Classical Differential Geometry

Peter Petersen

Preface

This is an evolving set of lecture notes on the classical theory of curves and surfaces.

The first 5 chapters can be covered in a one quarter class. The first 6 chapters in a one semester class. Chapters 6 and 7 can be covered in a second quarter class.

An excellent reference for the classical treatment of differential geometry is the book by Struik [**2**]. There is another more descriptive guide by Hilbert and Cohn-Vossen [**1**]. This book covers both geometry and differential geometry essentially without the use of calculus. It contains many interesting results and gives excellent descriptions of many of the constructions and results in differential geometry.

Contents

CHAPTER 1

General Curve Theory

One of the key aspects in geometry is *invariance*. This can be somewhat difficult to define, but the idea is that the properties or measurements under discussion should be described in such a way that they they make sense without reference to a special coordinate system. This can be quite difficult, but it has been a guiding principle since the ancients Greeks started formulating geometry. We'll often take for granted that we work in a Euclidean space where we know how to compute distances, angles, areas, and even volumes of simple geometric figures. Descartes discovered that these types of geometries could be described by what we call Cartesian space through coordinatizing the Euclidean space with Cartesian coordinates. This is the general approach we shall use, but it is still worthwhile to occasionally try to understand measurements not just algebraically or analytically, but also purely descriptively in geometric terms.

For example, how does one define a circle? It can defined as a set of points given by a specific type of equation, it can be given as a parametric curve, or it can be described as the collection of points at a fixed distance from the center. Using the latter definition without referring to coordinates is often a very useful tool in solving many problems.

There are as yet no figures in the text. It is however easy to find pictures of all the curves and surfaces mentioned in the text. A simple web search for "cissoid" or "conchoid of Nicomedes" or "minimal surfaces" or "Enneper's surface" will quickly bring up lots of information. Wikipedia is a very good general source. There are also other good sources that are commercial and can therefore not be mentioned directly.

1.1. Curves

The primary goal in the geometric theory of *curves* is to find ways of measuring that do not take in to account how a given curve is parametrized or how Euclidean space is coordinatized. However, it is generally hard to measure anything without coordinatizing space and parametrizing the curve. Thus the idea will be to see if some sort of canonical parametrization might exist and secondly to also show that our measurements can be defined using whatever parametrization the curve comes with. We will also try to make sure that our formulas are so that they do not necessarily refer to a specific set of Cartesian coordinates. To understand more general types of coordinates requires quite a bit of work and this will not be done until we introduce surfaces later in these notes.

Imagine traveling in a car or flying an airplane. You can keep track of time and/or the odometer and for each value of the time and/or odometer reading do a GPS measurement of where you are. This will give you a curve that is parametrized by time or distance traveled. The goal of curve theory is to decide what further measurements are needed to retrace the precise path traveled. Clearly one must also measure how one turns and that becomes the important thing to describe mathematically.

The fundamental dynamical vectors of a curve whose position is denoted $\mathbf{q}$ are the *velocity* $\mathbf{v} = \frac{d\mathbf{q}}{dt}$, *acceleration* $\mathbf{a} = \frac{d^2\mathbf{q}}{dt^2}$, and *jerk* $\mathbf{j} = \frac{d^3\mathbf{q}}{dt^3}$. The *tangent line* to a curve $\mathbf{q}$ at $\mathbf{q}(t)$ is the line through $\mathbf{q}(t)$ with direction $\mathbf{v}(t)$. The goal is to find geometric quantities that depend on velocity (or tangent lines), acceleration, and jerk that completely determine the path of the curve when we use some parameter t to travel along it.

Most of the curves we shall study will given as *parametrized curves*, i.e.,

$$\mathbf{q}(t) = \begin{bmatrix} x(t) \\ y(t) \\ \vdots \end{bmatrix} : I \to \mathbb{R}^n,$$

where $I \subset \mathbb{R}$ is an interval. Such a curve might be constant, which is equivalent to its velocity vanishing everywhere.

Occasionally curves are given to us in a more implicit form. They could come as solutions to first order differential equations

$$\frac{d\mathbf{q}}{dt} = F(\mathbf{q}(t), t).$$

In this case we obtain a unique solution (also called an integral curve) as long as we have an initial position $\mathbf{q}(t_0) = \mathbf{q}_0$ at some initial time t_0. In case the function $F(\mathbf{q})$ only depends on the position we can visualize it as a vector field as it gives a vector at each position. The solutions are then seen as curves whose velocity at each position $\mathbf{q}$ is the vector $\mathbf{v} = F(\mathbf{q})$.

Very often the types of differential equations are of second or even higher order

$$\frac{d^2\mathbf{q}}{dt^2} = F\left(\mathbf{q}(t), \frac{d\mathbf{q}}{dt}, t\right).$$

In this case we have to prescribe both the initial position $\mathbf{q}(t_0) = \mathbf{q}_0$ and velocity $\mathbf{v}(t_0) = \mathbf{v}_0$ in order to obtain a unique solution curve.

Another very general method for generating curves is through equations. In general one function $F(x, y) : \mathbb{R}^2 \to \mathbb{R}$ gives a collection of planar curves via the level sets

$$F(x, y) = c.$$

The implicit function theorem guarantees us that we get a unique curve as a graph over either x or y when the gradient of F doesn't vanish. The gradient is the vector

$$\nabla F = \begin{bmatrix} \frac{\partial F}{\partial x} \\ \frac{\partial F}{\partial y} \end{bmatrix}.$$

Geometrically the gradient is perpendicular to the level sets. This means that the level sets themselves have tangents that are given by the directions

$$\begin{bmatrix} -\frac{\partial F}{\partial y} \\ \frac{\partial F}{\partial x} \end{bmatrix}$$

as this vector is orthogonal to the gradient. This in turn offers us a different way of finding these levels as parametrized curves since they now also appear as solutions to the differential equation

$$\begin{bmatrix} \frac{dx}{dt} \\ \frac{dy}{dt} \end{bmatrix} = \begin{bmatrix} -\frac{\partial F}{\partial y}(x(t), y(t)) \\ \frac{\partial F}{\partial x}(x(t), y(t)) \end{bmatrix}.$$

In three variables we need two functions as such functions have level sets that are surfaces:

$$\begin{aligned} F_1(x, y, z) &= c_1, \\ F_2(x, y, z) &= c_2. \end{aligned}$$

In this case we also have a differential equation approach. Both of the gradients ∇F_1 and ∇F_2 are perpendicular to their level sets. Thus the cross product $\nabla F_1 \times \nabla F_2$ is tangent to the intersection of these two surfaces and we can describe the curves as solutions to

$$\frac{d\mathbf{q}}{dt} = (\nabla F_1 \times \nabla F_2)(\mathbf{q}).$$

It is important to realize that when we are looking for solutions to a first order system

$$\frac{d\mathbf{q}}{dt} = F(\mathbf{q}(t)),$$

then we geometrically obtain the same curves if we consider

$$\frac{d\mathbf{q}}{dt} = \lambda(\mathbf{q}(t)) F(\mathbf{q}(t)),$$

where λ is some scalar function, as the directions of the velocities stay the same. However, the parametrizations of the curves will change.

Classically curves were given descriptively in terms of geometric or even mechanical constructions. Thus a circle is the set of points in the plane that all have a fixed distance R to a fixed center. It became more common starting with Descartes to describe them by equations. Only about 1750 did Euler switch to considering parametrized curves. It is also worth mentioning that what we call curves used to be referred to as *lines*. This terminology still has remnants in some of the terms we introduce later.

We present a few classical examples of these constructions in the plane.

Example 1.1.1. Consider the equation

$$F(x, y) = x^2 + y^2 = c.$$

When $c > 0$ this describes a circle of radius $\sqrt{c}$. When $c = 0$ we only get the origin, while when $c < 0$ there are no solutions. The gradient is given by $(2x, 2y)$ and only vanishes at the origin.

The differential equation describing the level sets is

$$\begin{bmatrix} \frac{dx}{dt} \\ \frac{dy}{dt} \end{bmatrix} = \begin{bmatrix} -2y \\ 2x \end{bmatrix}.$$

The solutions are given by $\mathbf{q}(t) = R(\cos(2(t+\varphi)), \sin(2(t+\varphi)))$ where the constants R and φ can be adjusted according to any given initial position. A more convenient parametrization happens when we scale the system to become

$$\begin{bmatrix} \frac{dx}{dt} \\ \frac{dy}{dt} \end{bmatrix} = \begin{bmatrix} -y \\ x \end{bmatrix}$$

so that the solutions are $\mathbf{q}(\theta) = R(\cos(\theta+\varphi), \sin(\theta+\varphi))$ with θ being the angle to the x-axis. Yet a further scaling is possible as long as we exclude the origin

$$\begin{bmatrix} \frac{dx}{dt} \\ \frac{dy}{dt} \end{bmatrix} = \frac{1}{\sqrt{x^2+y^2}} \begin{bmatrix} -y \\ x \end{bmatrix}.$$

This time the solutions are given by

$$\mathbf{q}(\theta) = R\left(\cos\left(\frac{\theta+\varphi}{R}\right), \sin\left(\frac{\theta+\varphi}{R}\right)\right)$$

and we have to assume that $R > 0$.

Example 1.1.2. Consider

$$F(x,y) = x^2 - y^2 = c.$$

When $c \neq 0$ the solution set consists of two hyperbolas. They'll be separated by the y-axis when $c > 0$ and by the x-axis when $c < 0$. When $c = 0$ the solution set consists of the two lines $y = \pm x$. A tangent direction is given by $(2y, 2x)$, which we observe only vanishes at the origin. Unlike the above example we seem to have a valid level set passing through the origin, however, it consists of two curves that pass through the point of contention.

A nicely scaled differential equation is given by

$$\begin{bmatrix} \frac{dx}{dt} \\ \frac{dy}{dt} \end{bmatrix} = \begin{bmatrix} y \\ x \end{bmatrix}$$

and the solutions are given by

$$\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} ae^t + be^{-t} \\ ae^t - be^{-t} \end{bmatrix},$$

where a, b can be adjusted according to the initial values. There are five separate solutions that together give us the solutions to $x^2 - y^2 = 0$. We get the origin when $a = 0$, $b = 0$. The two parts of $y = x$ when $b = 0$ with the part in the first quadrant when $a > 0$ and in the third quadrant when $a < 0$. The two parts of $y = -x$ similarly come from $a = 0$.

Example 1.1.3. Consider the second order equation

$$\frac{d^2\mathbf{q}}{dt^2} = 0.$$

The solutions are straight lines $\mathbf{q}(t) = \mathbf{q}_0 + \mathbf{v}_0(t - t_0)$.

The next two examples show that scaling second order equations can, in contrast to first order equations, change the solutions drastically.

Example 1.1.4. The first example is given by the harmonic oscillator

$$\frac{d^2\mathbf{q}}{dt^2} = -\mathbf{q}.$$

This is easy to solve if we look at each coordinate separately. The solutions are:

$$\mathbf{q}(t) = \mathbf{q}_0 \cos(t - t_0) + \mathbf{v}_0 \sin(t - t_0).$$

Example 1.1.5. A far more subtle problem is Newton's inverse square law:

$$\frac{d^2\mathbf{q}}{dt^2} = -\frac{\mathbf{q}}{|\mathbf{q}|^3} = -\frac{1}{|\mathbf{q}|^2}\frac{\mathbf{q}}{|\mathbf{q}|}.$$

Newton showed that the solutions are conic sections (intersections of planes and a cone) and can be lines, circles, ellipses, parabolas, or hyperbolas.

Example 1.1.6. Finally we mention a less well known ancient example. This is the *conchoid (conch-like) of Nicomedes*. It is given by a quartic (degree 4) equation:

$$\left(x^2+y^2\right)(y-b)^2 - R^2y^2 = 0.$$

(It is not clear why this curve looks like a conch.) Descriptively it consists of two curves that are given as points (x, y) whose distance along radial lines to the line $y = b$ is R. The radial line is simply the line that passes through the origin and (x, y). So we are measuring the distance from (x, y) to the intersection of this radial line with the line $y = b$. As that intersection is $\left(\frac{x}{y}b, b\right)$ the condition is

$$\left(x - \frac{x}{y}b\right)^2 + (y-b)^2 = R^2,$$

which after multiplying both sides by y^2 easily reduces to the above equation.

The two parts of the curve correspond to points that are either above or below $y = b$. Note that no point on $y = b$ solves the equation as long as $b \neq 0$.

A simpler formula appears if we use polar coordinates. The line $y = b$ is described as

$$(x, y) = (b\cot\theta, b) = \frac{b}{\sin\theta}(\cos\theta, \sin\theta)$$

and the point (x, y) by

$$(x, y) = \left(\frac{b}{\sin\theta} \pm R\right)(\cos\theta, \sin\theta).$$

This gives us a natural parametrization of these curves.

Another parametrization is obtained if we intersect the curve with the lines $y = tx$ and use t as the parameter. This corresponds to $t = \tan\theta$ in polar coordinates. Thus we obtain the parameterized form

$$(x, y) = \left(\frac{b}{t} \pm \frac{R}{\sqrt{1+t^2}}\right)(1, t).$$

As we have seen, what we consider the same curve might have several different parametrizations.

Definition 1.1.7. Two parametrized curves $\mathbf{q}(t)$ and $\mathbf{q}^*(t^*)$ are *reparametrizations* of each other if it is possible to write $t = t(t^*)$ as a function of t^* and $t^* = t^*(t)$ such that

$$\mathbf{q}(t) = \mathbf{q}^*(t^*(t)) \text{ and } \mathbf{q}(t(t^*)) = \mathbf{q}^*(t^*).$$

If both of the functions $t(t^*)$ and $t^*(t)$ are differentiable, then it follows from the chain rule that

$$\frac{dt}{dt^*}\frac{dt^*}{dt} = 1.$$

In particular, these derivatives never vanish and have the same sign. We shall almost exclusively consider such reparametrizations. In fact we shall usually assume that these derivatives are positive so that the the direction of the curve is preserved under the reparametrization.

LEMMA 1.1.8. *If* $\mathbf{q}^*(t^*) = \mathbf{q}(t(t^*))$ *where* $t(t^*)$ *is differentiable with positive derivative, then* $\mathbf{q}^*$ *is a reparametrization of* $\mathbf{q}$.

PROOF. The missing piece in the definition of reparametrization is to show that we can also write t^* as a differentiable function of t. However, by assumption $\frac{dt}{dt^*} > 0$ so the function $t(t^*)$ is strictly increasing. This means that for a given value of the function there is at most one point in the domain yielding this value (horizontal line test). This shows that we can find the inverse function $t^*(t)$. Graphically, simply take the graph of $t(t^*)$ and consider its mirror image reflected in the diagonal line $t = t^*$. This function is also differentiable with derivative at $t = t_0$ is given by

$$\frac{1}{\frac{dt}{dt^*}(t^*(t_0))}.$$

□

It is generally too cumbersome to use two names for curves that are reparametrizations of each other. Thus we shall simply write $\mathbf{q}(t^*)$ for a reparametrization of $\mathbf{q}(t)$ with the meaning being that

$$\mathbf{q}(t) = \mathbf{q}(t^*(t)) \text{ and } \mathbf{q}(t(t^*)) = \mathbf{q}(t^*).$$

With that in mind we shall always think of two curves as being the same if they are reparametrizations of each other.

Finally we define two concepts that are easy to understand but not so easy to define rigorously.

DEFINITION 1.1.9. We say that a curve $\mathbf{q} : I \to \mathbb{R}^k$ is *simple* if I is an open interval; the curve is regular; and for each point q on the curve there is an open set $q \in O \subset \mathbb{R}^k$ such that $\{t \in I \mid \mathbf{q}(t) \in O\}$ is a union of intervals I_i and the curves $\mathbf{q} : I_i \to \mathbb{R}^k$ are all reparametrizations of each other.

REMARK 1.1.10. The special case where $\mathbf{q}(t_1) \neq \mathbf{q}(t_2)$ whenever $t_1 \neq t_2$ is quite easy to check and understand. In effect the curve must be regular and for each $q \in \mathbf{q}(I)$ there is an open set $q \in O \subset \mathbb{R}^k$ such that $\{t \in I \mid \mathbf{q}(t) \in O\}$ is an open interval. The situation when the curve intersects itself is somewhat delicate as it is not obvious how to check that two curves are reparametrizations of each other. The next proposition gives the simplest way of checking that a curve is simple.

PROPOSITION 1.1.11. *If a curve* $\mathbf{q}$ *is a solution to a first order equation* $\mathbf{v} = F(\mathbf{q})$*, then it is simple. In particular, if a curve is part of a regular level set for a function, then it is simple.*

PROOF. This is a consequence of uniqueness of solutions given initial values and the fact that the velocity only depends on position and not time. Assume $\mathbf{q}(t_1) = \mathbf{q}(t_2)$, then the two curves $\mathbf{q}_i(t) = \mathbf{q}(t + t_i)$ are both solutions to the first order equation and they have the same initial value at $t = 0$. Thus they are equal and clearly reparametrizations of each other. □

EXAMPLE 1.1.12. A circle $(\cos t, \sin t)$ is clearly simple no matter what interval we use for a domain.

DEFINITION 1.1.13. We say that a curve $\mathbf{q} : I \to \mathbb{R}^k$ is *closed* if there is an interval $[a, b] \subset I$ such that $\mathbf{q}(a) = \mathbf{q}(b)$ and $\mathbf{q}(I) = \mathbf{q}([a, b])$.

EXAMPLE 1.1.14. The figure ∞ is an example of a curve that is closed, but not simple. It can be described by an equation

$$\left(1 - x^2\right) x^2 = y^2.$$

Note that as the right hand side is non-negative it follows that $x^2 \leq 1$. When $x = -1, 0, 1$ we get that $y = 0$. For other values of x there are two possibilities for $y = \pm\sqrt{(1 - x^2)\, x^2}$.

Exercises.

(1) Show that the following properties for a curve are equivalent.
  (a) The curve is part of a line
  (b) All its tangent lines pass through a fixed point.
  (c) All its tangent lines are parallel.

(2) Show that lines in the plane satisfy $r \cos(\theta - \theta_0) = r_0$ in polar coordinates and characterize the two constants θ_0, r_0.

(3) Show that if we parametrize the sphere
$$x^2 + y^2 + z^2 = R^2$$
by
$$\begin{aligned} x &= R \sin\phi \cos\theta, \\ y &= R \sin\phi \sin\theta, \\ z &= R \cos\phi, \end{aligned}$$
then great circles satisfy $\tan\phi \cos(\theta - \theta_0) = \tan\phi_0$.

(4) Show that a curve $\mathbf{q}(t) : I \to \mathbb{R}^2$ lies on a line if and only if there is a vector $\mathbf{n} \in \mathbb{R}^2$ such that $\mathbf{q}(t) \cdot \mathbf{n}$ is constant.

(5) Show that for a curve $\mathbf{q}(t) : I \to \mathbb{R}^3$ the following properties are equivalent:
  (a) The curve lies in a plane.
  (b) There is a vector $\mathbf{n} \in \mathbb{R}^3$ such that $\mathbf{q}(t) \cdot \mathbf{n}$ is constant.
  (c) There is a vector $\mathbf{n} \in \mathbb{R}^3$ such that $\mathbf{v}(t) \cdot \mathbf{n} = 0$ for all t.

(6) Show that a curve $\mathbf{q}(t) : I \to \mathbb{R}^3$ lies on a line if and only if there are two linearly independent vectors $\mathbf{n}_1, \mathbf{n}_2 \in \mathbb{R}^3$ such that $\mathbf{q}(t) \cdot \mathbf{n}_1$ and $\mathbf{q}(t) \cdot \mathbf{n}_2$ are constant.

(7) Show that for a curve $\mathbf{q}(t) : I \to \mathbb{R}^n$ the following properties are equivalent.
  (a) The curve lies on a circle $(n = 2)$ or sphere $(n > 2.)$
  (b) There is a vector $\mathbf{c}$ such that $|\mathbf{q} - \mathbf{c}|$ is constant.
  (c) There is a vector $\mathbf{c}$ such that $(\mathbf{q} - \mathbf{c}) \cdot \mathbf{v} = 0$.

(8) Consider a curve of the form $\mathbf{q}(t) = r(t)(\cos t, \sin t)$ where r is a function of both $\cos t$ and $\sin t$
$$r(t) = p(\cos t, \sin t).$$
  (a) Show that this curve is closed.
  (b) Show that if $r(t) > 0$, then it is a regular and simple curve.
  (c) Let $t_1, t_2 \in [0, 2\pi)$ be distinct. Show that if $r(t_1) = r(t_2) = 0$ and $\dot{r}(t_1) \neq 0 \neq \dot{r}(t_2)$, then it is not simple.

(d) Show that if $r(t_0) = \dot{r}(t_0) = 0$, then its velocity vanishes at t_0.
(e) By adjusting a in $r(t) = 1 + a\cos t$ give examples of curves that satisfy the conditions in (b), (c), and (d).

(9) Consider a curve of the form $\mathbf{q}(t) = r(t)(1, t)$
(a) Show that if $r(t) > 0$, then it is a simple curve.
(b) Let t_1, t_2 be distinct. Show that if $r(t_1) = r(t_2) = 0$ and $\dot{r}(t_1) \neq 0 \neq \dot{r}(t_2)$, then it is not simple.
(c) Show that if $r(t_0) = \dot{r}(t_0) = 0$, then its velocity vanishes at t_0.
(d) By adjusting a in
$$r(t) = \frac{a + t^2}{1 + t^2}$$
give examples of curves that satisfy the conditions in (a), (b), and (c).

(10) Consider a curve in $\mathbb{R}^2$ whose velocity never vanishes and intersects all the radial lines from the origin at a constant angle θ_0. These are also called *loxodromes*. Determine what this curve must be if $\theta_0 = 0$ or $\frac{\pi}{2}$. Show that *logarithmic spirals*
$$\mathbf{q}(t) = ae^{bt}(\cos t, \sin t)$$
have this property.

