) \quad \text{for all } t.$$
   b. Conclude that if $f(a)=0$, then $f(t)=0$ for all t.
   c. Suppose now $\mathbf{v}\colon[a,b]\to\mathbb{R}^n$ is a differentiable vector function, and $M(t)$ is a continuous $n\times n$ matrix function for $t\in[a,b)$, and $\mathbf{v}'(t)=M(t)\mathbf{v}(t)$. Apply the result of part b to conclude that if $\mathbf{v}(a)=\mathbf{0}$, then $\mathbf{v}(t)=\mathbf{0}$ for all t. Deduce uniqueness of solutions to *linear* first order differential equations for vector functions. (Hint: Let $f(t)=\|\mathbf{v}(t)\|^2$ and $g(t)=2n\max\{|m_{ij}(t)|\}$.)
   d. Use part c to deduce uniqueness of solutions to linear n^{th} order differential equations. (Hint: Introduce new variables corresponding to higher derivatives.)

ANSWERS TO SELECTED EXERCISES

1.1.1 $\boldsymbol{\alpha}(t) = \left(\frac{1-t^2}{1+t^2}, \frac{2t}{1+t^2}\right)$.

1.1.4 We parametrize the curve by $\boldsymbol{\alpha}(t) = (t, f(t))$, $a \le t \le b$, and so length$(\boldsymbol{\alpha}) = \int_a^b \|\boldsymbol{\alpha}'(t)\|\,dt = \int_a^b \sqrt{1+(f'(t))^2}\,dt$.

1.1.6 $\boldsymbol{\beta}(s) = \left(\frac{1}{2}(\sqrt{s^2+4}+s), \frac{1}{2}(\sqrt{s^2+4}-s), \sqrt{2}\ln((\sqrt{s^2+4}+s)/2)\right)$.

1.2.1 **c.** $\kappa = \frac{1}{2\sqrt{2}\sqrt{1-s^2}}$

1.2.3 **a.** $\mathbf{T} = \frac{1}{2}(\sqrt{1+s}, -\sqrt{1-s}, \sqrt{2})$, $\kappa = \frac{1}{2\sqrt{2}\sqrt{1-s^2}}$, $\mathbf{N} = 1/\sqrt{2}(\sqrt{1-s}, \sqrt{1+s}, 0)$, $\mathbf{B} = \frac{1}{2}(-\sqrt{1+s}, \sqrt{1-s}, \sqrt{2})$, $\tau = \frac{1}{2\sqrt{2}\sqrt{1-s^2}}$; **c.** $\mathbf{T} = \frac{1}{\sqrt{2}\sqrt{1+t^2}}(t, \sqrt{1+t^2}, 1)$, $\kappa = \tau = 1/2(1+t^2)$, $\mathbf{N} = \frac{1}{\sqrt{1+t^2}}(1, 0, -t)$, $\mathbf{B} = \frac{1}{\sqrt{2}\sqrt{1+t^2}}(-t, \sqrt{1+t^2}, -1)$

1.2.5 $\kappa = 1/\sinh t$ (which we see, once again, is the absolute value of the slope).

1.2.6 $\mathbf{B}' = (\mathbf{T}\times\mathbf{N})' = \mathbf{T}'\times\mathbf{N} + \mathbf{T}\times\mathbf{N}' = (\kappa\mathbf{N})\times\mathbf{N} + \mathbf{T}\times(-\kappa\mathbf{T}+\tau\mathbf{B}) = \tau(\mathbf{T}\times\mathbf{B}) = \tau(-\mathbf{N})$, as required.

1.2.9 **b.** If all the osculating planes pass through the origin, then there are scalar functions λ and μ so that $\mathbf{0} = \boldsymbol{\alpha} + \lambda\mathbf{T} + \mu\mathbf{N}$. Differentiating and using the Frenet formulas, we obtain $\mathbf{0} = \mathbf{T} + \kappa\lambda\mathbf{N} + \lambda'\mathbf{T} + \mu\big(-\kappa\mathbf{T} + \tau\mathbf{B}\big) + \mu'\mathbf{N}$; collecting terms, we have $\mathbf{0} = \big(1 + \lambda' - \kappa\mu\big)\mathbf{T} + \big(\kappa\lambda + \mu'\big)\mathbf{N} + \mu\tau\mathbf{B}$. Since $\{\mathbf{T}, \mathbf{N}, \mathbf{B}\}$ is a basis for $\mathbb{R}^3$, we infer, in particular, that $\mu\tau = 0$. (We could also just have taken the dot product of the entire expression with $\mathbf{B}$.) $\mu(s) = 0$ leads to a contradiction, so we must have $\tau = 0$ and so the curve is planar.

1.2.11 We have $\boldsymbol{\alpha}'\times\boldsymbol{\alpha}'' = \kappa\upsilon^3\mathbf{B}$, so $\boldsymbol{\alpha}'\times\boldsymbol{\alpha}''' = (\boldsymbol{\alpha}'\times\boldsymbol{\alpha}'')' = (\kappa\upsilon^3\mathbf{B})' = (\kappa\upsilon^3)'\mathbf{B} + (\kappa\upsilon^3)(-\tau\upsilon\mathbf{N})$, so $(\boldsymbol{\alpha}'\times\boldsymbol{\alpha}''')\cdot\boldsymbol{\alpha}'' = -\kappa^2\tau\upsilon^6$. Therefore, $\tau = \boldsymbol{\alpha}'\cdot(\boldsymbol{\alpha}''\times\boldsymbol{\alpha}''')/(\kappa^2\upsilon^6)$, and inserting the formula of Proposition 2.2 gives the result.

1.2.25 **a.** Consider the unit normal $\mathbf{A}_{s,t}$ to the plane through $P = \mathbf{0}$, $Q = \boldsymbol{\alpha}(s)$, and $R = \boldsymbol{\alpha}(t)$. Choosing coordinates so that $\mathbf{T}(0) = (1,0,0)$, $\mathbf{N}(0) = (0,1,0)$, and $\mathbf{B}(0) = (0,0,1)$, we apply Proposition 2.6 to obtain

$$\boldsymbol{\alpha}(s)\times\boldsymbol{\alpha}(t) = \frac{st(s-t)}{12}\left(-\kappa_0^2\tau_0 st + \ldots, 2\kappa_0\tau_0(s+t) + \ldots, -6\kappa_0 + 2\kappa_0'(s+t) - \kappa_0^3 st + \ldots\right),$$

so $\mathbf{A}_{s,t} = \dfrac{\boldsymbol{\alpha}(s)\times\boldsymbol{\alpha}(t)}{\|\boldsymbol{\alpha}(s)\times\boldsymbol{\alpha}(t)\|} \to \mathbf{A} = (0,0,-1)$ as $s,t \to 0$. Thus, the plane through P with normal $\mathbf{A}$ is the osculating plane.

1.2.25 **a. cont.** Alternatively, let the equation of the plane through P, Q, and R be $\mathbf{A}_{s,t}\cdot\mathbf{x}=0$ (where we choose $\mathbf{A}_{s,t}$ to vary continuously with length 1). We want to determine $\mathbf{A}=\lim_{s,t\to 0}\mathbf{A}_{s,t}$. For fixed s and t, consider the function $F_{s,t}(u)=\mathbf{A}_{s,t}\cdot\boldsymbol{\alpha}(u)$. Then $F_{s,t}(0)=F_{s,t}(s)=F_{s,t}(t)=0$, so, by the mean value theorem, there are ξ_1 and ξ_2 so that $F'_{s,t}(\xi_1)=F'_{s,t}(\xi_2)=0$, hence η so that $F''_{s,t}(\eta)=0$. Now $F'_{s,t}(0)=\mathbf{A}_{s,t}\cdot\mathbf{T}(0)$ and $F''_{s,t}(0)=\mathbf{A}_{s,t}\cdot\kappa_0\mathbf{N}(0)$. Since $\xi_i\to 0$ and $\eta\to 0$ as $s,t\to 0$, we obtain $\mathbf{A}\cdot\mathbf{T}(0)=\mathbf{A}\cdot\mathbf{N}(0)=0$, so $\mathbf{A}=\pm\mathbf{B}(0)$, as desired.