(11) Show that the two equations
$$\begin{aligned} x^2 + y^2 + z^2 &= 4R^2, \\ (x - R)^2 + y^2 &= R^2 \end{aligned}$$
define a closed space curve that intersects itself at $x = 2R$ by showing that it can be parametrized as
$$\mathbf{q}(t) = R\left(\cos(t) + 1, \sin(t), 2\sin\left(\frac{t}{2}\right)\right).$$

(12) The *cissoid (ivy-like) of Diocles* is given by the equation
$$x\left(x^2 + y^2\right) = 2Ry^2.$$
(It is not clear why this curve looks like ivy.)
(a) Show that this can always be parametrized by y, but that this parametrization is not smooth at $y = 0$.
(b) Show that if $y = tx$, then we obtain a parametrization
$$(x, y) = \frac{2Rt^2}{1 + t^2}(1, t).$$
(c) Show that in polar coordinates
$$r = 2R\left(\frac{1}{\cos\theta} - \cos\theta\right).$$

(13) The *folium (leaf) of Descartes* is given by the equation
$$x^3 + y^3 - 3Rxy = 0.$$
In this case the curve really does describe a leaf in the first quadrant.
(a) Show that it can not be parameterized by x or y near the origin.

(b) Show that if $y = tx$, then we obtain a parametrization

$$(x, y) = \frac{3Rt}{1+t^3}(1, t)$$

that is valid for $t \neq -1$. What happens when $t = -1$?

(c) Show that in polar coordinates we have

$$r = \frac{3R\sin\theta\cos\theta}{\sin^3\theta + \cos^3\theta}.$$

(14) Given two planar curves $\mathbf{q}_1$ and $\mathbf{q}_2$ we can construct a *cissoid* $\mathbf{q}$ as follows: Assume that the line $y = tx$ intersects the curves in $\mathbf{q}_1 = (x_1(t), tx_1(t))$ and $\mathbf{q}_2 = (x_2(t), tx_2(t))$, then define $\mathbf{q}(t) = (x(t), tx(t))$ so that $|\mathbf{q}(t)| = |\mathbf{q}_1(t) - \mathbf{q}_2(t)|$.

(a) Show that $x(t) = \pm(x_1(t) - x_2(t))$.

(b) Show that the conchoid of Nicomedes is a cissoid.

(c) Show that the cissoid of Diocles is a cissoid.

(d) Show that the folium of Descartes is a cissoid.

(15) Let $\mathbf{q}$ be a cissoid where $\mathbf{q}_1$ is the circle of radius R centered at $(R, 0)$ and $\mathbf{q}_2$ a vertical line $x = b$.

(a) Show that when $b = 2R$ we obtain the cissoid of Diocles

$$x\left(x^2 + y^2\right) = 2Ry^2.$$

(b) Show that when $b = \frac{R}{2}$ we obtain the *trisectrix (trisector) of Maclaurin*

$$2x\left(x^2 + y^2\right) = -R\left(3x^2 - y^2\right).$$

(c) Show that when $b = R$ we obtain a *strophoid*

$$y^2(R - x) = x^2(x + R).$$

(d) Show that the change of coordinates $x = u + v$, $y = \sqrt{3}(u - v)$ turns the trisectrix of Maclaurin into Descartes folium.

1.2. Arclength and Linear Motion

The *arclength* is the distance traveled along the curve. One way of measuring the arclength geometrically is by imagining the curve as a thread that can be stretched out and measured. This however doesn't really help in formulating how it should be measured mathematically. Archimedes succeeded in understanding the arclength length of circles. The idea of measuring the length of general curves is relatively recent going back only to about 1600. Newton was the first to give the general definition that we shall use below. As we shall quickly discover, it is generally impossible to calculate the arclength of a curve as it involves finding anti-derivatives of fairly complicated functions.

From a dynamical perspective the change in arclength measures how fast the motion is along the curve. Specifically, if there is no change in arclength, then the curve is stationary, i.e., you stopped. More precisely, if the distance traveled is denoted by s (we can't use d for distance as it is used for differentiation), then the relative change with respect to the general parameter is the speed

$$\frac{ds}{dt} = \left|\frac{d\mathbf{q}}{dt}\right| = |\mathbf{v}|.$$

This means that s is the anti-derivative of speed and is defined up to an additive constant. The constant is determined by where we start measuring from. This means that we should define the *length of a curve* on $[a,b]$ as follows

$$L(\mathbf{q})_a^b = \int_a^b |\mathbf{v}|\, dt = s(b) - s(a).$$

Using substitution this is easily shown to be independent of the parameter t as long as the reparametrization is in the same direction. One also easily checks that a curve on $[a,b]$ is stationary if and only if its speed vanishes on $[a,b]$. We usually suppress the interval and instead simply write $L(\mathbf{q})$.

EXAMPLE 1.2.1. If $\mathbf{q}(t) = \mathbf{q}_0 + \mathbf{v}_0 t$ is a straight line, then its speed is constant $|\mathbf{v}_0|$ and so the arclength over an interval $[a,b]$ is $|\mathbf{v}_0|(b-a)$.

EXAMPLE 1.2.2. If $\mathbf{q}(t) = R(\cos t, \sin t) + \mathbf{c}$ is a circle of radius R centered at $\mathbf{c}$, then the speed is also constant R and so again it becomes easy to calculate the arclength.

EXAMPLE 1.2.3. Consider the hyperbola $x^2 - y^2 = 1$. It consists of two components separated by the y-axis. The component with $x > 0$ can be parametrized using hyperbolic functions $\mathbf{q}(t) = (\cosh t, \sinh t)$. The speed is

$$\frac{ds}{dt} = \sqrt{\sinh^2 t + \cosh^2 t} = \sqrt{2\sinh^2 t + 1} = \sqrt{\cosh 2t}$$

While this is both a fairly simple curve and a not terribly difficult expression for the speed it does not appear in any way easy to find the arclength explicitly.

DEFINITION 1.2.4. A curve is called *regular* if it is never stationary, or in other words the speed is always positive. A curve is said to be *parametrized by arclength* if its speed is alway 1. Such a parametrization is also called a *unit speed parametrization.*

LEMMA 1.2.5. *A regular curve* $\mathbf{q}(t)$ *can be reparametrized by arclength.*

PROOF. If we have a reparametrization $\mathbf{q}(s)$ of $\mathbf{q}(t)$ with $\frac{ds}{dt} > 0$ that has unit speed, then

$$\frac{d\mathbf{q}}{ds}\frac{ds}{dt} = \frac{d\mathbf{q}}{dt} = \mathbf{v}$$

so it follows that

$$\frac{ds}{dt} = \left|\frac{d\mathbf{q}}{dt}\right| = |\mathbf{v}|$$

must be the speed of $\mathbf{q}(t)$.

This tells us that we should define the reparametrization $s = s(t)$ as the anti-derivative of the speed:

$$s(t_1) = s(t_0) + \int_{t_0}^{t_1} \left|\frac{d\mathbf{q}}{dt}\right| dt.$$

It then follows that

$$\frac{ds}{dt} = \left|\frac{d\mathbf{q}}{dt}\right| > 0.$$

Thus it is also possible to find the inverse relationship $t = t(s)$ and we can define the reparametrized curve as $\mathbf{q}(s) = \mathbf{q}(s(t)) = \mathbf{q}(t)$.

This reparametrization depends on specifying an initial value $s(t_0)$ at some specific parameter t_0. For simplicity one often uses $s(0) = 0$ if that is at all reasonable. □

To see that arclength really is related to our usual concept of distance we show:

THEOREM 1.2.6. *The straight line is the shortest curve between any two points in Euclidean space.*

PROOF. We shall give two almost identical proofs. Without loss of generality assume that we have a curve $\mathbf{q}(t) : [a, b] \to \mathbb{R}^k$ where $\mathbf{q}(a) = 0$ and $\mathbf{q}(b) = p$. We wish to show that $L(\mathbf{q}) \geq |p|$. To that end select a unit vector field X which is also a gradient field $X = \nabla f$. Two natural choices are possible: For the first, simply let $f(x) = x \cdot \frac{p}{|p|}$, and for the second $f(x) = |x|$. In the first case the gradient is simply a parallel field and defined everywhere, in the second case we obtain the radial field which is not defined at the origin. When using the second field we need to restrict the domain of the curve to $[a_0, b]$ such that $\mathbf{q}(a_0) = 0$ but $\mathbf{q}(t) \neq 0$ for $t > a_0$. This is clearly possible as the set of points where $\mathbf{q}(t) = 0$ is a closed subset of $[a, b]$, so a_0 is just the maximum value where $\mathbf{q}$ vanishes.

This allows us to perform the following calculation using Cauchy-Schwarz, the chain rule, and the fundamental theorem of calculus. When we are in the second case the integrals are possibly improper at $t = a_0$, but clearly turn out to be perfectly well defined since the integrand has a continuous limit as t approaches a_0

$$\begin{aligned} L(\mathbf{q}) &= \int_{a_0}^{b} |\mathbf{v}|\, dt \\ &= \int_{a_0}^{b} |\dot{\mathbf{q}}|\, |\nabla f|\, dt \\ &\geq \int_{a_0}^{b} |\dot{\mathbf{q}} \cdot \nabla f|\, dt \\ &= \int_{a_0}^{b} \left| \frac{d(f \circ \mathbf{q})}{dt} \right| dt \\ &\geq \left| \int_{a_0}^{b} \frac{d(f \circ \mathbf{q})}{dt} dt \right| \\ &= |f(\mathbf{q}(b)) - f(\mathbf{q}(a_0))| \\ &= |f(p) - f(0)| \\ &= |f(p)| \\ &= |p|. \end{aligned}$$

We can even go backwards and check what happens when $L(\mathbf{q}) = |p|$. It appears that we must have equality in two places where we had inequality. Thus we have $\frac{d(f \circ \mathbf{q})}{dt} \geq 0$ everywhere and $\dot{\mathbf{q}}$ is proportional to ∇f everywhere. This implies that $\mathbf{q}$ is a possibly singular reparametrization of the straight line from 0 to p. □

We can also characterize lines through their velocities.

PROPOSITION 1.2.7. *A curve is a straight line if and only if all of its velocities are parallel.*

PROOF. The most general type of parametrization of a line is

$$\mathbf{q}(t) = \mathbf{q}_0 + \mathbf{v}_0 u(t)$$

where $u(t) : \mathbb{R} \to \mathbb{R}$ is a scalar valued function and $\mathbf{q}_0, \mathbf{v}_0 \in \mathbb{R}^n$ vectors. The velocity of such a curve is $\mathbf{v}_0 \frac{du}{dt}$ and so all velocities are indeed parallel.

Conversely if some general curve has the property that all velocities are parallel then we can write

$$\frac{d\mathbf{q}}{dt} = \mathbf{v} = v(t)\,\mathbf{v}_0$$

for some function $v(t) : \mathbb{R} \to \mathbb{R}$ and a fixed vector $\mathbf{v}_0 \in \mathbb{R}^n$. Then the curve it self can be found by integration

$$\mathbf{q}(t_1) = \mathbf{q}(t_0) + \mathbf{v}_0 \int_{t_0}^{t_1} v(t)\,dt.$$

So we obtain the general form of a line by letting $u(t)$ be an antiderivative of $v(t)$. $\square$

PROPOSITION 1.2.8. *The shortest distance from a point to a curve (if it exists) is realized by a line segment that is perpendicular to the curve.*

PROOF. Let $\mathbf{q} : [a, b] \to \mathbb{R}^k$ be a curve and assume that there is a $t_0 \in (a, b)$ such that

$$|\mathbf{q}(t) - p| \geq |\mathbf{q}(t_0) - p| \text{ for all } t \in [a, b].$$

This implies that

$$\frac{1}{2}|\mathbf{q}(t) - p|^2 \geq \frac{1}{2}|\mathbf{q}(t_0) - p|^2.$$

As the left hand side reaches a minimum at an interior point its derivative must vanish at t_0, i.e.,

$$(\mathbf{q}(t_0) - p) \cdot \frac{d\mathbf{q}}{dt}(t_0) = 0.$$

As the vector $\mathbf{q}(t_0) - p$ represents the segment from p to $\mathbf{q}(t_0)$ we have shown that it is perpendicular to the velocity of the curve. $\square$

With just a little more effort one can also find the shortest curves on spheres.

THEOREM 1.2.9. *The shortest curve between two points on a round sphere* $S^2(R) = \left\{q \in \mathbb{R}^3 \mid |q|^2 = R^2\right\}$ *is the shortest segment of the great circle through the two points.*

PROOF. Great circles on spheres centered at the origin are given as the intersections of the sphere with 2-dimensional planes through the origin. Note that if two points are antipodal then there are infinitely many great circles passing through them and all of the corresponding segments have length πR. If the two points are not antipodal, then there is a unique great circle between them and the shortest arc on this circle joining the points has length $< \pi R$.

Let us assume for simplicity that $R = 1$. The great circle that lies in the plane span $\{\mathbf{q}_0, \mathbf{v}_0\}$ where $\mathbf{q}_0 \perp \mathbf{v}_0$ and $|\mathbf{q}_0| = |\mathbf{v}_0| = 1$ can be parametrized as follows

$$\mathbf{q}(t) = \mathbf{q}_0 \cos t + \mathbf{v}_0 \sin t.$$

This curve passes through the point $\mathbf{q}_0 \in S^2(1)$ at $t = 0$ and has velocity $\mathbf{v}_0$ at that point. It also passes through the antipodal point $-\mathbf{q}_0$ at time $t = \pi$. Finally, it is also parametrized by arclength.

To find the great circle that passes through two points $\mathbf{q}_0, \mathbf{q}_1 \in S^2(1)$ that are not antipodal we simply select the initial velocity $\mathbf{v}_0$ to be the vector in the plane $\operatorname{span}\{\mathbf{q}_0, \mathbf{q}_1\}$ that is perpendicular to $\mathbf{q}_0$ and has length 1, i.e.,

$$\begin{aligned}
\mathbf{v}_0 &= \frac{\mathbf{q}_1 - (\mathbf{q}_1 \cdot \mathbf{q}_0)\,\mathbf{q}_0}{|\mathbf{q}_1 - (\mathbf{q}_1 \cdot \mathbf{q}_0)\,\mathbf{q}_0|} \\
&= \frac{\mathbf{q}_1 - (\mathbf{q}_1 \cdot \mathbf{q}_0)\,\mathbf{q}_0}{\sqrt{1 - (\mathbf{q}_1 \cdot \mathbf{q}_0)^2}}.
\end{aligned}$$

Then the great circle

$$\mathbf{q}(t) = \mathbf{q}_0 \cos t + \mathbf{v}_0 \sin t$$

passes through $\mathbf{q}_1$ when

$$t = \arccos(\mathbf{q}_1 \cdot \mathbf{q}_0).$$

The velocity of this great circle at $\mathbf{q}_1$ is

$$\mathbf{v}_1 = \frac{-\mathbf{q}_0 + (\mathbf{q}_0 \cdot \mathbf{q}_1)\,\mathbf{q}_1}{|-\mathbf{q}_0 + (\mathbf{q}_0 \cdot \mathbf{q}_1)\,\mathbf{q}_1|}$$

since it is the initial velocity of the great circle that starts at $\mathbf{q}_1$ and goes through $-\mathbf{q}_0$.

The goal now is to show that any curve $\mathbf{q}(t) : [0, L] \to S^2(1)$ between $\mathbf{q}_0$ and $\mathbf{q}_1$ has length $\geq \arccos(\mathbf{q}_1 \cdot \mathbf{q}_0)$. The proof of this follows the same pattern as the proof for lines. We start by assuming that $\mathbf{q}(t) \neq \mathbf{q}_0, \mathbf{q}_1$ when $t \in (0, L)$ and define

$$\mathbf{v}_1(t) = \frac{-\mathbf{q}_0 + (\mathbf{q}_0 \cdot \mathbf{q}(t))\,\mathbf{q}(t)}{|-\mathbf{q}_0 + (\mathbf{q}_0 \cdot \mathbf{q}(t))\,\mathbf{q}(t)|}.$$

Before the calculation note that since $\mathbf{q}(t)$ is a unit vector it follows that $\mathbf{q}\cdot\frac{d\mathbf{q}}{dt}=0$. With that in mind we obtain

$$\begin{aligned}
L(\mathbf{q}) &= \int_0^L |\mathbf{v}|\,dt \\
&= \int_0^L |\mathbf{v}|\,|\mathbf{v}_1(t)|\,dt \\
&\geq \int_0^L \left|\mathbf{v}_1(t)\cdot\frac{d\mathbf{q}}{dt}\right| dt \\
&= \int_0^L \left|\frac{-\mathbf{q}_0+(\mathbf{q}_0\cdot\mathbf{q}(t))\,\mathbf{q}(t)}{|-\mathbf{q}_0+(\mathbf{q}_0\cdot\mathbf{q}(t))\,\mathbf{q}(t)|}\cdot\frac{d\mathbf{q}}{dt}\right| dt \\
&= \int_0^L \left|\frac{-\mathbf{q}_0}{\sqrt{1-(\mathbf{q}_0\cdot\mathbf{q}(t))^2}}\cdot\frac{d\mathbf{q}}{dt}\right| dt \\
&= \int_0^L \left|\frac{d\arccos(\mathbf{q}_0\cdot\mathbf{q}(t))}{dt}\right| dt \\
&\geq \left|\int_0^L \frac{d\arccos(\mathbf{q}_0\cdot\mathbf{q}(t))}{dt}dt\right| \\
&= |\arccos(\mathbf{q}_0\cdot\mathbf{q}(L))-\arccos(\mathbf{q}_0\cdot\mathbf{q}(0))| \\
&= |\arccos(\mathbf{q}_0\cdot\mathbf{q}_1)-\arccos(\mathbf{q}_0\cdot\mathbf{q}_0)| \\
&= |\arccos(\mathbf{q}_0\cdot\mathbf{q}_1)| \\
&= \arccos(\mathbf{q}_0\cdot\mathbf{q}_1)
\end{aligned}$$

This proves that the segment of the great circle always has the shortest length.

In case the original curve was parametrized by arclength and also has this length we can backtrack the argument and observe that this forces $\mathbf{v}=\mathbf{v}_1$ or in other words

$$\frac{d\mathbf{q}}{dt}=\frac{-\mathbf{q}_0+(\mathbf{q}_0\cdot\mathbf{q}(t))\,\mathbf{q}(t)}{|-\mathbf{q}_0+(\mathbf{q}_0\cdot\mathbf{q}(t))\,\mathbf{q}(t)|}$$

This is a differential equation for the curve and we know that great circles solve this equation as the right hand side is the velocity of the great circle at $\mathbf{q}(t)$. So it follows from uniqueness of solutions to differential equations that any curve of minimal length is part of a great circle. □

Exercises.

(1) Compute the arclength parameter of $y=x^{\frac{3}{2}}$.

(2) Compute the arclength parameter of the parabolas $y=\sqrt{x}$ and $y=x^2$.

(3) Redefine the concept of closed and simple curves using arclength parametrization.

(4) Compute the arclength of the logarithmic spiral
$$ae^{bt}(\cos t,\sin t)$$
and explain why it is called logarithmic.

(5) Show that every regular planar curve that makes a constant angle $\theta_0>0$ with all radial lines can be reparametrized to be a *logarithmic spiral*
$$ae^{bt}(\cos t,\sin t).$$

Hint: One way of proving this is to show that all such curves satisfy a differential equation if we assume that they have unit speed.

(6) Compute the arclength of the *spiral of Archimedes*:

$$(a+bt)(\cos t, \sin t).$$

(7) Show that the parametrization of the folium of Descartes given by

$$(x,y) = \frac{3Rt}{1+t^3}(1,t)$$

is regular.

(8) Show that it is not possible to parametrize the cissoid of Diocles

$$x\left(x^2+y^2\right) = 2Ry^2$$

so that it is regular at the origin.

(9) Find the arclength parameter for the *twisted cubic*

$$\mathbf{q}(t) = \left(t, t^2, t^3\right).$$

(10) A *cycloid* is a planar curve that follows a point on a circle of radius R as it rolls along a straight line without slipping.

(a) Show that

$$\mathbf{q}(t) = tR\mathbf{e}_1 + R\mathbf{e}_2 - R(\mathbf{e}_2\cos t + \mathbf{e}_1 \sin t)$$

is a parametrization of a cycloid, when $\mathbf{e}_1, \mathbf{e}_2$ are orthonormal.

(b) Show that all cycloids can be parametrized to have the form

$$\mathbf{q}(t) = tR\mathbf{e}_1 + R\mathbf{e}_2 - R(\mathbf{e}_2\cos t + \mathbf{e}_1 \sin t) + \mathbf{q}_0$$

where $\mathbf{q}(0) = \mathbf{q}_0$.

(c) Show that any such cycloid stays on one side of the line $\mathbf{q}_0 + tR\mathbf{e}_1$ and has zero velocity cusps when it hits this line.

(d) Show that a cycloid hits the line at points that are $2\pi R$ apart.

(11) Let $\mathbf{q}(t) : I \to \mathbb{R}^2$ be a closed planar curve.

(a) Show that the curve is contained in a circle of smallest radius and that this circle is unique.

(b) Show that it either touches this circle in two antipodal points or in three points that form an acute triangle. Note that in either of these two cases it might still touch the circle in many other points as well.

(c) Show that if the radius of this circle is R, then $L(\mathbf{q}) \geq 4R$. Hint: This is clear when $\mathbf{q}$ hits the circle in antipodal points. When this does not happen there are several possible strategies. One is to use the acute triangle and show that the circumference of this triangle is $\geq 4R$. This can be shown using the law of sines that relates the side length to the opposite angle and the diameter of the circle

$$\begin{aligned} a &= 2R\sin\alpha, \\ b &= 2R\sin\beta, \\ c &= 2R\sin\gamma. \end{aligned}$$

The goal then is to show that

$$\sin\alpha + \sin\beta + \sin\gamma \geq 2,$$

when
$$\alpha, \beta, \gamma \in \left[0, \frac{\pi}{2}\right], \ \alpha + \beta + \gamma = \pi.$$

(12) (Spherical law of cosines) Consider three points $\mathbf{q}_i$, $i = 1, 2, 3$ on a unit sphere centered at the origin. Join these points by great circle segments to obtain a triangle. Let the side lengths be a_{ij} and the interior angle θ_i at $\mathbf{q}_i$.

(a) Show that
$$\cos a_{ij} = \mathbf{q}_i \cdot \mathbf{q}_j$$
and
$$\cos \theta_1 = \left(\frac{\mathbf{q}_2 - (\mathbf{q}_2 \cdot \mathbf{q}_1)\,\mathbf{q}_1}{\sqrt{1 - (\mathbf{q}_2 \cdot \mathbf{q}_1)^2}} \right) \cdot \left(\frac{\mathbf{q}_3 - (\mathbf{q}_3 \cdot \mathbf{q}_1)\,\mathbf{q}_1}{\sqrt{1 - (\mathbf{q}_3 \cdot \mathbf{q}_1)^2}} \right).$$

(b) Show that
$$\cos a_{23} = \cos a_{12} \cos a_{13} + \cos \theta_1 \sin a_{12} \sin a_{13}.$$

(c) Compare this with the Euclidean law of cosines
$$a_{23}^2 = a_{12}^2 + a_{13}^2 - 2a_{12}a_{23} \cos \theta$$
for a triangle with the same sides and conclude that $\theta_1 > \theta$.