1.3.4 Let $L=\text{length}(C)$. Then by Theorem 3.5 we have $2\pi=\int_0^L\kappa(s)\,ds\le\int_0^L c\,ds=cL$, so $L\ge 2\pi/c$.

2.1.3 **a.** $E=a^2$, $F=0$, $G=a^2\sin^2 u$; **d.** $E=G=a^2\cosh^2 u$, $F=0$

2.1.4 **a.** $4\pi^2 ab$

2.1.5 Say all the normal lines pass through the origin. Then there is a function λ so that $\mathbf{x}=\lambda\mathbf{n}$. Differentiating, we have $\mathbf{x}_u=\lambda\mathbf{n}_u+\lambda_u\mathbf{n}$ and $\mathbf{x}_v=\lambda\mathbf{n}_v+\lambda_v\mathbf{n}$. Dotting with $\mathbf{n}$, we get $0=\lambda_u=\lambda_v$. Therefore, λ is a constant and so $\|\mathbf{x}\|=\text{const}$. Alternatively, from the statement $\mathbf{x}=\lambda\mathbf{n}$ we proceed as follows. Since $\mathbf{n}\cdot\mathbf{x}_u=\mathbf{n}\cdot\mathbf{x}_v=0$, we have $\mathbf{x}\cdot\mathbf{x}_u=\mathbf{x}\cdot\mathbf{x}_v=0$. Therefore, $(\mathbf{x}\cdot\mathbf{x})_u=(\mathbf{x}\cdot\mathbf{x})_v=0$, so $\|\mathbf{x}\|^2$ is constant.

2.1.7 For $\mathbf{x}$ to be conformal, we must have $E=G$ and $F=0$; for it to preserve area we must have $1=\sqrt{EG-F^2}$, so $E=G=1$ and $F=0$, which characterizes a local isometry with the plane. The converse is immediate.

2.1.8 We check that $E=G=4/(1+u^2+v^2)^2$ and $F=0$, so the result follows from Exercise 6.

2.1.11 **b.** One of these is: $\mathbf{x}(u,v)=(\cos u+v\sin u,\sin u-v\cos u,v)$.

2.1.16 **a.** If $a\cosh(1/a)=R$, the area is $2\pi\big(a+R\sqrt{R^2-a^2}\big)$.

2.2.1 If u- and v-curves are lines of curvature, then $F=0$ (because principal directions are orthogonal away from umbilic points) and $m=S_P(\mathbf{x}_u)\cdot\mathbf{x}_v=k_1\mathbf{x}_u\cdot\mathbf{x}_v=0$. Moreover, if $S_P(\mathbf{x}_u)=k_1\mathbf{x}_u$ and $S_P(\mathbf{x}_v)=k_2\mathbf{x}_v$, we dot with $\mathbf{x}_u$ and $\mathbf{x}_v$, respectively, to obtain $\ell=Ek_1$ and $n=Gk_2$. Conversely, setting $S_P(\mathbf{x}_u)=a\mathbf{x}_u+b\mathbf{x}_v$, we infer that if $F=m=0$, then $0=S_P(\mathbf{x}_u)\cdot\mathbf{x}_v=Fa+Gb=Gb$, and so $b=0$. Therefore, $\mathbf{x}_u$ (and, similarly, $\mathbf{x}_v$) is an eigenvector of S_P.

2.2.3 **b.** $\ell=b$, $m=0$, $n=\cos u(a+b\cos u)$, $S_P=\begin{bmatrix}1/b & 0\\ 0 & \cos u/(a+b\cos u)\end{bmatrix}$, $H=\frac{1}{2}\big(\frac{1}{b}+\frac{\cos u}{a+b\cos u}\big)$, $K=\frac{\cos u}{b(a+b\cos u)}$; **d.** $\ell=-a$, $m=0$, $n=a$, $S_P=\begin{bmatrix}-(1/a)\operatorname{sech}^2 u & 0\\ 0 & (1/a)\operatorname{sech}^2 u\end{bmatrix}$, $H=0$, $K=-(1/a)^2\operatorname{sech}^4 u$.

2.2.5 We know from Example 1 of Chapter 1, Section 2 that the principal normal of the helix points along the ruling and is therefore orthogonal to $\mathbf{n}$. As we move along a ruling, $\mathbf{n}$ twists in a plane orthogonal to the ruling, so its directional derivative in the direction of the ruling is orthogonal to the ruling.

2.2.6 $E = \tanh^2 u$, $F = 0$, $G = \operatorname{sech}^2 u$, $-\ell = \operatorname{sech} u \tanh u = n$, $m = 0$

2.3.2 **d.** $\Gamma^v_{uv} = \Gamma^v_{vu} = f'(u)/f(u)$, $\Gamma^u_{vv} = -f(u)f'(u)$, all others 0.

2.4.4 $\kappa_g = \cot u_0$; we can also deduce this from Figure 3.1, as the curvature vector $\kappa\mathbf{N} = (1/\sin u_0)\mathbf{N}$ has tangential component $-(1/\sin u_0)\cos u_0 \mathbf{x}_u = \cot u_0(\mathbf{n}\times\mathbf{T})$.

2.4.9 Only circles. By Exercise 2 such a curve will also have constant curvature, and by Meusnier's Formula, Proposition 2.5, the angle ϕ between $\mathbf{N}$ and $\mathbf{n} = \boldsymbol{\alpha}$ is constant. Differentiating $\boldsymbol{\alpha}\cdot\mathbf{N} = \cos\phi = \text{const}$ yields $\tau(\boldsymbol{\alpha}\cdot\mathbf{B}) = 0$. Either $\tau = 0$, in which case the curve is planar, or else $\boldsymbol{\alpha}\cdot\mathbf{B} = 0$, in which case $\boldsymbol{\alpha} = \pm\mathbf{N}$, so $\tau = \mathbf{N}'\cdot\mathbf{B} = \pm\boldsymbol{\alpha}'\cdot\mathbf{B} = \pm\mathbf{T}\cdot\mathbf{B} = 0$. (In the latter case, the curve is a great circle.)

2.4.18 **a.** Obviously, the meridians are geodesics and the central circle $r = r_0$ is the only parallel that is a geodesic. Observe that if we have some other geodesic, then $r\cos\phi = c$ and $c < r_0$. The geodesic with $r\cos\phi = c$ will cross the central circle and then either approach one of the parallels $r = c$ asymptotically or hit one of the parallels $r = c$ tangentially and bounce back and forth between those two parallels. In either event, such a geodesic is bounded. (In fact, if a geodesic approaches a parallel asymptotically, that parallel *must* be a geodesic; see Exercise 27.)

2.4.24 The geodesics are of the form $\cosh^2 u + (v + c_1)^2 = c_2^2$ for constants c_1 and c_2.

3.1.1 **a.** $2\pi\sin u_0$

3.1.2 **a.** yes, yes, **b.** yes, yes, **c.** yes, no.

3.2.1 **b.** The semicircle centered at $(2,0)$ of radius $\sqrt{5}$; $d(P,Q) = \ln\big((3+\sqrt{5})/2\big) \approx 0.962$.