1.3. Curvature

We saw that arclength measures how far a curve is from being stationary. Our preliminary concept of *curvature* is that it should measure how far a curve is from being a line. For a planar curve the idea used to be to find a circle that best approximates the curve at a point (just like a tangent line is the line that best approximates the curve). The radius of this circle then gives a measure of how the curve bends with larger radius implying less bending. Huygens did quite a lot to clarify this idea for fairly general curves using purely geometric considerations (no calculus) and applied it to the study involutes and evolutes. Newton seems to have been the first to take the reciprocal of this radius to create curvature as we now define it. He also generated some of the formulas in both Cartesian and polar coordinates that are still in use today.

To formalize the idea of how a curve deviates from being a line we define the *unit tangent vector* of a regular curve $\mathbf{q}(t) : [a, b] \to \mathbb{R}^k$ as the direction of the velocity:
$$\mathbf{v} = \dot{\mathbf{q}} = |\mathbf{v}|\,\mathbf{T} = \frac{ds}{dt}\mathbf{T}.$$
When the unit tangent vector $\mathbf{T} = \mathbf{v}/|\mathbf{v}|$ is stationary, then the curve is evidently a straight line. So the degree to which the unit tangent is stationary is a measure of how fast it changes and in turn how far the curve is from being a line. We let θ be the arclength parameter for $\mathbf{T}$. The relative change between the arclength parameters for the unit tangent and the curve is by definition the *curvature*
$$\kappa = \frac{d\theta}{ds}$$
and for a general parametrization
$$\kappa = \frac{dt}{ds}\frac{d\theta}{dt}.$$

We shall see that the curvature is related to the part of the acceleration that is orthogonal to the unit tangent vector. Note that $\kappa \geq 0$ as θ increases with s.

PROPOSITION 1.3.1. *A regular curve is part of a line if and only if its curvature vanishes.*

PROOF. The unit tangent of a line is clearly stationary. Conversely if the curvature vanishes, then the unit tangent is stationary. This means that when the curve is parametrized by arclength, then it will be a straight line. □

Next we show how the curvature can be calculated with general parametrizations.

PROPOSITION 1.3.2. *The curvature of a regular curve is given by*

$$\begin{aligned}\kappa &= \frac{|\mathbf{v}|\,|\mathbf{a}-(\mathbf{a}\cdot\mathbf{T})\,\mathbf{T}|}{|\mathbf{v}|^3}\\ &= \frac{\textit{area of parallelogram}\,(\mathbf{v},\mathbf{a})}{|\mathbf{v}|^3}.\end{aligned}$$

PROOF. We calculate

$$\begin{aligned}\kappa &= \frac{d\theta}{ds}\\ &= \frac{d\theta}{dt}\frac{dt}{ds}\\ &= \left|\frac{d\mathbf{T}}{dt}\right||\mathbf{v}|^{-1}\\ &= \left|\frac{d}{dt}\frac{\mathbf{v}}{|\mathbf{v}|}\right||\mathbf{v}|^{-1}\\ &= \left|\frac{\mathbf{a}}{|\mathbf{v}|^2}-\frac{\mathbf{v}\,(\mathbf{a}\cdot\mathbf{v})}{|\mathbf{v}|^4}\right|\\ &= \frac{1}{|\mathbf{v}|^2}\left|\mathbf{a}-\frac{(\mathbf{a}\cdot\mathbf{v})\,\mathbf{v}}{|\mathbf{v}|^2}\right|\\ &= \frac{1}{|\mathbf{v}|^2}\,|\mathbf{a}-(\mathbf{a}\cdot\mathbf{T})\,\mathbf{T}|\,.\end{aligned}$$

The area of the parallelogram spanned by $\mathbf{v}$ and $\mathbf{a}$ is given by the product of the length of the base represented by $\mathbf{v}$ and the height represented by the component of $\mathbf{a}$ that is normal to the base, i.e., $\mathbf{a}-(\mathbf{a}\cdot\mathbf{T})\,\mathbf{T}$. Thus we obtain the formula

$$\begin{aligned}\kappa &= \frac{|\mathbf{v}|\,|\mathbf{a}-(\mathbf{a}\cdot\mathbf{T})\,\mathbf{T}|}{|\mathbf{v}|^3}\\ &= \frac{\text{area of parallelogram}\,(\mathbf{v},\mathbf{a})}{|\mathbf{v}|^3}.\end{aligned}$$

□

REMARK 1.3.3. For 3-dimensional curves the curvature can also be written as

$$\kappa = \frac{|\mathbf{v}\times\mathbf{a}|}{|\mathbf{v}|^3}.$$

Further note that when the unit tangent vector is regular it too has a unit tangent vector called the *normal* $\mathbf{N}$ to the curve. Specifically

$$\frac{d\mathbf{T}}{d\theta} = \mathbf{N}.$$

The unit normal is the unit tangent to the unit tangent. This vector is in fact perpendicular to $\mathbf{T}$ as

$$0 = \frac{d\,|\mathbf{T}|^2}{d\theta} = 2\left(\mathbf{T}\cdot\frac{d\mathbf{T}}{d\theta}\right) = 2\,(\mathbf{T}\cdot\mathbf{N})\,.$$

This normal vector is also called the *principal normal* for $\mathbf{q}$ when the curve is a space curve as there are also other vectors that are normal to the curve in that case. The line through a point on a curve in the direction of the principal normal is called the *principal normal line*.

In terms of the arclength parameter s for $\mathbf{q}$ we obtain

$$\frac{d\mathbf{T}}{ds} = \frac{d\theta}{ds}\frac{d\mathbf{T}}{d\theta} = \kappa\mathbf{N}$$

and

$$\kappa = \frac{d\mathbf{T}}{ds}\cdot\mathbf{N} = -\mathbf{T}\cdot\frac{d\mathbf{N}}{ds},$$

where the last equality follows from

$$0 = \frac{d\mathbf{T}\cdot\mathbf{N}}{ds} = \frac{d\mathbf{T}}{ds}\cdot\mathbf{N} + \mathbf{T}\cdot\frac{d\mathbf{N}}{ds}.$$

Proposition 1.3.4. *For a regular curve we have*

$$\mathbf{v} = (\mathbf{v}\cdot\mathbf{T})\,\mathbf{T} = |\mathbf{v}|\,\mathbf{T},$$

$$\mathbf{a} = (\mathbf{a}\cdot\mathbf{T})\,\mathbf{T} + (\mathbf{a}\cdot\mathbf{N})\,\mathbf{N} = (\mathbf{a}\cdot\mathbf{T})\,\mathbf{T} + \kappa\,|\mathbf{v}|^2\,\mathbf{N},$$

and

$$\boldsymbol{N} = \frac{\mathbf{a}-(\mathbf{a}\cdot\mathbf{T})\,\mathbf{T}}{|\mathbf{a}-(\mathbf{a}\cdot\mathbf{T})\,\mathbf{T}|}.$$

Thus the unit normal is the direction of the part of the acceleration that is perpendicular to the velocity.

Proof. The first formula follows directly from the definition of $\mathbf{T}$. For the second we note that

$$\begin{aligned}
\mathbf{a} &= \frac{d\mathbf{v}}{dt} \\
&= \frac{d\theta}{dt}\frac{d\mathbf{v}}{d\theta} \\
&= \frac{d\theta}{dt}\left(\frac{d\,|\mathbf{v}|}{d\theta}\mathbf{T} + |\mathbf{v}|\,\frac{d\mathbf{T}}{d\theta}\right) \\
&= \frac{d\theta}{dt}\left(\frac{d\,|\mathbf{v}|}{d\theta}\mathbf{T} + |\mathbf{v}|\,\mathbf{N}\right) \\
&= \frac{d\,|\mathbf{v}|}{dt}\mathbf{T} + \frac{d\theta}{dt}\,|\mathbf{v}|\,\mathbf{N}.
\end{aligned}$$

This shows that $\mathbf{a}$ is a linear combination of $\mathbf{T},\mathbf{N}$. It also shows that

$$\begin{aligned}\mathbf{a}\cdot\mathbf{N} &= \frac{d\theta}{dt}\left|\mathbf{v}\right| \\ &= \frac{d\theta}{ds}\frac{ds}{dt}\left|\mathbf{v}\right| \\ &= \kappa\left|\mathbf{v}\right|^{2}.\end{aligned}$$

So we obtain the second equation. The last formula then follows from the fact that $\mathbf{N}$ is the direction of the normal component of the acceleration. □

DEFINITION 1.3.5. An *involute* of a curve $\mathbf{q}(t)$ is a curve $\mathbf{q}^{*}(t)$ that lies on the corresponding tangent lines to $\mathbf{q}(t)$ and intersects these tangent lines orthogonally.

We can always construct involutes to regular curves. First of all

$$\mathbf{q}^{*}(t)=\mathbf{q}(t)+u(t)\,\mathbf{T}(t)$$

as it is forced to lie on the tangent lines to $\mathbf{q}$. Secondly the velocity $\mathbf{v}^{*}$ must be parallel to $\mathbf{N}$. Since

$$\frac{d\mathbf{q}^{*}}{dt}=\frac{d\mathbf{q}}{dt}+\frac{du}{dt}\mathbf{T}+u\kappa\frac{ds}{dt}\mathbf{N}$$

this forces us to select u so that

$$\frac{du}{dt}=-\frac{ds}{dt}.$$

Thus

$$\mathbf{q}^{*}(t)=\mathbf{q}(t)-s(t)\,\mathbf{T}(t),$$

where s is any arclength parametrization of $\mathbf{q}$. Note that s is only determined up to a constant so we always get infinitely many involutes to a given curve.

EXAMPLE 1.3.6. If we strip a length of masking tape glued to a curve keeping it taut while doing so, then the end of the tape will trace an involute.

Assume the original curve is unit speed $\mathbf{q}(s)$. The process of stripping the tape from the curve forces the endpoint of the tape to have an equation of the form

$$\mathbf{q}^{*}(s)=\mathbf{q}(s)+u(s)\,\mathbf{T}(s)$$

since for each value of s the tape has two parts, the first being the curve up to $\mathbf{q}(s)$ and the second the line segment from $\mathbf{q}(s)$ to $\mathbf{q}(s)+u(s)\,\mathbf{T}(s)$. The length of this is up to a constant given by

$$s+u(s).$$

As the piece of tape doesn't change length this is constant. This shows that $u=c-s$ for some constant c and thus that the curve is an involute.

EXAMPLE 1.3.7. Huygens designed pendulums using involutes. His idea was to take two planar convex curves that are mirror images of each other in the y-axis and are tangent to the y-axis with the unit tangent at this cusp pointing downwards. Suspend a string from this cusp point of length L with a metal disc attached at the bottom end to keep the string taut. Now displace the metal disc horizontally and release it. Gravity will then force the disc to swing back and forth. The trajectory will depend on the shape of the chosen convex curve and will be an involute of that curve.

Huygens was interested in creating a pendulum with the property that its period does not depend on the amplitude of the swing. Thus the period will remain constant even though the pendulum slows down with time. A curve with this property is called *tautochronic* and Huygens showed that it has to be a cycloid that looks like

$$R(\sin t, \cos t) + R(t, 0).$$

The involute is also a cycloid (see also exercises below).

EXAMPLE 1.3.8. Consider the unit circle $\mathbf{q}(s) = (\cos s, \sin s)$. This parametrization is by arclength so we obtain the involutes

$$\mathbf{q}^*(s) = (\cos s, \sin s) + (c - s)(-\sin s, \cos s).$$

In polar coordinates we have

$$r(s) = |\mathbf{q}^*(s)| = \sqrt{1 + (c - s)^2}.$$

When $c = 0$ we see that r increases with s and that the involute looks like a spiral.

DEFINITION 1.3.9. An *evolute* of a curve $\mathbf{q}(t)$ is a curve $\mathbf{q}^*(t)$ such that the tangent lines to $\mathbf{q}^*$ are orthogonal to $\mathbf{q}$ at corresponding values of t.

REMARK 1.3.10. Note that if $\mathbf{q}^*$ is an involute to $\mathbf{q}$, then conversely $\mathbf{q}$ is an evolute to $\mathbf{q}^*$. It is however quite complicated to construct evolutes in general, but, as we shall see, there are formulas for both planar and space curves.

Evolutes must look like

$$\mathbf{q}^*(t) = \mathbf{q}(t) + \mathbf{V}(t),$$

where $\mathbf{V} \cdot \mathbf{T} = 0$ and also have the property that

$$0 = \mathbf{T} \cdot \frac{d\mathbf{q}^*}{dt} = \mathbf{T} \cdot \left(\frac{d\mathbf{q}}{dt} + \frac{d\mathbf{V}}{dt}\right),$$

which is equivalent to

$$\mathbf{T} \cdot \frac{d\mathbf{V}}{dt} = -\frac{ds}{dt}.$$

Exercises.

(1) Show that a curve is part of a line if all its tangent lines pass through a fixed point.
(2) Show that for all vectors $v, w \in \mathbb{R}^n$ we have
$$\begin{aligned}\text{area of parallelogram}(v, w) &= \sqrt{|v|^2 |w|^2 - (v \cdot w)^2} \\ &= |v| |w| \sin\measuredangle(v, w).\end{aligned}$$
(3) Show that the curvature of a planar circle of radius R is $\frac{1}{R}$ by parametrizing this curve in the following way $\mathbf{q}(t) = R(\cos t, \sin t) + \mathbf{c}$.
(4) Find the curvature for the twisted cubic
$$\mathbf{q}(t) = \left(t, t^2, t^3\right).$$
(5) Compute the curvature of the *tractrix*
$$\mathbf{q}(t) = \left(\frac{1}{\cosh t}, t - \frac{\sinh t}{\cosh t}\right).$$
(6) Let $\mathbf{q}(t)$ be a regular curve with positive curvature. Define two vector fields whose integral curves are involutes to $\mathbf{q}$.

(7) (Huygens, 1673) Consider the cycloid

$$\mathbf{q}(t) = R(t + \sin t, 1 + \cos t)$$

(see also section 1.2 exercise 10 and note that this cycloid comes with a different parametrization and initial position).

(a) Show that the speed satisfies

$$\left|\frac{d\mathbf{q}}{dt}\right|^2 = 2R^2 \frac{\sin^2 t}{1 - \cos t}.$$

(b) Show that the arclength parameter s with initial value $s(0) = 0$ satisfies

$$s^2 = 8R^2(1 - \cos t).$$

(c) Show that the curvature satisfies

$$\kappa^2 = \frac{1}{8R^2(1 + \cos t)}.$$

(d) Show that for a general cycloid it is always possible to find $a \in \mathbb{R}$ such that

$$(s - a)^2 + \frac{1}{\kappa^2} = 16R^2.$$

(e) Show $a = 4R$ for the cycloid

$$\mathbf{q}(t) = R(t - \sin t, 1 - \cos t)$$

if we assume that $s(0) = 0$.

(8) If a curve in $\mathbb{R}^2$ is given as a graph $y = f(x)$ show that the curvature is given by

$$\kappa = \frac{|f''|}{\left(1 + (f')^2\right)^{\frac{3}{2}}}.$$

(9) Let $\mathbf{q}(t) = r(t)(\cos t, \sin t)$. Show that the speed is given by

$$\left(\frac{ds}{dt}\right)^2 = \left(\frac{dr}{dt}\right)^2 + r^2$$

and the curvature

$$\kappa = \frac{\left|2\left(\frac{dr}{dt}\right)^2 + r^2 - r\frac{d^2r}{dt^2}\right|}{\left(\left(\frac{dr}{dt}\right)^2 + r^2\right)^{\frac{3}{2}}}.$$

(10) Let $\mathbf{q}(t) : I \to \mathbb{R}^3$ be a regular curve with speed $\frac{ds}{dt} = \left|\frac{d\mathbf{q}}{dt}\right|$, where s is the arclength parameter. Prove that

$$\kappa = \frac{\sqrt{\frac{d^2\mathbf{q}}{dt^2} \cdot \frac{d^2\mathbf{q}}{dt^2} - \left(\frac{d^2s}{dt^2}\right)^2}}{\left(\frac{ds}{dt}\right)^2}.$$

(11) Compute the curvature of the logarithmic spiral

$$ae^{bt}(\cos t, \sin t).$$

(12) Compute the curvature of the spiral of Archimedes:

$$(a + bt)(\cos t, \sin t).$$

(13) Show that the involute to a straight line is a point.
(14) Show that a planar circle has its center as an evolute.
(15) The circular helix is given by

$$\mathbf{q}(t) = R(\cos t, \sin t, 0) + h(0,0,t).$$

Reparametrize this curve to be unit speed and show that its involutes lie in planes given by $z = c$ for some constant c.
(16) Let $\mathbf{q}(s)$ be a planar unit speed curve with positive curvature. Show that the curvature of the involute

$$\mathbf{q}^*(s) = \mathbf{q}(s) + (L - s)\mathbf{T}(s)$$

satisfies

$$\kappa^* = \frac{1}{|L - s|}$$

and compute the evolute of $\mathbf{q}^*$.
(17) For a regular curve $\mathbf{q}(t) : I \to \mathbb{R}^n$ we say that a field $\mathbf{X}$ is *parallel* along $\mathbf{q}$ if $\mathbf{X} \cdot \mathbf{T} = 0$ and $\frac{d\mathbf{X}}{dt}$ is parallel to $\mathbf{T}$, i.e.,

$$\frac{d\mathbf{X}}{dt} = \left(\frac{d\mathbf{X}}{dt} \cdot \mathbf{T}\right)\mathbf{T} = -\left(\frac{d\mathbf{T}}{dt} \cdot \mathbf{X}\right)\mathbf{T}.$$

(a) Show that for a fixed t_0 and $\mathbf{X}(t_0) \perp \mathbf{T}(s_0)$ there is a unique parallel field $\mathbf{X}$ that has the value $\mathbf{X}(t_0)$ at t_0.
(b) Show that if $\mathbf{X}_1$ and $\mathbf{X}_2$ are both parallel along $\mathbf{q}$, then $\mathbf{X}_1 \cdot \mathbf{X}_2$ is constant.
(c) A *Bishop frame* consists of an orthonormal frame $\mathbf{T}, \mathbf{N}_1, \mathbf{N}_2, ..., \mathbf{N}_{n-1}$ along the curve so that all $\mathbf{N}_i$ are parallel along $\mathbf{q}$. For such a frame show that

$$\frac{d}{dt}\begin{bmatrix} \mathbf{T} & \mathbf{N}_1 & \mathbf{N}_2 & \cdots & \mathbf{N}_{n-1} \end{bmatrix} = \frac{ds}{dt}\begin{bmatrix} \mathbf{T} & \mathbf{N}_1 & \mathbf{N}_2 & \cdots & \mathbf{N}_{n-1} \end{bmatrix}\begin{bmatrix} 0 & \kappa_1 & \kappa_2 & \cdots & \kappa_{n-1} \\ -\kappa_1 & 0 & 0 & \cdots & 0 \\ -\kappa_2 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ -\kappa_{n-1} & 0 & 0 & \cdots & 0 \end{bmatrix}.$$

Note that such frames always exist, even when the curve doesn't have positive curvature everywhere.
(d) Show further for such a frame that

$$\kappa^2 = \kappa_1^2 + \kappa_2^2 + \cdots + \kappa_{n-1}^2.$$

The collection $(\kappa_1, \kappa_2, ..., \kappa_{n-1})$ can in turn be thought of as a curve going into $\mathbb{R}^{n-1}$ and be investigated for higher order behavior of $\mathbf{q}$. When $\kappa > 0$ one generally divides this curve by κ and considers the spherical curve into S^{n-2}.
(e) Give an example of a closed space curve where the parallel fields don't close up.

1.4. Integral Curves

In this section we shall try to calculate the curvature of curves that are solutions to differential equations. As it is rarely possible to find explicit formulas for such solutions the goal is to use the fact that we know they exist and calculate their curvatures using only the data that the differential equation gives us. Recall that curves that are solutions to equations can also be considered as solutions to differential equations.

We start by considering a solution to a first order equation

$$\mathbf{v} = \frac{d\mathbf{q}}{dt} = F\left(\mathbf{q}\left(t\right)\right).$$

The first observation is that the speed is given by

$$\left|\mathbf{v}\right| = \left|\frac{d\mathbf{q}}{dt}\right| = \left|F\left(\mathbf{q}\left(t\right)\right)\right|.$$

The acceleration is computed using the chain rule

$$\mathbf{a} = \frac{d\mathbf{v}}{dt} = \frac{dF\left(\mathbf{q}\left(t\right)\right)}{dt} = DF\left(\frac{d\mathbf{q}}{dt}\right) = DF\left(F\left(\mathbf{q}\left(t\right)\right)\right).$$

The curvature is then given by

$$\begin{aligned}\kappa^2\left(t\right) &= \frac{\left|\mathbf{v}\right|^2\left|\mathbf{a}\right|^2 - \left(\mathbf{v}\cdot\mathbf{a}\right)^2}{\left|\mathbf{v}\right|^6}\\ &= \frac{\left|F\left(\mathbf{q}\left(t\right)\right)\right|^2\left|DF\left(F\left(\mathbf{q}\left(t\right)\right)\right)\right|^2 - \left(F\left(\mathbf{q}\left(t\right)\right)\cdot DF\left(F\left(\mathbf{q}\left(t\right)\right)\right)\right)^2}{\left|F\left(\mathbf{q}\left(t\right)\right)\right|^6}.\end{aligned}$$

So if we wish to calculate the curvature for a solution that passes through a fixed point q_0 at time $t = t_0$, then we have

$$\kappa^2\left(t_0\right) = \frac{\left|F\left(q_0\right)\right|^2\left|DF\left(F\left(q_0\right)\right)\right|^2 - \left(F\left(q_0\right)\cdot DF\left(F\left(q_0\right)\right)\right)^2}{\left|F\left(q_0\right)\right|^6}.$$

This is a formula that does not require us to solve the equation.