3.2.11 $\kappa_g = \coth R$

3.3.4 We have $\kappa_n = \mathrm{II}(\mathbf{e}_1,\mathbf{e}_1) = -d\mathbf{e}_3(\mathbf{e}_1)\cdot\mathbf{e}_1 = \omega_{13}(\mathbf{e}_1)$. Since $\mathbf{e}_3 = \sin\theta\bar{\mathbf{e}}_2 + \cos\theta\bar{\mathbf{e}}_3$, the calculations of Exercise 3 show that $\omega_{13} = \sin\theta\overline{\omega}_{12} + \cos\theta\overline{\omega}_{13}$, so $\omega_{13}(\mathbf{e}_1) = \sin\theta\overline{\omega}_{12}(\mathbf{e}_1) = \kappa\sin\theta$. Here θ is the angle between $\mathbf{e}_3$ and $\bar{\mathbf{e}}_3$, so this agrees with our previous result.

3.3.8 We have $\omega_1 = b\,du$ and $\omega_2 = (a + b\cos u)\,dv$, so $\omega_{12} = -\sin u\,dv$ and $d\omega_{12} = -\cos u\,du\wedge dv = -\left(\dfrac{\cos u}{b(a+b\cos u)}\right)\omega_1\wedge\omega_2$, so $K = \dfrac{\cos u}{b(a+b\cos u)}$.

3.4.2 **a.** Taking $\xi = f$ gives us $\int_0^1 f(t)^2\,dt = 0$. Since $f(t)^2 \geq 0$ for all t, if $f(t_0) \neq 0$, we have an interval $[t_0-\delta, t_0+\delta]$ on which $f(t)^2 \geq f(t_0)^2/2$, and so $\int_0^1 f(t)^2\,dt \geq f(t_0)^2\delta > 0$.

3.4.9 $y = \frac{1}{2}\cosh(2x)$

A.1.1 Consider $\mathbf{z} = \mathbf{x} - \mathbf{y}$. Then we know that $\mathbf{z}\cdot\mathbf{v}_i = 0$, $i = 1, 2$. Since $\{\mathbf{v}_1,\mathbf{v}_2\}$ is a basis for $\mathbb{R}^2$, there are scalars a and b so that $\mathbf{z} = a\mathbf{v}_1 + b\mathbf{v}_2$. Then $\mathbf{z}\cdot\mathbf{z} = \mathbf{z}\cdot(a\mathbf{v}_1 + b\mathbf{v}_2) = a(\mathbf{z}\cdot\mathbf{v}_1) + b(\mathbf{z}\cdot\mathbf{v}_2) = 0$, so $\mathbf{z} = \mathbf{0}$, as desired.

A.1.2 Hint: Take $\mathbf{u} = (\sqrt{a}, \sqrt{b})$ and $\mathbf{v} = (\sqrt{b}, \sqrt{a})$.

A.2.4 Let $\mathbf{v} = \int_a^b \mathbf{f}(t)\,dt$. Note that the result is obvious if $\mathbf{v} = \mathbf{0}$. We have $\|\mathbf{v}\|^2 = \mathbf{v}\cdot\int_a^b \mathbf{f}(t)\,dt = \int_a^b \mathbf{v}\cdot\mathbf{f}(t)\,dt \leq \int_a^b \|\mathbf{v}\|\|\mathbf{f}(t)\|\,dt = \|\mathbf{v}\|\int_a^b \|\mathbf{f}(t)\|\,dt$ (using the Cauchy-Schwarz inequality $\mathbf{u}\cdot\mathbf{v} \leq \|\mathbf{u}\|\,\|\mathbf{v}\|$), so, if $\mathbf{v} \neq \mathbf{0}$, we have $\|\mathbf{v}\| \leq \int_a^b \|\mathbf{f}(t)\|\,dt$, as needed.