For a second order equation

$$\mathbf{a} = \frac{d^2\mathbf{q}}{dt^2} = F\left(\mathbf{q}\left(t\right), \frac{d\mathbf{q}}{dt}\right) = F\left(\mathbf{q}\left(t\right), \mathbf{v}\left(t\right)\right)$$

there isn't much to compute as we now have to be given both position q_0 and velocity v_0 at time t_0. The curvature is given by

$$\begin{aligned}\kappa^2\left(t\right) &= \frac{\left|\mathbf{v}\right|^2\left|\mathbf{a}\right|^2 - \left(\mathbf{v}\cdot\mathbf{a}\right)^2}{\left|\mathbf{v}\right|^6}\\ &= \frac{\left|v_0\right|^2\left|F\left(q_0, v_0\right)\right|^2 - \left(v_0\cdot F\left(q_0, v_0\right)\right)^2}{\left|v_0\right|^6}.\end{aligned}$$

However, note that we can also calculate the change in speed by observing that

$$\frac{d\left|\mathbf{v}\right|^2}{dt} = 2\mathbf{v}\cdot\mathbf{a} = 2\mathbf{v}\cdot F\left(\mathbf{q}, \mathbf{v}\right).$$

A few examples will hopefully clarify this a little better.

EXAMPLE 1.4.1. First an example were we know that the solutions are circles.

$$F(x,y) = (-y,x)$$

and

$$\begin{aligned} DF(F(x,y)) &= \begin{bmatrix} \frac{\partial(-y)}{\partial x} & \frac{\partial(-y)}{\partial y} \\ \frac{\partial(x)}{\partial x} & \frac{\partial(x)}{\partial y} \end{bmatrix} \begin{bmatrix} -y \\ x \end{bmatrix} \\ &= \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} -y \\ x \end{bmatrix} \\ &= \begin{bmatrix} -x \\ -y \end{bmatrix}. \end{aligned}$$

So at $q_0 = (x_0, y_0)$ we have

$$\kappa^2 = \frac{\left(x_0^2 + y_0^2\right)^2 - (x_0 y_0 - x_0 y_0)^2}{\left(x_0^2 + y_0^2\right)^3} = \frac{1}{|q_0|^2},$$

which agrees with our knowledge that the curvature is the reciprocal of the radius.

EXAMPLE 1.4.2. Next we look at the second order equation

$$\mathbf{a} = -\frac{\mathbf{q}}{|\mathbf{q}|^3}.$$

The curvature is

$$\begin{aligned} \kappa^2 &= \frac{|\mathbf{v}|^2 \left| -\frac{\mathbf{q}}{|\mathbf{q}|^3} \right|^2 - \left(-\frac{\mathbf{q}}{|\mathbf{q}|^3} \cdot \mathbf{v} \right)^2}{|\mathbf{v}|^6} \\ &= \frac{|\mathbf{q}|^2 |\mathbf{v}|^2 - (\mathbf{q} \cdot \mathbf{v})^2}{|\mathbf{q}|^6 |\mathbf{v}|^6}. \end{aligned}$$

Yielding

$$\kappa = \frac{\text{area of parallelogram}\,(\mathbf{q}, \mathbf{v})}{|\mathbf{q}|^3 |\mathbf{v}|^3}.$$

So the curvature vanishes when the velocity is radial (proportional to position), this conforms with the fact that radial lines are solutions to this equation. Otherwise all other solutions must have nowhere vanishing curvature. In general the numerator is constant along solutions as

$$\begin{aligned} \frac{d}{dt}\left(|\mathbf{q}|^2 |\mathbf{v}|^2 - (\mathbf{q} \cdot \mathbf{v})^2 \right) &= 2\mathbf{q} \cdot \mathbf{v} \, |\mathbf{v}|^2 + 2 |\mathbf{q}|^2 \mathbf{v} \cdot \mathbf{a} \\ &\quad -2\mathbf{q} \cdot \mathbf{v} \left(|\mathbf{v}|^2 + \mathbf{q} \cdot \mathbf{a} \right) \\ &= 2\mathbf{q} \cdot \mathbf{v} \, |\mathbf{v}|^2 - 2 \frac{1}{|\mathbf{q}|} \mathbf{v} \cdot \mathbf{q} \\ &\quad -2\mathbf{q} \cdot \mathbf{v} \left(|\mathbf{v}|^2 - \frac{1}{|\mathbf{q}|} \right) \\ &= 0. \end{aligned}$$

This is better known as *Kepler's second law.* The triangle with constant area in Kepler's second law has $\mathbf{q}$ and $\mathbf{q} + \mathbf{v}$ as sides. Thus its area is half the area of the parallelogram we just calculated to be constant.

Exercises.

(1) Consider a second order equation
$$\mathbf{a} = F(\mathbf{q}, \mathbf{v}),$$
where $F(\mathbf{q}, \mathbf{v}) \in \operatorname{span}\{\mathbf{q}, \mathbf{v}\}$ for all vectors $\mathbf{q}, \mathbf{v} \in \mathbb{R}^n$. Show that the solutions are planar, i.e., $\operatorname{span}\{\mathbf{q}(t), \mathbf{v}(t)\}$ does not depend on t.

(2) Consider the equation
$$\mathbf{a} = -\frac{\mathbf{q}}{|\mathbf{q}|^3}.$$
Show that
$$-\frac{\mathbf{q}}{|\mathbf{q}|^3} = \nabla \frac{1}{|\mathbf{q}|}$$
and conclude that the total energy
$$E = \frac{1}{2}|\mathbf{v}|^2 - \frac{1}{|\mathbf{q}|}$$
is constant along solutions.

(3) Consider the equation
$$\mathbf{a} = -\frac{\mathbf{q}}{|\mathbf{q}|^3}$$
with $\mathbf{q} \in \mathbb{R}^2$. Show that along a solution the two equations
$$\begin{aligned} A^2 &= |\mathbf{q}|^2 |\mathbf{v}|^2 - (\mathbf{q} \cdot \mathbf{v})^2, \\ E &= \frac{1}{2}|\mathbf{v}|^2 - \frac{1}{|\mathbf{q}|} \end{aligned}$$
allow us to compute the tangent line as a function of position $\mathbf{q}$ and the two constants A, E.

(4) Consider an equation
$$\mathbf{a} = f(|\mathbf{q}|)\,\mathbf{q}$$
coming from a radial force field. Show that
$$A^2 = |\mathbf{q}|^2 |\mathbf{v}|^2 - (\mathbf{q} \cdot \mathbf{v})^2$$
is constant along solutions.

(5) Assume a planar curve is given as a level set $F(x, y) = c$, where $\nabla F \neq 0$ everywhere along the curve. We orient and parametrize the curve so that $\mathbf{v} = \left(-\frac{\partial F}{\partial y}, \frac{\partial F}{\partial x}\right)$. Use the chain rule to show that the acceleration is
$$\begin{aligned} \mathbf{a} &= \begin{bmatrix} -\frac{\partial^2 F}{\partial x \partial y} & -\frac{\partial^2 F}{\partial y^2} \\ \frac{\partial^2 F}{\partial x^2} & \frac{\partial^2 F}{\partial y \partial x} \end{bmatrix} \begin{bmatrix} -\frac{\partial F}{\partial y} \\ \frac{\partial F}{\partial x} \end{bmatrix} \\ &= \left[\frac{\partial \mathbf{v}}{\partial (x, y)}\right][\mathbf{v}]. \end{aligned}$$

CHAPTER 2

Planar Curves

2.1. General Frames

Our approach to planar curves follows very closely the concepts that we shall also use for space curves. This is certainly not the way the subject developed historically, but it has shown itself to be a very useful strategy.

Before delving into the theory the keen reader might be interested in a few generalities about taking derivatives of a basis $U(t), V(t)$ that depends on t, and viewed as a choice of basis at $\mathbf{q}(t)$. We shall normally use $U(t) = \dot{c}(t)$ or $U(t) = \mathbf{T}(t)$. Given any choice for $U(t)$, a natural choice for $V(t)$ would be the unit vector orthogonal to $U(t)$. The goal is to identify the matrix $[D]$ that appears in

$$\frac{d}{dt}\begin{bmatrix} U & V \end{bmatrix} = \begin{bmatrix} \frac{d}{dt}U & \frac{d}{dt}V \end{bmatrix} = \begin{bmatrix} U & V \end{bmatrix}[D]$$

There is a complicated formula (see theorem A.1.1)

$$[D] = \left(\begin{bmatrix} U & V \end{bmatrix}^t \begin{bmatrix} U & V \end{bmatrix}\right)^{-1} \begin{bmatrix} U & V \end{bmatrix}^t \begin{bmatrix} \frac{d}{dt}U & \frac{d}{dt}V \end{bmatrix}$$

that can be simplifies to

Theorem 2.1.1. *Let $U(t), V(t)$ be an orthonormal frame that depends on a parameter t, then*

$$\begin{aligned} \frac{d}{dt}\begin{bmatrix} U & V \end{bmatrix} &= \begin{bmatrix} U & V \end{bmatrix}\begin{bmatrix} 0 & \lambda \\ -\lambda & 0 \end{bmatrix}, \\ \lambda &= U \cdot \frac{d}{dt}V = -V \cdot \frac{d}{dt}U \end{aligned}$$

or

$$\begin{aligned} \frac{d}{dt}U &= \lambda V, \\ \frac{d}{dt}V &= -\lambda U \end{aligned}$$

Proof. We use that

$$\begin{bmatrix} U & V \end{bmatrix}^t \begin{bmatrix} U & V \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

The derivative of this then gives

$$\begin{aligned} \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} &= \begin{bmatrix} \frac{d}{dt}U & \frac{d}{dt}V \end{bmatrix}^t \begin{bmatrix} U & V \end{bmatrix} + \begin{bmatrix} U & V \end{bmatrix}^t \begin{bmatrix} \frac{d}{dt}U & \frac{d}{dt}V \end{bmatrix} \\ &= \begin{bmatrix} \left(\frac{d}{dt}U\right)\cdot U & \left(\frac{d}{dt}U\right)\cdot V \\ \left(\frac{d}{dt}V\right)\cdot U & \left(\frac{d}{dt}V\right)\cdot V \end{bmatrix} + \begin{bmatrix} U \cdot \frac{d}{dt}U & U \cdot \frac{d}{dt}V \\ V \cdot \frac{d}{dt}U & V \cdot \frac{d}{dt}V \end{bmatrix}. \end{aligned}$$

Thus

$$\begin{aligned}
\left(\frac{d}{dt}U\right)\cdot U &= 0=\left(\frac{d}{dt}V\right)\cdot V,\\
\left(\frac{d}{dt}V\right)\cdot U &= -V\cdot\frac{d}{dt}U.
\end{aligned}$$

Our formula for $[D]$ then becomes

$$\begin{aligned}
[D] &= \begin{bmatrix} U & V \end{bmatrix}^t \begin{bmatrix} \frac{d}{dt}U & \frac{d}{dt}V \end{bmatrix}\\
&= \begin{bmatrix} U\cdot\frac{d}{dt}U & U\cdot\frac{d}{dt}V \\ V\cdot\frac{d}{dt}U & V\cdot\frac{d}{dt}V \end{bmatrix}\\
&= \begin{bmatrix} 0 & \lambda \\ -\lambda & 0 \end{bmatrix}.
\end{aligned}$$

□

Occasionally we need one more derivative

$$\begin{aligned}
\frac{d^2}{dt^2}\begin{bmatrix} U & V \end{bmatrix} &= \begin{bmatrix} U & V \end{bmatrix}\begin{bmatrix} -\lambda^2 & \frac{d\lambda}{dt} \\ -\frac{d\lambda}{dt} & -\lambda^2 \end{bmatrix},\\
\frac{d^2U}{dt^2} &= -\lambda^2U-\frac{d\lambda}{dt}V,\\
\frac{d^2V}{dt^2} &= \frac{d\lambda}{dt}U-\lambda^2V.
\end{aligned}$$

2.2. The Fundamental Equations

For a planar regular curve $\mathbf{q}(t):[a,b]\to\mathbb{R}^2$ we have as for general curves

$$\frac{d\mathbf{q}}{dt}=|\mathbf{v}|\frac{\mathbf{v}}{|\mathbf{v}|}=\frac{ds}{dt}\mathbf{T}.$$

Instead of the choice of normal that depended on the acceleration (see section 1.3) we select an *oriented normal* $\mathbf{N}_\pm$ such that $\mathbf{T}$ and $\mathbf{N}_\pm$ are positively oriented, i.e., if $\mathbf{T}=(a,b)$, then $\mathbf{N}_\pm=(-b,a)$. Thus $\mathbf{N}_\pm=\pm\mathbf{N}$.

This leads us to a *signed curvature* defined by

$$\kappa_\pm=\mathbf{N}_\pm\cdot\frac{d\mathbf{T}}{ds}.$$

PROPOSITION 2.2.1. *The signed curvature can be calculated using the formula*

$$\kappa_\pm=\frac{\textit{signed area of parallelogram}\,(\mathbf{v},\mathbf{a})}{|\mathbf{v}|^3}=\frac{\det\begin{bmatrix}\mathbf{v} & \mathbf{a}\end{bmatrix}}{|\mathbf{v}|^3}$$

THEOREM 2.2.2. *(Euler, 1736) The fundamental equations that govern planar curves are*

$$\begin{aligned}
\frac{d\mathbf{q}}{dt} &= \frac{ds}{dt}\mathbf{T},\\
\frac{d\mathbf{T}}{dt} &= \frac{ds}{dt}\kappa_\pm\mathbf{N}_\pm,\\
\frac{d\mathbf{N}_\pm}{dt} &= -\frac{ds}{dt}\kappa_\pm\mathbf{T}.
\end{aligned}$$

Moreover, given an initial position $\mathbf{q}(0)$ *and unit direction* $\mathbf{T}(0)$ *the curve* $\mathbf{q}(t)$ *is uniquely determined by its speed and signed curvature.*

Proof. The three equations are simple to check as $\mathbf{T}, \mathbf{N}_{\pm}$ form an orthonormal basis. For fixed speed and signed curvature functions these equations form a differential equation which has a unique solution given the initial values $\mathbf{q}(0)$, $\mathbf{T}(0)$ and $\mathbf{N}_{\pm}(0)$. The normal vector is determined by the unit tangent so we have all of that data. $\square$

Geometrically we say that the planar curve $\mathbf{q}(t)$ is determined by the planar curve $\left(\frac{ds}{dt}, \kappa_{\pm}\right)$. If it is possible to find the arc-length parametrization, then the data $(s(t), \kappa_{\pm}(t))$ can equally well be used to describe the geometry of a planar curve.

We offer a combined characterization of lines and circles as the curves that are horizontal lines in $(s, \kappa_{\pm})$ coordinates, i.e., they have constant curvature.

Theorem 2.2.3. *A planar curve is part of a line if and only if its signed curvature vanishes. A planar curve is part of a circle if and only if its signed curvature is non-zero and constant.*

Proof. If the curvature vanishes then we already know that it has to be a straight line.

If the curve is a circle of radius R with center $\mathbf{c}$, then

$$|\mathbf{q}(s) - \mathbf{c}|^2 = R^2.$$

Differentiating this yields

$$\mathbf{T} \cdot (\mathbf{q}(s) - \mathbf{c}) = 0.$$

Thus the unit tangent is perpendicular to the radius vector $\mathbf{q}(s) - \mathbf{c}$. Differentiating again yields

$$\kappa_{\pm}\mathbf{N}_{\pm} \cdot (\mathbf{q}(s) - \mathbf{c}) + 1 = 0.$$

However the normal and radius vectors must be parallel so their inner product is $\pm R$. This shows that the curvature is constant. We also obtain the equation

$$\mathbf{q} = \mathbf{c} - \frac{1}{\kappa_{\pm}}\mathbf{N}_{\pm}.$$

This indicates that, if we take a curve with constant curvature, then we should attempt to show that

$$\mathbf{c} = \mathbf{q} + \frac{1}{\kappa_{\pm}}\mathbf{N}_{\pm}$$

is constant. Since $\kappa_{\pm}$ is constant the derivative of this curve is

$$\frac{d\mathbf{c}}{ds} = \mathbf{T} + \frac{1}{\kappa_{\pm}}(-\kappa_{\pm}\mathbf{T}) = 0.$$

So $\mathbf{c}$ is constant and

$$|\mathbf{q}(s) - \mathbf{c}|^2 = \left|\frac{1}{\kappa_{\pm}}\mathbf{N}_{\pm}\right|^2 = \frac{1}{\kappa_{\pm}^2}.$$

thus showing that $\mathbf{q}$ is a circle of radius $\frac{1}{\kappa_{\pm}}$ centered at $\mathbf{c}$. $\square$

Proposition 2.2.4. *The evolute of a unit speed planar curve $\mathbf{q}(s)$ of non-zero curvature is given by*

$$\mathbf{q}^* = \mathbf{q} + \frac{1}{\kappa_{\pm}}\mathbf{N}_{\pm} = \mathbf{q} + \frac{1}{\kappa}\mathbf{N}.$$

PROOF. This follows from remark 1.3.10 and

$$\frac{d\mathbf{q}^*}{ds} = \mathbf{T} + \frac{1}{\kappa_\pm}\left(-\kappa_\pm \mathbf{T}\right) + \frac{d}{ds}\left(\frac{1}{\kappa_\pm}\right)\mathbf{N}_\pm = \frac{d}{ds}\left(\frac{1}{\kappa_\pm}\right)\mathbf{N}_\pm.$$

□

Exercises.

(1) Show that a planar curve is part of a line if all its tangent lines pass through a fixed point.

(2) Compute the signed curvature of $\mathbf{q}(t) = \left(t, t^3\right)$ and show that it has a critical point at $t = 0$, is negative for $t < 0$, and positive for $t > 0$.

(3) Let $\mathbf{q}(s) = (x(s), y(s)) : [0, L] \to \mathbb{R}^2$ be a unit speed planar curve with signed curvature $\kappa_\pm(s)$ and $\mathbf{q}^*(s) = x(s)e_1 + y(s)e_2 + \mathbf{x}$ another planar curve where e_1, e_2 is a positively oriented orthonormal basis and $\mathbf{x}$ a point.
  (a) Show that $\mathbf{q}^*$ is a unit speed curve with curvature $\kappa_\pm^*(s) = \kappa_\pm(s)$.
  (b) Show that a planar unit speed curve with the same curvature as $\mathbf{q}$ is of the form $\mathbf{q}^*$.

(4) Compute the signed curvature of the logarithmic spiral

$$ae^{bt}(\cos t, \sin t)$$

(5) Compute the signed curvature of the spiral of Archimedes:

$$(a + bt)(\cos t, \sin t)$$

(6) Show that if a planar unit speed curve $\mathbf{q}(s)$ satisfies:

$$\kappa_\pm(s) = \frac{1}{es + f}$$

for constants $e, f > 0$, then it is a logarithmic spiral.

(7) Show that a planar curve is part of a circle if all its normal lines pass through a fixed point.

(8) Show that $\kappa_\pm \frac{ds}{dt} = \det\begin{bmatrix} \mathbf{T} & \frac{d\mathbf{T}}{dt} \end{bmatrix}$.

(9) Let $\mathbf{q}(t) = r(t)(\cos t, \sin t)$. Show that the speed is given by

$$\left(\frac{ds}{dt}\right)^2 = \left(\frac{dr}{dt}\right)^2 + r^2$$

and the curvature

$$\kappa_\pm = \frac{2\left(\frac{dr}{dt}\right)^2 + r^2 - r\frac{d^2r}{dt^2}}{\left(\left(\frac{dr}{dt}\right)^2 + r^2\right)^{\frac{3}{2}}}.$$

Parametrize the curve $\left(1 - x^2\right)x^2 = y^2$ in this way and compute its curvature. Note that such a parametrization won't be valid for all t.

(10) Assume a planar curve is given as a level set $F(x, y) = c$ where $\nabla F \neq 0$ everywhere along the curve. We orient and parametrize the curve so that $\mathbf{v} = \left(-\frac{\partial F}{\partial y}, \frac{\partial F}{\partial x}\right)$.
  (a) Show that the signed normal is given by

$$\mathbf{N}_\pm = -\frac{\nabla F}{|\nabla F|}.$$

(b) Use the chain rule to show that the acceleration is

$$\begin{aligned}\mathbf{a} &= \begin{bmatrix} -\frac{\partial^2 F}{\partial x \partial y} & -\frac{\partial^2 F}{\partial y^2} \\ \frac{\partial^2 F}{\partial x^2} & \frac{\partial^2 F}{\partial y \partial x} \end{bmatrix} \begin{bmatrix} -\frac{\partial F}{\partial y} \\ \frac{\partial F}{\partial x} \end{bmatrix} \\ &= \left[\frac{\partial \mathbf{v}}{\partial (x, y)} \right] [\mathbf{v}] .\end{aligned}$$

(c) Show that

$$\begin{aligned}\kappa_{\pm} &= \frac{1}{|\nabla F|^3} \begin{bmatrix} -\frac{\partial F}{\partial x} & -\frac{\partial F}{\partial y} \end{bmatrix} \begin{bmatrix} -\frac{\partial^2 F}{\partial x \partial y} & -\frac{\partial^2 F}{\partial y^2} \\ \frac{\partial^2 F}{\partial x^2} & \frac{\partial^2 F}{\partial y \partial x} \end{bmatrix} \begin{bmatrix} -\frac{\partial F}{\partial y} \\ \frac{\partial F}{\partial x} \end{bmatrix} \\ &= \frac{1}{|\nabla F|^3} \begin{bmatrix} -\frac{\partial F}{\partial y} & \frac{\partial F}{\partial x} \end{bmatrix} \begin{bmatrix} \frac{\partial^2 F}{\partial x^2} & \frac{\partial^2 F}{\partial x \partial y} \\ \frac{\partial^2 F}{\partial x \partial y} & \frac{\partial^2 F}{\partial y^2} \end{bmatrix} \begin{bmatrix} -\frac{\partial F}{\partial y} \\ \frac{\partial F}{\partial x} \end{bmatrix} .\end{aligned}$$

(11) (Jerrard, 1961) With notation as in the previous exercise show that

$$\kappa_{\pm} = \operatorname{div} \frac{\nabla F}{|\nabla F|}.$$

(12) Compute the curvature of $\left(1 - x^2\right) x^2 = y^2$ at the points where the above formula works. What can you say about the curvature at the origin where the curve intersects itself.

(13) Compute the curvature of the cissoid of Diocles $x\left(x^2 + y^2\right) = 2Ry^2$.

(14) Compute the curvature of the conchoid of Nicomedes $\left(x^2 + y^2\right)(y - b)^2 - R^2 y^2 = 0$.