INDEX

angle excess, 83
arclength, 6
asymptotic curve, 48, 51, 55
asymptotic direction, 48, 54, 56

Bäcklund, 106
Bertrand mates, 21
binormal vector, 11
Bishop frame, 33

$\mathcal{C}^k$, 1, 35, 116
catenary, 5
catenoid, 43, 66
Cauchy-Schwarz inequality, 116
chain rule, 116
characteristic polynomial, 115
Christoffel symbols, 57
Clairaut's relation, 73, 77, 78
Codazzi equations, 59, 63, 104
compact, 61
cone angle, 90
conformal, 40
connection form, 104
convex, 28
covariant constant, 67
covariant derivative, 67
Crofton's formula, 25, 33
cross ratio, 99
cubic
  cuspidal, 2
  nodal, 2
  twisted, 3
curvature, 11
curve, simple closed, 26
cycloid, 3
cylindrical projection, 42

Darboux frame, 70, 103, 105
developable, *see* ruled surface, developable
directrix, 38
Dupin indicatrix, 56

eigenvalue, 115
eigenvector, 115
elliptic point, 50
Euler characteristic, 85
exterior angle, 33, 83

first fundamental form, 39
flat, 49, 60, 61, 65, 77, 84, 90, 104
Foucault pendulum, 69
Frenet formulas, 11
Frenet frame, 11
functional, 108

Gauss equation, 60, 63, 104
Gauss map, 24, 44
Gauss-Bonnet formula, 83, 86, 96, 105
Gauss-Bonnet Theorem, 95
  global, 86
  local, 83
Gaussian curvature, 49, 51, 53, 57, 60, 82, 104
  constant, 62, 92
generalized helix, 15
geodesic, 70
geodesic curvature, 71
globally isometric, 75
gradient, 116
Green's Theorem, 82, 117
Gronwall inequality, 120

h, 93
H, 49
helicoid, 36, 48, 65, 66
helix, 3
holonomy, 79, 82
hyperbolic plane, 91
  Klein-Beltrami model, 101

Poincaré model, 100
hyperbolic point, 50

inversion, 99
involute, 19
isometry, 114

K, 49
k-point contact, 56
knot, 26

Laplacian, 64
line of curvature, 47
linear fractional transformation, 94
locally isometric, 39

mean curvature, 49
meridian, 38, 52
metric, 74
Meusnier's Formula, 51
minimal surface, 49, 64, 65, 110
moving frame, 101

normal curvature, 51
normal field, 32
normal plane, 17

oriented, 84
orthogonal, 114
orthonormal, 114
osculating circle, 22
osculating plane, 17, 22
osculating sphere, 22

pacman, 42, 75, 90
parabolic point, 50
parallel, 38, 52, 67, 75, 95
parallel translate, 68
parametrization
regular, 1, 35
parametrized by arclength, 7
parametrized curve, 1
pedal property, 112
planar point, 50
Poincaré disk, 100
positively oriented, 114
principal curvature, 47
constant, 65, 66
principal direction, 47, 54
principal normal vector, 11
profile curve, 38
pseudosphere, 51

rectifying plane, 17
reflection, 98
regular, 1
regular parametrization, 35
rotation index, 27
ruled surface, 38
developable, 61, 65, 77
ruling, 38

second fundamental form, 46, 53
shape operator, 45, 53
smooth, 1, 35
spherical coordinates, 37
stereographic projection, 37
support line, 32
surface, 35
surface area, 41
surface of revolution, 37
symmetric, 45

tangent indicatrix, 24
tangent plane, 38
Theorema Egregium, 60, 104
torsion, 11
torus, 36
total curvature, 24, 88
total twist, 32
tractrix, 5, 13
triply orthogonal system, 55
Tschebyschev net, 43
twist, 32

u-, v-curves, 35
ultraparallels, 95
umbilic, 50
unit normal, 39
unit tangent vector, 11

variation, 108
vector field, 67
velocity, 1
vertex, 29

zone, 41