(15) Consider a unit speed curve $\mathbf{q}(s)$ with non-vanishing curvature and use the notation $\frac{df}{ds} = f'$. Show that $\mathbf{q}$ satisfies the third order equation

$$\mathbf{q}''' - \frac{\kappa_{\pm}'}{\kappa_{\pm}} \mathbf{q}'' + \kappa_{\pm}^2 \mathbf{q}' = 0.$$

(16) Show that the curvature of the evolute $\mathbf{q}^*$ of a unit speed curve $\mathbf{q}(s)$ satisfies

$$\frac{1}{\kappa_{\pm}^*} = \frac{1}{2} \frac{d}{ds} \left(\frac{1}{\kappa_{\pm}^2} \right).$$

(17) (Huygens, 1673) Consider the cycloid

$$\mathbf{q}(t) = R(t + \sin t, 1 + \cos t).$$

It traces a point on a circle of radius R that rolls along the x-axis. Any curve that is constructed by tracing a point on a circle rolling along a line is called a cycloid (see also section 1.3 exercise 7).

(a) Show that the signed curvature is given by

$$\kappa_{\pm} = \frac{-1}{2R\sqrt{2(1 + \cos t)}}.$$

(b) Show that the evolute is also a cycloid.

(c) Show that any curve that satisfies

$$(s - a)^2 + \frac{1}{\kappa^2} = 16R^2$$

for a constant $a \in \mathbb{R}$ is a cycloid. In other words cycloids are circles centered on the first axis in $\left(s, \frac{1}{\kappa_{\pm}}\right)$ coordinates.

(d) Show that any cycloid is the involute of a cycloid.

(18) (Newton, Principia) Consider the equation

$$\mathbf{a} = -\frac{\mathbf{q}}{|\mathbf{q}|^3}$$

with $\mathbf{q} \in \mathbb{R}^2$.

(a) Show that solutions have the property that

$$A = \mathbf{q} \cdot \mathbf{N}_\pm |\mathbf{v}|$$

is constant in time.

(b) Show that

$$\kappa_\pm = \frac{A}{|\mathbf{q}|^3 |\mathbf{v}|^3}$$

and

$$\mathbf{a} = \frac{-\mathbf{q} \cdot \mathbf{v}}{|\mathbf{q}|^3 |\mathbf{v}|}\mathbf{T} + \frac{A}{|\mathbf{q}|^3 |\mathbf{v}|}\mathbf{N}_\pm.$$

(c) Show that when $A \neq 0$ then

$$\mathbf{k} = A |\mathbf{v}| \mathbf{N}_\pm + \frac{\mathbf{q}}{|\mathbf{q}|}$$

is constant and

$$A^2 = |\mathbf{q}| \left(1 - \frac{\mathbf{q} \cdot \mathbf{k}}{|\mathbf{q}|}\right).$$

(d) Show that if we parametrize the solution with respect to the angle with the axis spanned by $\mathbf{k}$, i.e., $|\mathbf{q}| |\mathbf{k}| \cos\phi = \mathbf{q} \cdot \mathbf{k}$, then this gives us the classical formula

$$|\mathbf{q}| = \frac{A^2}{1 - |\mathbf{k}| \cos\phi},$$

which for $|\mathbf{k}| < 1$ describes an ellipse, $|\mathbf{k}| = 1$ a parabola, and $|\mathbf{k}| > 1$ a hyperbola.

(19) For a planar unit speed curve $\mathbf{q}(s)$ consider the parallel curve

$$\mathbf{q}_\epsilon = \mathbf{q} + \epsilon \mathbf{N}_\pm$$

for some fixed ϵ.

(a) Show that this curve is regular as long as $\epsilon\kappa_\pm \neq 1$.

(b) Show that the curvature is

$$\frac{\kappa_\pm}{|1 - \epsilon\kappa_\pm|}.$$

(20) If a curve in $\mathbb{R}^2$ is given as a graph $y = f(x)$ show that the curvature is given by

$$\kappa_\pm = \frac{f''}{\left(1 + (f')^2\right)^{\frac{3}{2}}}.$$

(21) (Newton, 1671 and Huygens, 1673) Consider a regular planar curve $\mathbf{q}(t)$ with $\kappa_\pm(t_0) \neq 0$. Let $l(t)$ denote the normal line to $\mathbf{q}$ at $\mathbf{q}(t)$.

(a) Show that $l(t)$ and $l(t_0)$ are not parallel for t near t_0.

(b) Let $\mathbf{x}(t)$ denote the intersection of $l(t)$ and $l(t_0)$. Show that $\lim_{t\to t_0} \mathrm{x}(t)$ exists and denote this limit $\mathbf{c}(t_0)$.

(c) Show that

$$(\mathbf{c}(t_0) - \mathbf{q}(t_0)) \cdot \mathbf{N}(t_0) = \frac{1}{\kappa_\pm(t_0)}.$$

Note the the left hand side is the signed distance from $\mathbf{c}(t_0)$ to $\mathbf{q}(t_0)$ along the normal through $\mathbf{q}(t_0)$. The circle of radius $\left|\frac{1}{\kappa_\pm(t_0)}\right|$ centered at $\mathbf{c}(t_0)$ is the circle that best approximates the curve at $\mathbf{q}(t_0)$.

(d) Show that the curve $\mathbf{c}(t)$ is the evolute of $\mathbf{q}(t)$.

(22) (Newton, 1671, but the idea is much older for specific curves. Kepler considered it well-known.) Consider a regular planar curve $\mathbf{q}(t)$. For 3 consecutive values $t - \epsilon < t < t + \epsilon$ let $\mathbf{c}(t, \epsilon)$ denote the center of the unique circle that goes through the three points $\mathbf{q}(t-\epsilon)$, $\mathbf{q}(t)$, $\mathbf{q}(t+\epsilon)$ with $\mathbf{c}(t, \epsilon) = \infty$ if the points lie on a line.

(a) Show that $\mathbf{c}(t, \epsilon)$ is the point of intersection between the normal lines to the segments between $\mathbf{q}(t)$ and $\mathbf{q}(t \pm \epsilon)$ that pass through the midpoint of these segments.

(b) Show that $\mathbf{q}(t-\epsilon)$, $\mathbf{q}(t)$, $\mathbf{q}(t+\epsilon)$ do not lie on a line for small ϵ if $\kappa_\pm(t) \neq 0$.

(c) Show that $\mathbf{c}(t, \epsilon)$ lies on the normal line through some point $\mathbf{q}(t_0)$ where $t_0 \in (t-\epsilon, t+\epsilon)$. Hint: Show that there is a point on the curve in the open interval closest to $\mathbf{c}(t, \epsilon)$ and use that as the desired point.

(d) Show that

$$\lim_{\epsilon \to 0} (\mathbf{c}(t, \epsilon) - \mathbf{q}(t)) = \lim_{\epsilon \to 0} (\mathbf{c}(t, \epsilon) - \mathbf{q}(t_0)) = \frac{1}{\kappa_\pm(t)} \mathbf{N}(t).$$

(23) (Normal curves) Consider a family of lines in the (x, y)-plane parametrized by t:

$$F(x, y, t) = a(t)x + b(t)y + c(t) = 0.$$

A *normal curve* to this family is a curve $(x(t), y(t))$ such that its tangents are precisely the lines of this family.

(a) Show that such a curve exists and can be determined by the equations:

$$\begin{aligned} F = a(t)x + b(t)y + c(t) &= 0, \\ \frac{\partial F}{\partial t} = \dot{a}(t)x + \dot{b}(t)y + \dot{c}(t) &= 0 \end{aligned}$$

when the Wronskian

$$\det \begin{bmatrix} a & b & c \\ \dot{a} & \dot{b} & \dot{c} \\ \ddot{a} & \ddot{b} & \ddot{c} \end{bmatrix} \neq 0.$$

(b) Show that for fixed x_0, y_0 the number of solutions or roots to the equation $F(x_0, y_0, t) = 0$ corresponds to the number of tangent lines to the normal curve that pass through (x_0, y_0).

(c) Consider the case where $a = 1$, $b = t$, $c = t^n$, $n = 2, 3, 4, \ldots$

(i) Determine the number of roots in relation to how (x_0, y_0) is placed relative to the normal curve.

(ii) Show that multiple roots only occur when (x_0, y_0) is on the normal curve.

2.3. Length and Area

In this section we establish some fundamental results about planar curves that relate the arclength to areas or more generally surface integrals.

Oriented lines can be described by a directional angle $\theta \in [0, 2\pi)$ and a distance $p \in [0, \infty)$ from the origin, the equation for such a line is

$$x \cos\theta + y \sin\theta = p.$$

We denote the space of oriented lines by OL. To eliminate the choice of angle we identify OL with with points on the cylinder $S^1 \times [0, \infty)$, where the first factor describes a unit direction in the plane and the second a distance. This eliminates the angle choice, but it still leaves us with having specified p as a distance to a specific point.

The cylinder model helps us put a measure on OL so that we can integrate functions. This is simply done as a natural surface integral:

$$\int_{l \in OL} f(l)\, dl = \int f(l(u,v)) \left| \frac{\partial l}{\partial u} \times \frac{\partial l}{\partial v} \right| dudv.$$

If $\mathbf{q} : [a,b] \to \mathbb{R}^2$ is a curve denote by $n_{\mathbf{q}}(l)$ the number of times the curve intersects the oriented line l.

Theorem 2.3.1. *(Cauchy-Crofton) The length of a curve can be computed via the formula*

$$L(\mathbf{q}) = \frac{1}{4} \int_{l \in OL} n_{\mathbf{q}}(l)\, dl.$$

Proof. Consider a regular curve $\mathbf{q}(t) : [a,b] \to \mathbb{R}^2$. At a point $\mathbf{q}(t)$ of the curve consider all lines passing through this point. If we denote the direction of each of these oriented lines by $(\cos\theta, \sin\theta)$ then the distance from this line to the origin is given by

$$r(\theta, t) = |(-\sin\theta, \cos\theta) \cdot \mathbf{q}(t)|.$$

This gives us a parametrization of the set of lines that intersect the curve

$$l(\theta, t) = (\cos\theta, \sin\theta, p(\theta, t)).$$

However, different values of t correspond to the same line if that line intersects the curve in several points, i.e., $n_{\mathbf{q}}(l)$ is the number of t values where $l = l(\theta, t)$. Thus the integral on the right hand side can be reinterpreted as

$$\begin{aligned}
\frac{1}{4}\int_{l\in OL} n_{\mathbf{q}}(l)\,dl &= \frac{1}{4}\int_{(\theta,t)\in[0,2\pi]\times[a,b]} dl\,(\theta,t)\\
&= \frac{1}{4}\int_a^b\int_0^{2\pi}\left|\frac{\partial l}{\partial\theta}\times\frac{\partial l}{\partial t}\right| d\theta dt\\
&= \frac{1}{4}\int_a^b\int_0^{2\pi}\left|\begin{bmatrix}-\sin\theta\\ \cos\theta\\ \frac{\partial r}{\partial\theta}\end{bmatrix}\times\begin{bmatrix}0\\0\\ \frac{\partial r}{\partial t}\end{bmatrix}\right| d\theta dt\\
&= \frac{1}{4}\int_a^b\int_0^{2\pi}\left|\frac{\partial r}{\partial t}\right| d\theta dt\\
&= \frac{1}{4}\int_a^b\int_0^{2\pi}\left|(-\sin\theta,\cos\theta)\cdot\frac{d\mathbf{q}}{dt}\right| d\theta dt\\
&= \frac{1}{4}\int_a^b\int_0^{2\pi}\left|(-\sin\theta,\cos\theta)\cdot\mathbf{T}\right|\frac{ds}{dt} d\theta dt\\
&= \frac{1}{4}\int_0^L\int_0^{2\pi}\left|(-\sin\theta,\cos\theta)\cdot\mathbf{T}\right| d\theta ds.
\end{aligned}$$

Now for a fixed s (or t) consider the integral

$$\int_0^{2\pi}\left|(-\sin\theta,\cos\theta)\cdot\mathbf{T}\right| d\theta.$$

As $(-\sin\theta,\cos\theta)$ is the unit normal to the line passing through $\mathbf{q}(s)$ we note that

$$\left|(-\sin\theta,\cos\theta)\cdot\mathbf{T}\right| = \left|\sin\phi\right|,$$

where ϕ denotes the angle between the line and $\mathbf{T}(s)$. Since s is fixed we have that the two angles ϕ,θ differ by a constant, so it follows that

$$\begin{aligned}
\int_0^{2\pi}\left|(-\sin\theta,\cos\theta)\cdot\mathbf{T}\right| d\theta &= \int_0^{2\pi}\left|\sin\phi\right| d\phi\\
&= 4.
\end{aligned}$$

This in turn implies the result we wanted to prove. □

Crofton's formula will reappear in section 3.3 for curves on spheres and at that point we will give a different proof.

Corollary 2.3.2. *If we reparametrize OL using (ϕ,s) instead of (θ,s) we obtain*

$$\int_{l\in OL} f(l)\,n_{\mathbf{q}}(l)\,dl = \int_0^L\int_0^{2\pi} f(l(\phi,s))\left|\sin\phi\right| d\phi ds.$$

Proof. This is simply the change of variables formula for functions on OL given our analysis in the proof of Crofton's formula. □

Before presenting another well known and classical result relating area and length we need to indicate a proof of an intuitively obvious theorem. This result allows us to speak of the inside and outside of a simple closed planar curve.

Theorem 2.3.3. *(Jordan Curve Theorem) A simple closed planar curve divides the plane in to two regions one that is bounded and one that is unbounded.*

Proof. Note that in case the curve is also part of the regular level set of a function, then the constructions that follow are much simpler.

Consider a simple closed curve $\mathbf{q}$ that is parametrized by arclength. We construct the parallel curves

$$\mathbf{q}^{\epsilon} = \mathbf{q} + \epsilon \mathbf{N}_{\pm}.$$

The velocity is

$$\frac{d\mathbf{q}^{\epsilon}}{dt} = (1 - \epsilon \kappa_{\pm}) \mathbf{T}.$$

So as long as ϵ is small they are clearly regular. We also know that they are closed. Finally we can also show that they are simple for small ϵ. This follows from a contradiction argument. Thus assume that there is a sequence $\epsilon_i \to 0$ and distinct parameter values $s_i \neq t_i$ such that

$$\mathbf{q}(s_i) + \epsilon_i \mathbf{N}_{\pm}(s_i) = \mathbf{q}(t_i) + \epsilon_i \mathbf{N}_{\pm}(t_i).$$

By compactness we can, after passing to subsequences, assume that $s_i \to s$ and $t_i \to t$. Since $\mathbf{q}$ is simple it follows that $s = t$. Now if $|\kappa_{\pm}| \leq K$ then the derivatives of $\mathbf{N}_{\pm}$ are bounded in absolute value by K so it follows that

$$\begin{aligned} |\mathbf{q}(s_i) - \mathbf{q}(t_i)| &= |\epsilon_i| \, |\mathbf{N}_{\pm}(t_i) - \mathbf{N}_{\pm}(s_i)| \\ &\leq |\epsilon_i| \, K \, |t_i - s_i| \,. \end{aligned}$$

But this implies that the derivative of $\mathbf{q}$ vanishes at $s = t$ which contradicts that the curve is regular.

Fix $\epsilon_0 > 0$ so that both $\mathbf{q}^{\epsilon_0}$ and $\mathbf{q}^{-\epsilon_0}$ are closed simple curves. We think of them as inside and outside curves, but we don't know yet which is which. Every point p not on $\mathbf{q}$ will either lie on a parallel curve $\mathbf{q}^{\epsilon}$ with $|\epsilon| \leq \epsilon_0$, in which case we can decide which side of $\mathbf{q}$ p lies on, or the shortest line from p to $\mathbf{q}$ will cross either $\mathbf{q}^{\epsilon_0}$ or $\mathbf{q}^{-\epsilon_0}$ in a point closest to p (note that such a line crosses $\mathbf{q}$ orthogonally) so again we can decide which side of $\mathbf{q}$ p lies on. Next, it is not too difficult to see that these two regions are open and connected. Finally, one of them is bounded and that will be the inside region. □

The *isoperimetric ratio* of a simple closed planar curve $\mathbf{q}$ is L^2/A where L is the perimeter, i.e., length of $\mathbf{q}$, and A is the area of the interior. We say that $\mathbf{q}$ *minimizes the isoperimetric ratio* if L^2/A is as small as it can be.

The isomerimetric inequality asserts that the isoperimetric ratio always exceeds 4π and is only minimal for circles. This will be established in the next theorem using a very elegant proof that does not assume the existence of a curve that realizes this ratio. Steiner in the 1830s devised several intuitive proofs of the isoperimetric inequality assuming that such minimizers exist. It is, however, not so simple to show that such curves exist as Dirichlet repeatedly pointed out to Steiner. Some of Steiner's ideas will be explored in the exercises.

The isoperimetric inequality would seem almost obvious and has been investigated for millennia. In fact a related problem, known as Dido's problem, appears in ancient legends. Dido founded Carthage and was faced with the problem of enclosing the largest possible area for the city with a long string (called a length of hide as the string had to be cut from a cow hide). However, the city was to be placed along the shoreline and so it was only necessary to enclose the city on the land side. In mathematical terms we can let the shore line be a line, and the curve that will enclose the city on the land side is a curve that begins and ends on the line and otherwise stays on one side of the line. It is not hard to imagine that a semicircle whose diameter is on the given line yields the largest area for a curve of fixed length.

THEOREM 2.3.4. *The isoperimetric inequality states that if a simple closed curve bounds an area A and has circumference L, then*

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

Moreover, equality can only happen when the curve is a circle.

PROOF. (Knothe, 1957) We give a very direct proof using Green's theorem in the form of the divergence theorem. Unlike many other proofs, this one also easily generalizes to higher dimensions.

Consider a simple closed curve $\mathbf{q}$ of length L that can be parametrized by arclength. The domain of area A is then the interior of this curve. Let the domain be denoted Ω. We wish to select a (Knothe) map $F : \Omega \to B(0, R)$ where $B(0, R)$ also has area A. More specifically we seek a map with the properties

$$F(u, v) = (x(u), y(u, v))$$

and

$$\det DF = \frac{\partial x}{\partial u} \frac{\partial y}{\partial v} = 1.$$

Such a map can be constructed if we select $x(u_0)$ and $y(u_0, v_0)$ for a specific $(u_0, v_0) \in \Omega$ to satisfy

$$\text{area}(\{u < u_0\} \cap \Omega) = \text{area}(\{x < x(u_0)\} \cap B(0, R))$$

and

$$\text{area}(\{u < u_0\} \cap \{v < v_0\} \cap \Omega) = \text{area}(\{x < x(u_0)\} \cap \{y < y(u_0, v_0)\} \cap B(0, R)).$$

The choice of $B(0, R)$ together with the intermediate value theorem guarantee that we can construct this map. By choice, this map is area preserving as it is forced to map any rectangle in Ω to a region of equal area in $B(0, R)$. To see this note that it preserves the area of sets $\{u_0 \leq u < u_1\} \cap \Omega$ as they can be written as a difference of sets

$$\{u_0 \leq u < u_1\} \cap \Omega = \{u < u_1\} \cap \Omega - (\{u < u_0\} \cap \Omega)$$

whose areas are preserved by definition of the map. We then obtain the rectangle $[u_0, u_1) \times [v_0, v_1)$ by intersecting this strip with the set $\{v_0 \leq v < v_1\} \cap \Omega$. Thus this rectangle is in turn written as a difference

$$\begin{aligned}
[u_0, u_1) \times [v_0, v_1) &= (\{u < u_1\} \cap \Omega - (\{u < u_0\} \cap \Omega)) \cap (\{v < v_1\} \cap \Omega - (\{v < v_0\} \cap \Omega)) \\
&= (\{u < u_1\} \cap \Omega - (\{u < u_0\} \cap \Omega)) \cap (\{v < v_1\} \cap \Omega) \\
&\quad - (\{u < u_1\} \cap \Omega - (\{u < u_0\} \cap \Omega)) \cap (\{v < v_0\} \cap \Omega)
\end{aligned}$$

between two sets whose areas are preserved by the map.

The two conditions additionally force $\frac{\partial x}{\partial u} > 0$, $\frac{\partial y}{\partial v} > 0$. To prove the isoperimetric inequality we use Green's theorem in the form of the divergence theorem in the plane. The vector field is given by the map F. Note that the outward unit normal for Ω is the vector $-\mathbf{N}_{\pm}$ if the curve $\mathbf{q}$ runs counter clockwise. Using that $|F| \leq R$ we obtain:

$$\begin{aligned} \int_{\Omega} \operatorname{div} F du dv &= -\int_0^L F \cdot \mathbf{N}_{\pm} ds \\ &\leq RL. \end{aligned}$$

On the other hand the geometric mean $\sqrt{ab}$ is always smaller than the arithmetic mean $\frac{1}{2}(a+b)$ so we also have:

$$\begin{aligned} \operatorname{div} F &= \frac{\partial x}{\partial u} + \frac{\partial y}{\partial v} \\ &\geq 2\sqrt{\frac{\partial x}{\partial u}\frac{\partial y}{\partial v}} \\ &= 2. \end{aligned}$$

Consequently

$$2A \leq RL,$$

which implies

$$4A^2 \leq R^2 L^2.$$

Now we constructed $B(0,R)$ so that $A = \pi R^2$. So we obtain the isoperimetric inequality

$$4\pi A \leq L^2.$$

The equality case can only occur when we have equality in all of the above inequalities. In particular

$$\frac{\partial x}{\partial u} = \frac{\partial y}{\partial v}$$

everywhere showing that

$$\frac{\partial x}{\partial u} = \frac{\partial y}{\partial v} = 1.$$

This tells us that the function takes the form: $F(u,v) = (u + u_0, v + g(u))$. We also used that $|F \cdot \mathbf{N}_{\pm}| \leq |F| \leq R$ when the function is restricted to the boundary curve. Thus we also have $F \circ \mathbf{q} = -R\mathbf{N}_{\pm}$, i.e.,

$$\mathbf{q} + (u_0, g(u(s))) = -R\mathbf{N}_{\pm},$$

where $\mathbf{q}(s) = (u(s), v(s))$. Differentiating with respect to s then implies that

$$(1 - R\kappa)\mathbf{T} = \left(0, \frac{\partial g}{\partial u}\frac{du}{ds}\right).$$

This means either that $1 = R\kappa$ or that $\mathbf{q}$ is constant in the first coordinate. In the latter case $\frac{du}{ds} = 0$, so it still follows that $1 = R\kappa$. Thus the curve has constant non-zero curvature which shows that it must be a circle. □

REMARK 2.3.5. We've used without justification that the Knothe map is smooth so that we can take its divergence. This may however not be the case. The partial derivative $\frac{\partial x}{\partial u}$, when it exists, is equal to the sum of lengths of the intervals that make up the set $\{u = u_0\} \cap \Omega$. So if we assume that part of the boundary is a vertical line at $u = u_0$ and that the domain contains points both to the right and left of this line, then $\frac{\partial x}{\partial u}$ is not continuous at $u = u_0$.

To get around this issue one can assume that the domain is convex. Or in general that the boundary curve has the property that its tangent lines at points where the curvature vanishes are not parallel to the axes. The latter condition can generally be achieved by rotating the curve and appealing to Sard's theorem. Specifically, we wish to ensure that the normal $\mathbf{N}_{\pm}$ is never parallel to the axes at places where $\frac{d\mathbf{N}_{\pm}}{dt} = 0$.

Alternately it is also possible to prove the divergence theorem under fairly weak assumptions about the derivatives of the function.

Exercises.

(1) Show that for a domain $\Omega \subset \mathbb{R}^2$ with smooth boundary curve $\mathbf{q}$, the divergence theorem

$$\int_{\Omega} \operatorname{div} F dudv = -\int_0^L F \cdot \mathbf{N}_{\pm} ds$$

follows from Green's theorem.

(2) Show that

$$A = \int_{\Omega} dudv = -\frac{1}{2}\int_0^L \mathbf{q} \cdot \mathbf{N}_{\pm} ds.$$

(3) Compute the area in the leaf of the folium of Descartes (see section 1.1 exercise 13).

(4) We say that a simple closed planar curve $\mathbf{q}$ has *convex interior* if the domain Ω bounded by $\mathbf{q}$ has the property that for any two points in Ω the line segments between the points also lie in Ω.
   (a) Show that if $\mathbf{q}$ minimizes the isoperimetric ratio, then its interior must be convex.
   (b) Show that if $\mathbf{q}$ minimizes the isoperimetric ratio and has perimeter L, then any section of $\mathbf{q}$ that has length $L/2$ solves Dido's problem.
   (c) Show that the isoperimetric problem is equivalent to Dido's problem.

(5) Consider all triangles where two sides a, b are fixed. Show that the triangle of largest area is the right triangle where a and b are perpendicular. Note that this triangle can be inscribed in a semicircle where the diameter is the hypotenuse. Use this to solve Dido's problem if we assume that there is a curve that solves Dido's problem.

(6) Show that among all quadrilaterals that have the same four side lengths $a, b, c, d > 0$ in order, the one with the largest area is the one that can be inscribed in a circle so that all four vertices are on the circle. Use this to solve the isoperimetric problem assuming that there is a curve that minimizes the isoperimetric ratio.

(7) Try to prove that the regular $2n$-gon maximizes the area among all $2n$-gons with the same perimeter.

2.4. The Rotation Index

We now turn to a geometric interpretation of the signed curvature.

THEOREM 2.4.1. *For a regular curve the angle between the unit tangent and the x-axis is an anti-derivative of the signed curvature with respect to arclength.*

PROOF. We start with an analysis of the problem. Assume that we have a parametrization (we don't know yet that it is possible to select such a parametrization) of the unit tangent by using the angle to the first axis:

$$\begin{aligned}\mathbf{T}(t) &= (\cos\theta(t), \sin\theta(t)),\\ \mathbf{N}_\pm(t) &= (-\sin\theta(t), \cos\theta(t)).\end{aligned}$$

In this case

$$\frac{ds}{dt}\kappa_\pm \mathbf{N}_\pm = \frac{d\mathbf{T}}{dt} = \frac{d\theta}{dt}\mathbf{N}_\pm.$$

So we should be able to declare that θ is an antiderivative of $\frac{ds}{dt}\kappa_\pm$. Note that as long as the signed curvature is non-negative this is consistent with the interpretation of θ as an arclength parameter for $\mathbf{T}$.

To verify that such a choice works, define

$$\begin{aligned}\theta(t_1) &= \theta_0 + \int_a^{t_1} \frac{ds}{dt}\kappa_\pm dt, \text{ where}\\ \mathbf{T}(a) &= (\cos\theta_0, \sin\theta_0)\end{aligned}$$

and consider the orthonormal unit fields

$$\begin{aligned}U &= (\cos\theta(t), \sin\theta(t)),\\ V &= (-\sin\theta(t), \cos\theta(t)).\end{aligned}$$

They are clearly related by

$$\frac{dU}{dt} = \frac{d\theta}{dt}V.$$

If we can show that $\mathbf{T}\cdot U \equiv 1$, then it follows that $\mathbf{T} = U$. Our choice of θ_0 forces the dot product to be 1 at $t = a$. To show that it is constant we show that the derivative vanishes

$$\begin{aligned}\frac{d}{dt}(\mathbf{T}\cdot U) &= \frac{d\mathbf{T}}{dt}\cdot U + \mathbf{T}\cdot\frac{dU}{dt}\\ &= \frac{ds}{dt}\kappa_\pm \mathbf{N}_\pm \cdot U + \frac{d\theta}{dt}\mathbf{T}\cdot V\\ &= \frac{ds}{dt}\kappa_\pm(\mathbf{N}_\pm\cdot U + \mathbf{T}\cdot V)\\ &= 0,\end{aligned}$$

where the last equality follows by noting that if $\mathbf{T} = (f, g)$, then $\mathbf{N}_\pm(-g, f)$ so

$$\mathbf{N}_\pm\cdot U + \mathbf{T}\cdot V = -g\cos + f\sin + -f\sin + g\cos = 0.$$

In other words, the two inner products define complementary angles. □

DEFINITION 2.4.2. The *total curvature* of a curve $\mathbf{q} : [a, b] \to \mathbb{R}^2$ is defined as

$$\int_a^b \kappa_\pm \frac{ds}{dt}dt.$$

When we reparametrize the curve by arclength this simplifies to

$$\int_0^L \kappa_\pm ds.$$

The total curvature measures the total change in the tangent since the curvature measures the infinitesimal change of the tangent.

The ancient Greeks actually used a similar idea to calculate the angle sum in a convex polygon. Specifically, the sum of the exterior angles in a polygon adds up to 2π. This is because we can imagine the tangent line at each vertex jumping from one side to the next and while turning measuring the angle it is turning. When we return to the side we started with we have completed a full circle. When the polygon has n vertices this gives us the formula $(n-2)\pi$ for the sum of the interior angles.

A similar result holds for closed planar curves as $\mathbf{T}(a) = \mathbf{T}(b)$ for such a curve.

Proposition 2.4.3. *The total curvature of a planar closed curve is an integer multiple of 2π.*

The integer is called the *rotation index* of the curve:

$$i_{\mathbf{q}} = \frac{1}{2\pi}\int_a^b \kappa_{\pm}\frac{ds}{dt}dt.$$

We can more generally define the *winding number* of a closed unit curve $\mathbf{t} : [a,b] \to S^1 \subset \mathbb{R}^2$. Being closed now simply means that $\mathbf{t}(a) = \mathbf{t}(b)$. The idea is to measure the number of times such a curve winds or rotates around the circle. The specific formula is very similar. First construct the positively oriented normal $\mathbf{n}(t)$ to $\mathbf{t}(t)$, i.e. the unit vector perpendicular to $\mathbf{t}(t)$ such that $\det\begin{bmatrix} \mathbf{t}(t) & \mathbf{n}(t) \end{bmatrix} = 1$ and then check the change of $\mathbf{t}$ against $\mathbf{n}$. Note that as $\mathbf{t}$ is a unit vector its derivative is proportional to $\mathbf{n}$. The winding number is given by

$$w_{\mathbf{t}} = \frac{1}{2\pi}\int_a^b \frac{d\mathbf{t}}{dt}\cdot \mathbf{n}dt.$$

With this definition

$$i_{\mathbf{q}} = w_{\mathbf{T}}.$$

Proposition 2.4.4. *The winding number of a closed unit curve is an integer. Moreover, it doesn't change under small changes in $\mathbf{t}$.*

Proof. The results holds for all continuous curves, but as we've used derivatives to define it we have to assume that it is smooth. However, the proof works equally well if we assume that the curve is piecewise smooth.

As above define

$$\theta(t_0) = \theta_0 + \int_a^{t_0} \frac{d\mathbf{t}}{dt}\cdot \mathbf{n}dt,$$

where

$$\mathbf{t}(a) = (\cos\theta_0, \sin\theta_0).$$

Then show that

$$\mathbf{t}(t) = (\cos\theta(t), \sin\theta(t)).$$

Next suppose that we have two $\mathbf{t}_1, \mathbf{t}_2$ parametrized on the same interval $[a,b]$ such that

$$|\mathbf{t}_1 - \mathbf{t}_2| \leq \epsilon < 2.$$

If in addition their derivatives are also close and bounded then it is not hard to see directly that the winding numbers are close. However, as they are integers, the only way in which they can be close is if they agree.

To prove the result without assumptions about derivatives we start with the crucial observation that if

$$|\theta_1 - \theta_2| < \pi,$$

then

$$|\theta_1 - \theta_2| < \frac{\pi}{2} \left|(\cos\theta_1, \sin\theta_1) - (\cos\theta_2, \sin\theta_2)\right|.$$

In other words if the difference in angles between two points on the circle is less that π then the difference in angles is bounded by a uniform multiple of the cord length between the points.

Now assume that we have

$$\begin{aligned}
\mathbf{t}_1(t) &= (\cos\theta_1(t), \sin\theta_1(t)), \\
\mathbf{t}_2(t) &= (\cos\theta_2(t), \sin\theta_2(t)),
\end{aligned}$$

with

$$|\theta_1(a) - \theta_2(a)| < \pi,$$

then we claim that

$$|\theta_1(t) - \theta_2(t)| < \frac{\pi}{2} |\mathbf{t}_1(t) - \mathbf{t}_2(t)|$$

for all t.

We know the claim holds for $t = a$ and as all the functions are continuous the set of parameters t that satisfy this condition is open (it is a strict inequality). Next we can show that this set is also closed. To see this assume that the inequality holds for t_n and that $t_n \to t$. We have

$$|\theta_1(t_n) - \theta_2(t_n)| < \frac{\pi}{2} |\mathbf{t}_1(t_n) - \mathbf{t}_2(t_n)| \leq \frac{\pi}{2}\epsilon,$$

so it follows from continuity that

$$|\theta_1(t) - \theta_2(t)| \leq \frac{\pi}{2}\epsilon < \pi.$$

This shows that

$$|\theta_1(t) - \theta_2(t)| < \frac{\pi}{2} |\mathbf{t}_1(t) - \mathbf{t}_2(t)|.$$

It now follows that

$$\begin{aligned}
|w_{\mathbf{t}_1} - w_{\mathbf{t}_2}| &\leq \frac{1}{2\pi} |(\theta_1(b) - \theta_1(a)) - (\theta_2(b) - \theta_2(a))| \\
&\leq \frac{1}{2\pi} |(\theta_1(b) - \theta_2(b)) - (\theta_1(a) - \theta_2(a))| \\
&\leq \frac{1}{2\pi} |(\theta_1(b) - \theta_2(b))| + \frac{1}{2\pi} |(\theta_1(a) - \theta_2(a))| \\
&\leq \frac{1}{2}\epsilon < 1.
\end{aligned}$$

This shows that the winding numbers are equal. □

The next theorem is often called the Umlaufsatz (going around theorem). It is universally credited to H. Hopf, however, the name and theorem is certainly due to A. Ostrowski. Ostrowski's papers were in fact published in the same journal in the same year as Hopf's paper. Hopf's proof was meant as a shorter more elegant version of Ostrowski's far longer version. Ostrowski himself refers to the theorem as Rolle's theorem.

THEOREM 2.4.5. *(Ostrowski, 1935) A simple closed curve has rotation index* ± 1.

PROOF. (Hopf, 1935) We assume that we have a simple closed curve $\mathbf{q}(s) : [0, l] \to \mathbb{R}^2$ that is parametrized by arclength. Moreover, after possibly rotating and translating the curve we'll assume that $\mathbf{q}(0) = (0, 0)$, $\mathbf{T}(0) = (\pm 1, 0)$, and $x(s) \geq 0$ for all s. The idea is to create a family of unit vectors on a triangle where $0 \leq s \leq t \leq l$.

$$\mathbf{T}(s,t) = \begin{cases} \mathbf{T}(s) & s = t, \\ -\mathbf{T}(0) & s = 0,\ t = l, \\ \frac{\mathbf{q}(t) - \mathbf{q}(s)}{|\mathbf{q}(t) - \mathbf{q}(s)|} & \text{for all other } s < t. \end{cases}$$

Since the curve is simple, closed, and smooth this will yield a well-defined function whose values are aways unit vectors. If we select any simple path in this triangle that passes from $(0, 0)$ to (l, l) then $\mathbf{T}$ will wind around the unit circle and end up where it began as $\mathbf{T}(0, 0) = \mathbf{T}(l, l)$. Moreover, if we make a slight change in this path it will wind around the same number of times. Along the diagonal the number of windings is the rotation index of the curve. However, if we move up the y-axis and then along the upper edge of the triangle, then we are first following $\mathbf{T}(0, t) = \frac{\mathbf{q}(t)}{|\mathbf{q}(t)|}$ and then $\mathbf{T}(s, l) = \frac{\mathbf{q}(l) - \mathbf{q}(s)}{|\mathbf{q}(l) - \mathbf{q}(s)|}$. Assume that $\mathbf{T}(0) = (1, 0)$, then $\mathbf{T}(0, t)$ rotates precisely π from right to left while it points upwards as $\mathbf{q}$ lies in the upper half plane, and $\mathbf{T}(s, l)$ rotates π from left to right while pointing downwards. Thus this rotation is precisely 2π. This shows that $\mathbf{q}$ also has rotation index 1. When instead $\mathbf{T}(0) = (-1, 0)$ the rotation index is -1. □

DEFINITION 2.4.6. The *total absolute curvature* is defined as

$$\int_a^b \kappa \frac{ds}{dt} dt = \int_a^b |\kappa_\pm| \frac{ds}{dt} dt.$$

Exercises.

(1) Let $\mathbf{q}(t) = r(t)(\cos(nt), \sin(nt))$ where is $t \in [0, 2\pi]$, $n \in \mathbb{Z}$, and $r(t) > 0$ is 2π-periodic. Show that $i_{\mathbf{q}} = n$.
(2) Show that the rotation index for $\left(1 - x^2\right) x^2 = y^2$ is zero. Show that the total absolute curvature is $> 2\pi$.
(3) Let $\mathbf{q}(s) : [0, L] \to \mathbb{R}^2$ be a unit speed curve that is piecewise smooth, i.e., the domain can be subdivided

$$0 = a_1 < a_2 < \cdots < a_{k+1} = L$$

such that the curve is smooth on each interval $[a_i, a_{i+1}]$, $i = 1, ..., k$. The *exterior angle* $\theta_i \in [-\pi, \pi]$ at a_i is defined by

$$\begin{aligned} \cos\theta_i &= \mathbf{T}\left(a_i^-\right) \cdot \mathbf{T}\left(a_i^+\right), \\ \sin\theta_i &= \mathbf{N}_\pm\left(a_i^-\right) \cdot \mathbf{T}\left(a_i^+\right), \end{aligned}$$

where

$$\mathbf{T}\left(a_i^\pm\right) = \frac{d\mathbf{q}}{ds^\pm}(a_i) = \lim_{h \to 0} \frac{\mathbf{q}(a_i \pm h) - \mathbf{q}(a_i)}{\pm h}$$

and $\mathbf{N}_\pm$ defined as the corresponding signed normal.

(a) If $\mathbf{q}$ is closed show that
$$\int_0^L \kappa_\pm ds + \sum_{i=1}^k \theta_i = i_{\mathbf{q}} 2\pi$$
for some $i_{\mathbf{q}} \in \mathbb{Z}$.
(b) If $\mathbf{q}$ is both closed and simple show that $i_{\mathbf{q}} = \pm 1$.
(c) Show that the sum of the exterior angles in a polygon is 2π if the polygon is oriented appropriately.

(4) Let
$$\mathbf{q}(t) = (1 + a\cos t)(\cos t, \sin t), \ t \in [0, 2\pi].$$
(a) Show that this is a simple curve when $|a| < 1$ and intersects it self once when $|a| > 1$. Hint: Show that if $r(t) > 0$, then $r(t)(\cos t, \sin t)$ defines a simple curve. When $r(t)$ changes sign investigate what happens when it vanishes.
(b) Show that
$$\frac{d\theta}{dt} = 1 + \frac{a(a + \cos t)}{1 + a^2 + 2a\cos t}$$
and conclude that
$$\int_0^{2\pi} \frac{a(a + \cos t)}{1 + a^2 + 2a\cos t} dt = \begin{cases} 0 & |a| < 1, \\ 2\pi & |a| > 1. \end{cases}$$

(5) Show that any closed planar curve satisfies
$$\int_a^b \kappa \frac{ds}{dt} dt \geq 2\pi.$$

(6) Show that by selecting a very flat ∞ shape where the tangents at the intersection are close to the x-axis we obtain examples with rotation index 0 and total absolute curvature close to 2π.

(7) Let $\mathbf{q} : [0, L] \to \mathbb{R}^2$ be a closed curve parametrized by arclength. Show that if $\int_0^L \kappa ds = 2\pi$, then $\kappa_\pm$ cannot change sign and the rotation index is ± 1. Later we will show that this implies that the curve is simple as well.

(8) Let $\mathbf{q}(t)$, $t \in [a, b]$ be a regular planar curve and $\theta(t) \in [\theta_0, \theta_1]$ an arclength parameter for $\mathbf{T}$. Define $v(t)$ as the distance from the origin to the tangent line through $\mathbf{q}(t)$.
(a) Show that
$$v(t) = -\mathbf{q}(t) \cdot \mathbf{N}_\pm(t).$$
(b) Show by an example (e.g., a straight line) that $\mathbf{q}$ is not necessarily a function of θ.
(c) Define the curve
$$\mathbf{q}^*(\theta) = \frac{dv}{d\theta}\mathbf{T} - v\mathbf{N}_\pm = \frac{dv}{d\theta}(\cos\theta, \sin\theta) - v(-\sin\theta, \cos\theta)$$
and show that
$$\frac{d\mathbf{q}^*}{d\theta} = \left(\frac{d^2 v}{d\theta^2} + v\right)(\cos\theta, \sin\theta).$$
(d) Show that when $\mathbf{q}^*$ is a regular curve then it is a reparametrization of $\mathbf{q}$.

(e) Under that assumption show further that

$$v + \frac{d^2 v}{d\theta^2} = \frac{1}{\kappa},$$

$$L(\mathbf{q}) = \int_{\theta_0}^{\theta_1} v(\theta)\, d\theta.$$

(f) How is this related to Crofton's formula?

2.5. Two Surprising Results

DEFINITION 2.5.1. A *vertex* of a curve is a point on the curve where the curvature is a local maximum or a local minimum.

THEOREM 2.5.2. *(Mukhopadhyaya, 1909 and Kneser, 1912) A simple closed curve has at least 4 vertices.*

We start with the following observation.

PROPOSITION 2.5.3. *Suppose we have a curve* $\mathbf{q}$ *that is tangent to a circle and lies inside (resp. outside) the circle, then its curvature is larger (resp. smaller) than or equal to the curvature of the circle at the points where they are tangent.*

PROOF. Assume the curve $\mathbf{q}$ is tangent to the circle of radius R centered at $\mathbf{c}$ at $s = s_0$. This implies that

$$|\mathbf{q}(s) - \mathbf{c}|^2 \leq R^2 \text{ and } |\mathbf{q}(s_0) - \mathbf{c}|^2 = R^2.$$

Thus the function $s \mapsto |\mathbf{q}(s) - \mathbf{c}|^2$ has a (local) maximum at $s = s_0$. This implies that its derivative at s_0 vanishes. This is simply the fact that the curve is tangent to the circle. Moreover, the second derivative is nonpositive. Assume that both circle and curve are parametrized to run counter clockwise. Thus they have the same unit tangents at s_0 and consequently also the same inward pointing normals. This normal is

$$\mathbf{N}_{\pm}(s_0) = -\frac{\mathbf{q}(s_0) - \mathbf{c}}{|\mathbf{q}(s_0) - \mathbf{c}|} = -\frac{\mathbf{q}(s_0) - \mathbf{c}}{R}.$$

The second derivative of $s \mapsto |\mathbf{q}(s) - \mathbf{c}|^2$ is

$$2 + 2(\mathbf{q} - \mathbf{c}) \cdot \ddot{\mathbf{q}} = 2 + 2\kappa_{\pm}\mathbf{N}_{\pm} \cdot (\mathbf{q} - \mathbf{c}).$$

Therefore, at s_0 we have

$$\begin{aligned} 0 &\geq 2 + 2\kappa_{\pm}(s_0)\, \mathbf{N}_{\pm}(s_0) \cdot (\mathbf{q}(s_0) - \mathbf{c}) \\ &= 2 + 2\kappa_{\pm}(s_0)\, \mathbf{N}_{\pm}(s_0) \cdot (-R\mathbf{N}_{\pm}(s_0)) \\ &= 2 - 2R\kappa_{\pm}(s_0). \end{aligned}$$

This implies our claim.

The proof of the second claim about the curve lying outside the circle is proved in the same way with all inequalities reversed. □

We are now ready to prove the four vertex theorem. Mukhopadhyaya proved this result for simple planar curves with strictly positive curvature and a few years later Kneser proved the general version, apparently without knowledge of Mukhopadhyaya's earlier contribution. An excellent account of the history of this fascinating result can be found here: http://www.ams.org/notices/200702/fea-gluck.pdf

Proof. (Osserman, 1985) Select the circle of smallest radius R circumscribing the simple closed curve. The points of contact between this circle and the curve cannot lie on one side of a diagonal. If they did, then it'd be possible to slide the circle in the orthogonal direction to the diagonal until it doesn't hit the curve. We could then find a circle of smaller radius that contains the curve. This means that we can find points $q_1, ..., q_{k+1}$ of contact where $q_{k+1} = q_1$ and either $k = 2$ and q_1 and q_2 are antipodal, or $k > 2$ and any two consecutive points q_i and q_{i+1} lie one one side of a diagonal. Note there might be more points of contact.

Now orient both circle and curve so that their normals always point inside. At points of contact where the tangent lines are equal, the normal vectors must then also be equal, as the curve is inside the circle. This forces the unit tangent vectors to be equal.

First observe that the curvature at these k points is $\geq R^{-1}$.

If the curve coincides with the circle between two consecutive points of contact q_i and q_{i+1}, then the curvature is constant and we have infinitely many vertices. Otherwise there will be a point q on the curve between q_i and q_{i+1} that is inside the circle. Then we can select a circle of radius $> R$ that passes through q_i and q_{i+1} and still contains q in its interior. Now slide this new circle orthogonally to the cord between q_i and q_{i+1} until the part of the curve between q_i and q_{i+1} lies outside the circle but still touches it somewhere. At this place the curvature will be $< R^{-1}$.

This shows that we can find k points where the curvature is $\geq R^{-1}$ and k points between these where the curvature is $< R^{-1}$. This implies that there must be at least k local maxima for κ where the curvature is $\geq R^{-1}$ and between each two consecutive local maxima a minimum where the curvature is $< R^{-1}$. Note that the maxima and minima don't have to be at the points of contact. Thus we have found $2k$ vertices. □

Definition 2.5.4. For a line and a curve consider the points on the line where the curve is tangent to the line. This set will generally be empty. We say that the line is a *double tangent* if this set is not empty and not a segment of the line. Thus the curve will have contact with the line in at least two places but will not have contact with the line at all of the points in between these two points of contact.

When a curve is not too wild it is possible to relate double tangents and self intersections.

A generic curve is defined as a regular curve such that:

(1) Tangent lines cannot be tangent to the curve at more than 2 points.
(2) At self-intersection points the curve intersects itself twice.
(3) The curve only has a finite number of inflection points where the curvature changes sign.
(4) Finally, no point on the curve can belong to more than one of these categories of points.

For a generic curve T_+ is the number of tangent lines that are tangent to the curve in two places such that the curve lies on the same side of the tangent line at the points of contact. T_- is the number of tangent lines that are tangent to the curve in two places such that the curve lies on opposite sides of the tangent line at the points of contact. I is the number of inflection points, i.e., points where the curvature changes sign. D is the number of self-intersections (double points).

THEOREM 2.5.5. *(Fabricius-Bjerre, 1962) For a generic closed curve we have*

$$2T_+ - 2T_- - 2D - I = 0.$$

PROOF. The proof proceeds by checking the number of intersections between the positive tangent lines $\mathbf{q}(t) + r\mathbf{v}(t)$, $r \geq 0$ and the curve as we move forwards along the curve. As we move along the curve this number will change but ultimately return to its initial value.

When we pass through an inflection point or a self-intersection this number will decrease by 1. When we pass a point that corresponds to a double tangent $T_\pm$ the change will be 0 or ± 2 with the sign being consistent with the type of tangent.

To keep track of what happens we subdivide the two types of double tangents into three categories denoted $\overset{\rightarrow\rightarrow}{T}_\pm, \overset{\rightarrow\leftarrow}{T}_\pm, \overset{\leftarrow\rightarrow}{T}_\pm$. Here $\overset{\rightarrow\rightarrow}{T}_\pm$ indicates that the tangent vectors at the double points have the same directions, $\overset{\rightarrow\leftarrow}{T}_\pm$ indicate that the tangent vectors at the double points have opposite directions but towards each other, and $\overset{\leftarrow\rightarrow}{T}_\pm$ indicate that the tangent vectors at the double points have opposite directions but away from each other.

For double tangents of the type $\overset{\leftarrow\rightarrow}{T}_\pm$ no intersections will be gained or lost as we pass through points of that type. For $\overset{\rightarrow\leftarrow}{T}_\pm$ the change is always ± 2 at both of the points of contact. For $\overset{\rightarrow\rightarrow}{T}_\pm$ the change is ± 2 for one of the points and 0 for the other. Thus as we complete one turn of the curve we must have

$$2\overset{\rightarrow\rightarrow}{T}_+ + 4\overset{\rightarrow\leftarrow}{T}_+ - 2\overset{\rightarrow\rightarrow}{T}_- - 4\overset{\rightarrow\leftarrow}{T}_- - I - 2D = 0.$$

We now reverse the direction of the curve and repeat the counting procedure. The points of type $I, D, \overset{\rightarrow\rightarrow}{T}_\pm$ remain the same, while the points of types $\overset{\rightarrow\leftarrow}{T}_\pm$ and $\overset{\leftarrow\rightarrow}{T}_\pm$ are interchanged. Thus we also have

$$2\overset{\rightarrow\rightarrow}{T}_+ + 4\overset{\leftarrow\rightarrow}{T}_+ - 2\overset{\rightarrow\rightarrow}{T}_- - 4\overset{\leftarrow\rightarrow}{T}_- - I - 2D = 0.$$

Adding these two equations and dividing by 2 now gives us the formula. □

Exercises.

(1) Show that a vertex is a critical point for the curvature. Draw an example where a critical point for the curvature does not correspond to a local maximum/minimum.
(2) Show that a simple closed planar curve $\mathbf{q}(t)$ has the property that its unit tangent $\mathbf{T}$ is parallel to $\frac{d^2\mathbf{T}}{ds^2}$ at at least four points.
(3) Show that concept of a vertex does not depend on the parametrization of the curve.
(4) Show that an ellipse that is not a circle has 4 vertices.
(5) Find the vertices of the curve
$$x^4 + y^4 = 1.$$
(6) Show that the curve
$$(1 - 2\sin\theta)(\cos\theta, \sin\theta)$$
is not simple and has exactly two vertices.

(7) Show that a vertex for a curve given by a graph $y = f(x)$ satisfies

$$\left(1+\left(\frac{df}{dx}\right)^2\right)\frac{d^3f}{dx^3} = 3\frac{df}{dx}\left(\frac{d^2f}{dx^2}\right)^2.$$

(8) Consider a curve

$$\mathbf{q}(t) = r(t)(\cos t, \sin t)$$

where $r > 0$ and is 2π-periodic. Draw pictures where maxima/minima for r correspond to vertices. Is it possible to find an example where the minimum for r corresponds to a local maximum for $\kappa_\pm$ and the maximum for r corresponds to a local minimum for $\kappa_\pm$?

Possible exercises: Define what it means for a curve to lie inside a simple closed curve. If the simple closed curve is a circle this is easy, in general, a similar definition can be adopted if the simple closed curve is defined as a level set of a function. Otherwise we have to use the Jordan curve thm.

Define what it means for a curve to locally be on one side of the other relative to a line. In this case think of the line as being the x-axis, then one curves lies above the other if y-coordinates that correspond to the same x-coordinate lie above each other. Note that this is strictly local and only works when the curves can be written as graphs over the x-axis.

Prove that if two curves have the same unit tangent and one lies more to the left than the other then it also has larger (or equal) curvature.

Prove that if a curve lies inside another curve and they are tangent at a point, then it also lies more to the left of the other curve.

2.6. Convex Curves

Definition 2.6.1. We say that a regular planar curve is *convex* if it always lies on one side of its tangent lines. We say that it is *strictly convex* if it only intersects its tangent lines at the point of contact. A closed strictly convex curve is also called an *oval*.

Note that we do not need to assume that the curve is closed for this definition to make sense.

Theorem 2.6.2. *A planar convex curve is simple and the signed curvature cannot change sign.*

Proof. First we show that the curvature can't change sign. We assume that the curve $\mathbf{q} : [0, L] \to \mathbb{R}^2$ is parametrized by arclength. Since the curve lies on one side of its tangent at any point $\mathbf{q}(s_0)$ it follows that

$$(\mathbf{q}(s) - \mathbf{q}(s_0)) \cdot \mathbf{N}_\pm(s_0)$$

is either non-negative or nonpositive for all s. If it vanishes, then the curve must be part of the tangent line through $\mathbf{q}(s_0)$. In this case it is clearly simple and the curvature vanishes. Otherwise we have two disjoint sets $I_\pm \subset [0, L]$, where

$$\begin{aligned} I_+ &= \{s_0 \in [0, L] \mid (\mathbf{q}(s) - \mathbf{q}(s_0)) \cdot \mathbf{N}_\pm(s_0) \geq 0 \text{ for all } s \in [0, L]\} \\ I_- &= \{s_0 \in [0, L] \mid (\mathbf{q}(s) - \mathbf{q}(s_0)) \cdot \mathbf{N}_\pm(s_0) \leq 0 \text{ for all } s \in [0, L]\} \end{aligned}$$

Both of these sets must be closed by the continuity of $(\mathbf{q}(s)-\mathbf{q}(s_0))\cdot\mathbf{N}_{\pm}(s_0)$. However, it is not possible to write an interval as the disjoint union of two closed sets unless one of these sets is empty.

Now assume that $I_+=[0,L]$. Thus $(\mathbf{q}(s)-\mathbf{q}(s_0))\cdot\mathbf{N}_{\pm}(s_0)\geq 0$ for all s,s_0 with equality for $s=s_0$. Then the second derivative with respect to s is also non-negative at s_0:

$$\begin{aligned} 0 &\leq \frac{d^2\mathbf{q}}{ds^2}(s_0)\cdot\mathbf{N}_{\pm}(s_0) \\ &= \frac{d\mathbf{T}}{ds}(s_0)\cdot\mathbf{N}_{\pm}(s_0) \\ &= \kappa_{\pm}(s_0). \end{aligned}$$

This shows that the signed curvature is always non-negative.

We still assume that the curve always lies to the left of its oriented tangent lines. If $\mathbf{q}(s_0)=\mathbf{q}(s_1)$, then the tangent lines must agree at this point as the curve otherwise can't be smooth. The unit tangents must also agree since the entire curve always lies to the left of its oriented tangents. Next we observe that if some nearby point $\mathbf{q}(s_0+\epsilon)$ lies on the common tangent line, then the curve must lie on the tangent line for all $s\in[s_0,s_0+\epsilon]$ as $\mathbf{q}(s_0+\epsilon)$ would otherwise lie to the right of one of the tangent lines through $\mathbf{q}(s)$. The same reasoning now shows that $\mathbf{q}(s)$ must lie on the tangent line for $s\in[s_1,s_1+\epsilon]$ as $\mathbf{q}(s_0+\epsilon)$ would otherwise lie to the right of one of the tangent lines through $\mathbf{q}(s)$. This means we can suppose that s_0,s_1 are chosen so that $\mathbf{q}(s_i+t)$ lies strictly to the left of the tangent line for all small $t\in[0,\epsilon]$. In other words the total curvature over $[s_i,s_i+\epsilon]$ is positive. This implies that the tangent line through $\mathbf{q}(s_0+\epsilon_0)$ for small ϵ_0 is always parallel to a tangent line through $\mathbf{q}(s_1+\epsilon_1)$ for some small ϵ_1. These tangent lines must agree as otherwise the curve couldn't be to the left of both tangent lines. This common tangent line is by construction not parallel to the tangent line through $\mathbf{q}(s_i)$. In particular, if $\mathbf{q}(s_0+\epsilon_0)\neq\mathbf{q}(s_1+\epsilon_1)$, then one of these points will be closer to the tangent line through $\mathbf{q}(s_i)$. Assume $\mathbf{q}(s_1+\epsilon_1)$ is closer. Then it follows that $\mathbf{q}(s_1+\epsilon_1)$ must lie to the right of the tangent lines through $\mathbf{q}(s_0+s)$ for $s<\epsilon_0$ close to ϵ_0. This shows that $\mathbf{q}(s_0+\epsilon_0)=\mathbf{q}(s_1+\epsilon_1)$ and that the curve is simple. □

THEOREM 2.6.3. *If a curve has non-negative signed curvature and total curvature $\leq\pi$, then it is convex.*

PROOF. Any curve with non-negative curvature always locally lies on the left of its tangent lines. So if it comes back to intersect a tangent l after having travelled to the left of l, then there will be a point of locally maximal distance to the left of l. At this local maximum the tangent line l^* must be parallel but not equal to l. If they are oriented in the same direction, then the curve will locally be on the right of l^*. As that does not happen they have opposite direction. This shows that the total curvature is $\geq\pi$. However, the curve will have strictly larger total curvature as it still has to make its way back to intersect l. □

THEOREM 2.6.4. *If a closed curve has non-negative signed curvature and total curvature $\leq 2\pi$, then it is convex.*

PROOF. The argument is similar to the one above. Assume that we have a tangent line l such that the curve lies on both sides of this line. As the curve is closed there'll be points one both sides of this tangent at maximal distance from the tangent. The tangent lines l^* and l^{**} at these points are then parallel to l. Thus we have three parallel tangent lines that are not equal. Two of these must correspond to unit tangents that point in the same direction. As the curvature does not change sign this implies that the total curvature of part of the curve is 2π. The total curvature must then be $> 2\pi$ as these two tangent lines are different and the curve still has to return to both of the points of contact. □

EXAMPLE 2.6.5. Euler's construction. The details are discussed in the exercises below. A piecewise smooth simple closed planar curve with $n > 2$ cusps and the property that tangents at different points are never parallel has the property that its involutes are curves of constant width. The cusps are the points where the curve is not smooth and we assume that the unit tangents are opposite at those points, i.e., the interior angles are zero at the non-smooth points.

Exercises.

(1) Let $\mathbf{q}(\theta)$ be a simple closed planar curve with $\kappa > 0$ parametrized by θ, where θ is defined as the arclength parameter of the unit tangent field $\mathbf{T}$. Show that

$$\begin{aligned} \frac{d\mathbf{q}}{d\theta} &= \frac{1}{\kappa}\mathbf{T}, \\ \frac{d\mathbf{T}}{d\theta} &= \mathbf{N}, \\ \frac{d\mathbf{N}}{d\theta} &= -\mathbf{T}, \\ \mathbf{T}(\theta+\pi) &= -\mathbf{T}(\theta). \end{aligned}$$

(2) Let $\mathbf{q}(\theta)$ be a simple closed planar curve with $\kappa > 0$ parametrized by θ, where θ is defined as the arclength parameter of the unit tangent field $\mathbf{T}$. Define $v(\theta)$ as the distance from the origin to the tangent line through $\mathbf{q}(\theta)$.

(a) Show that
$$v(\theta) = -\mathbf{q}(\theta)\cdot\mathbf{N}(\theta).$$

(b) Show that the width (distance) between the parallel tangent lines through $\mathbf{q}(\theta)$ and $\mathbf{q}(\theta+\pi)$ is
$$w(\theta) = v(\theta) + v(\theta+\pi) = \mathbf{N}(\theta)\cdot(\mathbf{q}(\theta+\pi) - \mathbf{q}(\theta)).$$

(c) Show that:
$$L(\mathbf{q}) = \int_0^{2\pi} v(\theta)\,d\theta.$$

(d) Show that
$$\frac{1}{\kappa} = v + \frac{d^2v}{d\theta^2}.$$

(e) Let A denote the area enclosed by the curve. Establish the following formulas for A
$$A = \frac{1}{2}\int_0^L v\,ds = \frac{1}{2}\int_0^{2\pi}\left(v^2 + v\frac{d^2v}{d\theta^2}\right)d\theta = \frac{1}{2}\int_0^{2\pi}\left(v^2 - \left(\frac{dv}{d\theta}\right)^2\right)d\theta.$$

(3) Let $\mathbf{q}(\theta)$ be a simple closed planar curve with $\kappa > 0$ parametrized by θ, where θ is defined as the arclength parameter of the unit tangent field $\mathbf{T}$. Show that the width from the previous problem satisfies:

$$\frac{d^2 w}{d\theta^2} + w = \frac{1}{\kappa(\theta)} + \frac{1}{\kappa(\theta + \pi)}.$$

(4) Let $\mathbf{q}(\theta)$ be a simple closed planar curve with $\kappa > 0$ parametrized by θ, where θ is defined as the arclength parameter of the unit tangent field $\mathbf{T}$. With the width defined as in the previous exercises show that:

$$\int_0^{2\pi} w d\theta = 2L(\mathbf{q}).$$

(5) Let $\mathbf{q}(\theta)$ be a simple closed planar curve of constant width with $\kappa > 0$. The curve is parametrized by θ, where θ is defined as the arclength parameter of the unit tangent field $\mathbf{T}$.
   (a) Show that if θ corresponds to a local maximum for κ, then the opposite point $\theta + \pi$ corresponds to a local minimum.
   (b) Assume for the remainder of the exercise that κ has a finite number of critical points and that they are all local maxima or minima. Show that the number of vertices is even and ≥ 6.
   (c) Show that each point on the evolute corresponds to two points on the curve.
   (d) Show that the evolute consists of n convex curves that are joined at n cusps that correspond to pairs of vertices on the curve.
   (e) Show that the evolute has no double tangents.

(6) (Euler) Reverse the construction in the previous exercise to create curves of constant width by taking involutes of suitable curves.

(7) Let $\mathbf{q}$ be a closed convex curve and l a line.
   (a) Show that l can only intersect $\mathbf{q}$ in one point, two points, or a line segment.
   (b) Show that if l is also a tangent line then it cannot intersect $\mathbf{q}$ in only two points.
   (c) Show that the interior of $\mathbf{q}$ is convex, i.e., the segment between any two points in the interior also lies in the interior.

(8) Let $\mathbf{q}$ be a planar curve with non-negative signed curvature. Show that if $\mathbf{q}$ has a double tangent, then its total curvature is $\geq 2\pi$. Note that it is possible for the double tangent to have opposite directions at the points of tangency.

(9) Give an example of a planar curve (not closed) with positive curvature and no double tangents that is not convex.

(10) Let $\mathbf{q}$ be a closed planar curve without double tangents. Show that $\mathbf{q}$ is convex. Hint: Consider the set A of points (parameter values) on $\mathbf{q}$ where $\mathbf{q}$ lies on one side of the tangent line. Show that A is closed and not empty. Show that boundary points of A (i.e., points in A that are limit points of sequences in the complement of A) correspond to double tangents.

CHAPTER 3

Space Curves

3.1. The Fundamental Equations

The theory of space curves dates back to Clairaut in 1731. He considered them as the intersection of two surfaces given by equations. Clairaut showed that space curves have two curvatures, but they did not corresponds exactly to the curvature and torsion we introduce below. The subject was later taken up by Euler who was the first to work with parametrized curves and use arclength as a parameter. Lancret in 1806 introduced the concepts of unit tangent, principal normal and bi-normal and with those curvature and torsion as we now understand them. It is possible that Monge had some inklings of what torsion was, but he never presented an explicit formula. Cauchy in 1826 considerably modernized the subject and formulated some of the relations that later became part of the Serret and Frenet equations that we shall introduce below.

In order to create a set of equations for space curves $\mathbf{q}(t) : [a,b] \to \mathbb{R}^3$ we need to not only assume that the curve is regular but also that its velocity and acceleration are always linearly independent. This is equivalent to assuming that $\mathbf{q}$ is regular and that the unit tangent $\mathbf{T}$ also defines a regular curve, i.e., that the curvature never vanishes. In this case it is possible to define a suitable positively oriented orthonormal frame $\mathbf{T}$, $\mathbf{N}$, and $\mathbf{B}$ by declaring

$$\begin{aligned}
\mathbf{T} &= \frac{\mathbf{v}}{|\mathbf{v}|}, \\
\mathbf{N} &= \frac{\mathbf{a} - (\mathbf{a} \cdot \mathbf{T})\,\mathbf{T}}{|\mathbf{a} - (\mathbf{a} \cdot \mathbf{T})\,\mathbf{T}|}, \\
\mathbf{B} &= \mathbf{T} \times \mathbf{N}.
\end{aligned}$$

The new normal vector $\mathbf{B}$ is called the *bi-normal*. We define the *curvature* and *torsion* by

$$\begin{aligned}
\kappa &= \mathbf{N} \cdot \frac{d\mathbf{T}}{ds}, \\
\tau &= \mathbf{B} \cdot \frac{d\mathbf{N}}{ds}.
\end{aligned}$$

We should check that these definitions for $\mathbf{N}$ and κ are consistent with our earlier definitions. In section 1.3 we started by defining θ as arclength parameter for $\mathbf{T}$ and then proceeded to show that the above formulas for κ and $\mathbf{N}$ hold. So it is a question of checking that our new definitions conversely imply the old ones. We'll do this after having established the next theorem using the new definitions.

THEOREM 3.1.1. *(Serret, 1851 and Frenet, 1852) If $\mathbf{q}(t)$ is a regular space curve with linearly independent velocity and acceleration, then*

$$\begin{aligned}
\frac{d\mathbf{q}}{dt} &= \frac{ds}{dt}\mathbf{T}, \\
\frac{d\mathbf{T}}{dt} &= \kappa\frac{ds}{dt}\mathbf{N}, \\
\frac{d\mathbf{N}}{dt} &= -\kappa\frac{ds}{dt}\mathbf{T} + \tau\frac{ds}{dt}\mathbf{B}, \\
\frac{d\mathbf{B}}{dt} &= -\tau\frac{ds}{dt}\mathbf{N},
\end{aligned}$$

or

$$\frac{d}{dt}\begin{bmatrix} \mathbf{q} & \mathbf{T} & \mathbf{N} & \mathbf{B} \end{bmatrix} = \frac{ds}{dt}\begin{bmatrix} \mathbf{q} & \mathbf{T} & \mathbf{N} & \mathbf{B} \end{bmatrix}\begin{bmatrix} 0 & 0 & 0 & 0 \\ 1 & 0 & -\kappa & 0 \\ 0 & \kappa & 0 & -\tau \\ 0 & 0 & \tau & 0 \end{bmatrix}.$$

Moreover,

$$\begin{aligned}
\kappa &= \frac{|\mathbf{v}\times\mathbf{a}|}{|\mathbf{v}|^3} = \frac{|\mathbf{a}-(\mathbf{a}\cdot\mathbf{T})\mathbf{T}|}{|\mathbf{v}|^2}, \\
\tau &= \frac{\det\begin{bmatrix} \mathbf{v} & \mathbf{a} & \mathbf{j} \end{bmatrix}}{|\mathbf{v}\times\mathbf{a}|^2} = \frac{(\mathbf{v}\times\mathbf{a})\cdot\mathbf{j}}{|\mathbf{v}\times\mathbf{a}|^2}, \\
\mathbf{N} &= \frac{\mathbf{a}-(\mathbf{a}\cdot\mathbf{T})\mathbf{T}}{|\mathbf{a}-(\mathbf{a}\cdot\mathbf{T})\mathbf{T}|}, \\
\mathbf{B} &= \frac{\mathbf{v}\times\mathbf{a}}{|\mathbf{v}\times\mathbf{a}|}.
\end{aligned}$$

PROOF. The explicit formula for $\mathbf{N}$ is our explicit formula for the principal normal. As $\mathbf{T}, \mathbf{N}, \mathbf{B}$ form an orthonormal basis we have

$$\frac{d\mathbf{T}}{dt} = \left(\frac{d\mathbf{T}}{dt}\cdot\mathbf{T}\right)\mathbf{T} + \left(\frac{d\mathbf{T}}{dt}\cdot\mathbf{N}\right)\mathbf{N} + \left(\frac{d\mathbf{T}}{dt}\cdot\mathbf{B}\right)\mathbf{B}.$$

Here

$$\begin{gathered}
\frac{d\mathbf{T}}{dt}\cdot\mathbf{T} = \frac{1}{2}\frac{d|\mathbf{T}|^2}{dt} = 0, \\
\frac{d\mathbf{T}}{dt}\cdot\mathbf{N} = \frac{ds}{dt}\frac{d\mathbf{T}}{ds}\cdot\mathbf{N} = \frac{ds}{dt}\kappa, \\
\frac{d\mathbf{T}}{dt}\cdot\mathbf{B} = \frac{d}{dt}\left(\frac{\mathbf{v}}{|\mathbf{v}|}\right)\cdot\mathbf{B} = \left(\frac{\mathbf{a}}{|\mathbf{v}|} - \frac{\mathbf{v}}{|\mathbf{v}|^2}\frac{d|\mathbf{v}|}{dt}\right)\cdot\mathbf{B} = 0,
\end{gathered}$$

as $\mathbf{B}$ is perpendicular to $\mathbf{T}, \mathbf{N}$ and thus also to $\mathbf{v}, \mathbf{a}$. The establishes

$$\frac{d\mathbf{T}}{dt} = \kappa\frac{ds}{dt}\mathbf{N}.$$

Next note that

$$0 = \mathbf{B}\cdot\frac{d\mathbf{T}}{dt} = -\frac{d\mathbf{B}}{dt}\cdot\mathbf{T}.$$

This together with

$$\mathbf{B}\cdot\frac{d\mathbf{B}}{dt} = 0$$

shows that

$$\frac{d\mathbf{B}}{dt} = \left(\frac{d\mathbf{B}}{dt}\cdot\mathbf{N}\right)\mathbf{N}.$$

However, we also have

$$0 = \frac{d\mathbf{B}}{dt}\cdot\mathbf{N} + \mathbf{B}\cdot\frac{d\mathbf{N}}{dt} = \frac{d\mathbf{B}}{dt}\cdot\mathbf{N} + \frac{ds}{dt}\tau.$$

This implies

$$\frac{d\mathbf{B}}{dt} = -\tau\frac{ds}{dt}\mathbf{N}.$$

Finally the equation

$$\frac{d\mathbf{N}}{dt} = -\kappa\frac{ds}{dt}\mathbf{T} + \tau\frac{ds}{dt}\mathbf{B}.$$

is a direct consequence of the other two equations.

The formula for the curvature follows from observing that

$$\begin{aligned}
\frac{d\mathbf{T}}{ds}\cdot\mathbf{N} &= \left(\frac{\mathbf{a}}{|\mathbf{v}|} - \frac{\mathbf{v}}{|\mathbf{v}|^2}\frac{d\,|\mathbf{v}|}{ds}\right)\cdot\mathbf{N} \\
&= \frac{\mathbf{a}}{|\mathbf{v}|}\cdot\frac{\mathbf{a}-(\mathbf{a}\cdot\mathbf{T})\,\mathbf{T}}{|\mathbf{a}-(\mathbf{a}\cdot\mathbf{T})\,\mathbf{T}|} \\
&= \frac{\mathbf{a}\cdot\mathbf{a}-(\mathbf{a}\cdot\mathbf{T})^2}{|\mathbf{v}|\,|\mathbf{a}-(\mathbf{a}\cdot\mathbf{T})\,\mathbf{T}|} \\
&= \frac{|\mathbf{a}|^2\,|\mathbf{v}|^2-(\mathbf{a}\cdot\mathbf{v})^2}{|\mathbf{v}|^3\,|\mathbf{a}-(\mathbf{a}\cdot\mathbf{T})\,\mathbf{T}|} \\
&= \frac{|\mathbf{a}-(\mathbf{a}\cdot\mathbf{T})\,\mathbf{T}|}{|\mathbf{v}|^2},
\end{aligned}$$

where $|\mathbf{v}|\,|\mathbf{a}-(\mathbf{a}\cdot\mathbf{T})\,\mathbf{T}| = \sqrt{|\mathbf{a}|^2\,|\mathbf{v}|^2-(\mathbf{a}\cdot\mathbf{v})^2}$.

The formula for the binormal $\mathbf{B}$ now follows directly from the calculation

$$\begin{aligned}
\mathbf{T}\times\mathbf{N} &= \frac{1}{|\mathbf{v}|}\mathbf{v}\times\left(\frac{\mathbf{a}-(\mathbf{a}\cdot\mathbf{T})\,\mathbf{T}}{|\mathbf{a}-(\mathbf{a}\cdot\mathbf{T})\,\mathbf{T}|}\right) \\
&= \frac{1}{|\mathbf{v}|}\mathbf{v}\times\left(\frac{\mathbf{a}}{|\mathbf{a}-(\mathbf{a}\cdot\mathbf{T})\,\mathbf{T}|}\right) \\
&= \frac{\mathbf{v}\times\mathbf{a}}{|\mathbf{v}|\,|\mathbf{a}-(\mathbf{a}\cdot\mathbf{T})\,\mathbf{T}|} \\
&= \frac{\mathbf{v}\times\mathbf{a}}{|\mathbf{v}\times\mathbf{a}|}.
\end{aligned}$$

In the last equality recall that the denominators are the areas of the same parallelogram spanned by $\mathbf{v}$ and $\mathbf{a}$.

To establish the general formula for τ we note (with more explanations to follow)

$$\begin{aligned}\mathbf{B}\cdot\frac{d\mathbf{N}}{dt} &= \frac{\mathbf{v}\times\mathbf{a}}{|\mathbf{v}\times\mathbf{a}|}\cdot\frac{d}{dt}\left(\frac{\mathbf{a}-(\mathbf{a}\cdot\mathbf{T})\,\mathbf{T}}{|\mathbf{a}-(\mathbf{a}\cdot\mathbf{T})\,\mathbf{T}|}\right)\\ &= \frac{\mathbf{v}\times\mathbf{a}}{|\mathbf{v}\times\mathbf{a}|}\cdot\frac{\mathbf{j}}{|\mathbf{a}-(\mathbf{a}\cdot\mathbf{T})\,\mathbf{T}|}\\ &= \frac{(\mathbf{v}\times\mathbf{a})\cdot\mathbf{j}}{|\mathbf{v}\times\mathbf{a}|^2}\,|\mathbf{v}|\,.\end{aligned}$$

In the third line all of the missing terms disappear as they are perpendicular to $\mathbf{v}\times\mathbf{a}$. The last line follows from our formulas for the area of the parallelogram spanned by $\mathbf{v}$ and $\mathbf{a}$. A slightly more convincing proof works by first noticing that

$$\begin{aligned}\mathbf{v} &= (\mathbf{v}\cdot\mathbf{T})\,\mathbf{T},\\ \mathbf{a} &= (\mathbf{a}\cdot\mathbf{T})\,\mathbf{T}+(\mathbf{a}\cdot\mathbf{N})\,\mathbf{N},\\ \mathbf{j} &= (\mathbf{j}\cdot\mathbf{T})\,\mathbf{T}+(\mathbf{j}\cdot\mathbf{N})\,\mathbf{N}+(\mathbf{j}\cdot\mathbf{B})\,\mathbf{B}.\end{aligned}$$

Thus

$$\det\begin{bmatrix}\mathbf{v} & \mathbf{a} & \mathbf{j}\end{bmatrix} = (\mathbf{v}\cdot\mathbf{T})\,(\mathbf{a}\cdot\mathbf{N})\,(\mathbf{j}\cdot\mathbf{B})\,.$$

Next we recall that

$$\begin{aligned}\mathbf{v}\cdot\mathbf{T} &= |\mathbf{v}|\,,\\ \mathbf{a}\cdot\mathbf{N} &= |\mathbf{v}|^2\,\kappa.\end{aligned}$$

So we have to calculate $\mathbf{j}\cdot\mathbf{B}$. Keeping in mind that $\mathbf{a}\cdot\mathbf{B}=0$ we obtain

$$\begin{aligned}\mathbf{j}\cdot\mathbf{B} &= -\mathbf{a}\cdot\frac{d\mathbf{B}}{dt}\\ &= \tau\,|\mathbf{v}|\,\mathbf{a}\cdot\mathbf{N}\\ &= \tau\kappa\,|\mathbf{v}|^3\end{aligned}$$

and finally combine this with

$$\kappa = \frac{|\mathbf{v}\times\mathbf{a}|}{|\mathbf{v}|^3}$$

to obtain the desired identity. □

The curvature and torsion can also be described by the formulas

$$\kappa = \frac{\text{area of parallelogram }(\mathbf{v},\mathbf{a})}{|\mathbf{v}|^3},$$

$$\tau = \frac{\text{signed volume of the parallepiped }(\mathbf{v},\mathbf{a},\mathbf{j})}{(\text{area of the parallelogram }(\mathbf{v},\mathbf{a}))^2}.$$

Corollary 3.1.2. *If $\mathbf{q}(t)$ is a regular space curve with linearly independent velocity and acceleration, then $\mathbf{T}$ is regular and if θ is its arclength parameter, then*

$$\frac{d\theta}{ds} = \frac{d\mathbf{T}}{ds}\cdot\mathbf{N}$$

and

$$\frac{d\mathbf{T}}{d\theta} = \frac{\mathbf{a}-(\mathbf{a}\cdot\mathbf{T})\,\mathbf{T}}{|\mathbf{a}-(\mathbf{a}\cdot\mathbf{T})\,\mathbf{T}|}.$$

PROOF. By assumption

$$0 < \kappa = \frac{d\mathbf{T}}{ds} \cdot \mathbf{N}.$$

This implies in particular that $\mathbf{T}$ is regular. We know from the chain rule that

$$\frac{d\mathbf{T}}{d\theta} = \frac{ds}{d\theta}\frac{d\mathbf{T}}{ds} = \frac{ds}{d\theta}\kappa\mathbf{N}.$$

Here both sides are unit vectors that are perpendicular to $\mathbf{T}$ and by definition $\frac{ds}{d\theta} > 0$ and $\kappa > 0$. This forces

$$\frac{d\theta}{ds} = \kappa$$

and

$$\frac{d\mathbf{T}}{d\theta} = \mathbf{N}.$$

This establishes the formulas □

There is a very elegant way of collecting the Serret-Frenet formulas.

COROLLARY 3.1.3. *(Darboux) For a space curve as above define the Darboux vector*

$$\mathbf{D} = \tau\mathbf{T} + \kappa\mathbf{B},$$

then

$$\frac{d}{dt}\begin{bmatrix} \mathbf{T} & \mathbf{N} & \mathbf{B} \end{bmatrix} = \frac{ds}{dt}\mathbf{D} \times \begin{bmatrix} \mathbf{T} & \mathbf{N} & \mathbf{B} \end{bmatrix}.$$

PROOF. We have

$$\begin{aligned} \mathbf{D} \times \mathbf{T} &= \kappa\mathbf{N}, \\ \mathbf{D} \times \mathbf{N} &= \tau\mathbf{B} - \kappa\mathbf{T}, \\ \mathbf{D} \times \mathbf{B} &= -\tau\mathbf{N}, \end{aligned}$$

so the equation follows directly from the Serret-Frenet formulas. □

Exercises.

(1) Find the curvature, torsion, normal, and binormal for the twisted cubic

$$\mathbf{q}(t) = \left(t, t^2, t^3\right).$$

(2) Consider a regular space curve $\mathbf{q}(t)$ with non-vanishing curvature and torsion. Let $\mathbf{k}$ be a fixed vector and denote by $\phi_{\mathbf{T}}$, $\phi_{\mathbf{N}}$, $\phi_{\mathbf{B}}$ the angles between $\mathbf{T}$, $\mathbf{N}$, $\mathbf{B}$ and $\mathbf{k}$. Show that

$$\begin{aligned} \kappa &= -\frac{\sin\phi_{\mathbf{T}}}{\cos\phi_{\mathbf{N}}}\frac{d\phi_{\mathbf{T}}}{dt}, \\ \frac{d\phi_{\mathbf{N}}}{dt}\sin\phi_{\mathbf{N}} &= \kappa\cos\phi_{\mathbf{T}} - \tau\cos\phi_{\mathbf{B}}, \\ \tau &= \frac{\sin\phi_{\mathbf{B}}}{\cos\phi_{\mathbf{N}}}\frac{d\phi_{\mathbf{B}}}{dt}, \end{aligned}$$

and

$$\frac{d\phi_{\mathbf{B}}}{dt}\sin\phi_{\mathbf{B}} = -\frac{\tau}{\kappa}\frac{d\phi_{\mathbf{T}}}{dt}\sin\phi_{\mathbf{T}}.$$

(3) Consider a cylindrical curve of the form

$$\mathbf{q}(\theta) = (\cos\theta, \sin\theta, z(\theta)).$$

Show that

$$\kappa = \frac{\left(1 + (z')^2 + (z'')^2\right)^{\frac{1}{2}}}{\left(1 + (z')^2\right)^{\frac{3}{2}}},$$

$$\tau = \frac{z' + z'''}{1 + (z')^2 + (z'')^2}.$$

(4) Show that for a unit speed curve $\mathbf{q}(s)$ with positive curvature

$$\det\left[\begin{array}{ccc} \mathbf{a} & \mathbf{j} & \frac{d^4\mathbf{q}}{ds^4} \end{array}\right] = \kappa^5 \frac{d}{ds}\left(\frac{\tau}{\kappa}\right).$$

(5) For a unit speed curve $\mathbf{q}(s)$ with positive curvature and torsion define $\mathbf{q}^*(s) = \int \mathbf{B}(s)\,ds$. Show that $\mathbf{q}^*$ is also unit speed and that $\mathbf{T}^* = \mathbf{B}$, $\mathbf{N}^* = -\mathbf{N}$, $\mathbf{B}^* = \mathbf{T}$, $\kappa^* = \tau$, and $\tau^* = \kappa$.

(6) Let $\mathbf{q}(t) : I \to \mathbb{R}^3$ be a regular curve such that its tangent field $\mathbf{T}(t)$ is also regular. Let s be the arclength parameter for $\mathbf{q}$ and θ the arclength parameter for $\mathbf{T}$. Show that

$$\det\left[\begin{array}{ccc} \mathbf{T} & \frac{d\mathbf{T}}{d\theta} & \frac{d^2\mathbf{T}}{d\theta^2} \end{array}\right] = \frac{\tau}{\kappa}.$$

(7) Show that $\mathbf{T}$ is regular when $\kappa > 0$ and that in this case the curvature of $\mathbf{T}$ is given by

$$\sqrt{1 + \left(\frac{\tau}{\kappa}\right)^2}$$

and the torsion by

$$\frac{1}{\kappa\left(1 + \left(\frac{\tau}{\kappa}\right)^2\right)} \frac{d}{ds}\left(\frac{\tau}{\kappa}\right).$$

(8) Show that the circular helix

$$(R\cos t, R\sin t, ht)$$

has constant curvature and torsion. Compute R, h in terms of the curvature and torsion. Conversely show that any unit speed space curve with constant curvature and torsion must look like

$$\mathbf{q}(s) = R\cos\left(\frac{s}{\sqrt{R^2 + h^2}}\right)e_1 + R\sin\left(\frac{s}{\sqrt{R^2 + h^2}}\right)e_2 + \frac{h}{\sqrt{R^2 + h^2}}se_3 + \mathbf{q}(0),$$

where e_1, e_2, e_3 is an orthonormal basis.

(9) Let $\mathbf{q}(s) = (x(s), y(s), z(s)) : [0, L] \to \mathbb{R}^3$ be a unit speed space curve with curvature $\kappa(s)$ and torsion $\tau(s)$. Construct another space curve $\mathbf{q}^*(s) = x(s)e_1 + y(s)e_2 + z(s)e_3 + \mathbf{x}$, where e_1, e_2, e_3 is a positively oriented orthonormal basis and $\mathbf{x}$ and point.

(a) Show that $\mathbf{q}^*$ is a unit speed curve with curvature $\kappa^*(s) = \kappa(s)$ and torsion $\tau^*(s) = \tau(s)$.

(b) Show that a unit speed space curve with the same curvature and torsion as $\mathbf{q}$ is of the form $\mathbf{q}^*$.

(10) Show that $\mathbf{B}$ is regular when $|\tau| > 0$ and that in this case the curvature of $\mathbf{B}$ is given by

$$\sqrt{1+\left(\frac{\kappa}{\tau}\right)^2}.$$

(11) Show that for a unit speed curve $\mathbf{q}(s)$ with positive curvature and non-zero torsion

$$\det\left[\begin{array}{ccc} \frac{d\mathbf{B}}{ds} & \frac{d^2\mathbf{B}}{ds^2} & \frac{d^3\mathbf{B}}{ds^3} \end{array}\right] = \tau^5 \frac{d}{ds}\left(\frac{\kappa}{\tau}\right).$$

(12) Show that $\mathbf{N}$ is regular when $\kappa^2+\tau^2 > 0$ and that in this case the curvature of $\mathbf{N}$ is given by

$$\sqrt{1+\frac{\left(\kappa\frac{d\tau}{ds}-\tau\frac{d\kappa}{ds}\right)^2}{\left(\kappa^2+\tau^2\right)^3}}.$$

(13) Define $\rho = \sqrt{\kappa^2+\tau^2}$ and ϕ by

$$\kappa = \rho\cos\phi,\ \tau = \rho\sin\phi.$$

Show that $\rho = |\mathbf{D}|$ and that ϕ is the natural arclength parameter for the unit field $\frac{1}{\rho}\mathbf{D}$.

(14) Show that a space curve is part of a line if all its tangent lines pass through a fixed point.

(15) Let $\mathbf{Q}(t)$ be a vector associated to a curve $\mathbf{q}(t)$ such that

$$\frac{d}{dt}\left[\begin{array}{ccc} \mathbf{T} & \mathbf{N} & \mathbf{B} \end{array}\right] = \frac{ds}{dt}\mathbf{Q}\times\left[\begin{array}{ccc} \mathbf{T} & \mathbf{N} & \mathbf{B} \end{array}\right].$$

Show that $\mathbf{Q} = \mathbf{D}$.

(16) Let $\mathbf{q}(s)$ be a unit speed space curve with non-vanishing curvature and torsion. Show that

$$\frac{d}{ds}\left(\frac{1}{\tau}\frac{d}{ds}\left(\frac{1}{\kappa}\frac{d^2\mathbf{q}}{ds^2}\right)\right)+\frac{d}{ds}\left(\frac{\kappa}{\tau}\frac{d\mathbf{q}}{ds}\right)+\frac{\tau}{\kappa}\frac{d^2\mathbf{q}}{ds^2} = 0.$$

(17) For a regular space curve $\mathbf{q}(t)$ we say that a normal field $\mathbf{X}$ is *parallel* along $\mathbf{q}$ if $\mathbf{X}\cdot\mathbf{T} = 0$ and $\frac{d\mathbf{X}}{dt}$ is parallel to $\mathbf{T}$.

(a) Show that for a fixed t_0 and $\mathbf{X}(t_0)\perp\mathbf{T}(s_0)$ there is a unique parallel field $\mathbf{X}$ that is $\mathbf{X}(t_0)$ at t_0.

(b) A *Bishop frame* consists of an orthonormal frame $\mathbf{T},\mathbf{N}_1,\mathbf{N}_2$ along the curve so that $\mathbf{N}_1,\mathbf{N}_2$ are both parallel along $\mathbf{q}$. For such a frame show that

$$\frac{d}{dt}\left[\begin{array}{ccc} \mathbf{T} & \mathbf{N}_1 & \mathbf{N}_2 \end{array}\right] = \frac{ds}{dt}\left[\begin{array}{ccc} \mathbf{T} & \mathbf{N}_1 & \mathbf{N}_2 \end{array}\right]\left[\begin{array}{ccc} 0 & \kappa_1 & \kappa_2 \\ -\kappa_1 & 0 & 0 \\ -\kappa_2 & 0 & 0 \end{array}\right].$$

Note that such frames always exist, even when the space curve doesn't have positive curvature everywhere.

(c) Show further that for such a frame

$$\kappa^2 = \kappa_1^2+\kappa_2^2.$$

(d) Show that if $\mathbf{q}$ has positive curvature so that $\mathbf{N}$ is well-defined, then

$$\mathbf{N} = \cos\phi\mathbf{N}_1+\sin\phi\mathbf{N}_2,$$

where

$$\frac{d\phi}{dt} = \frac{ds}{dt}\tau,$$
$$\kappa_1 = \kappa\cos\phi,\ \kappa_2 = \kappa\sin\phi.$$

(e) Give an example of a closed space curve where the parallel curves don't close up.

3.2. Characterizations of Space Curves

We show that the tangent lines determine a space curve, but that the (principal) normal lines do not necessarily characterize the curve.

THEOREM 3.2.1. *If* $\mathbf{q}(t)$ *and* $\mathbf{q}^*(t)$ *are two regular curves that admit a common parametrization such that their tangent lines agree at corresponding points, then* $\mathbf{q}(t) = \mathbf{q}^*(t)$ *for all* t *where either* $\kappa(t) \neq 0$ *or* $\kappa^*(t) \neq 0$.

PROOF. Note that the common parametrization is not necessarily the arclength parametrization for either curve. These arclength parametrizations are denoted s, s^*. The assumption implies that corresponding velocity vectors are always parallel and that

$$\mathbf{q}^*(t) = \mathbf{q}(t) + u(t)\,\mathbf{T}(t)$$

for some function $u(t)$. We obtain by differentiation

$$\frac{d\mathbf{q}^*}{dt} = \frac{d\mathbf{q}}{dt} + \frac{du}{dt}\mathbf{T} + u\frac{ds}{dt}\kappa\mathbf{N}.$$

This forces

$$u\frac{ds}{dt}\kappa = 0$$

as $\mathbf{N}$ is perpendicular to the other vectors. So whenever $\kappa \neq 0$ it follows that $u = 0$. This means that the curves agree on the set where $\kappa \neq 0$. Reversing the roles of the curves we similarly obtain that the curves agree when $\kappa^* \neq 0$. □

The analogous question for principal normal lines requires that these normal lines are defined and thus that the curvatures never vanish. Nevertheless it is easy to find examples of pairs of curves that have the same normal lines without being the same curve. The double helix is in fact a great example of this. This corresponds to the two pairs of circular helices

$$\mathbf{q} = (R\cos t, R\sin t, ht) \text{ and } \mathbf{q}^* = (-R\cos t, -R\sin t, ht)$$

More generally for fixed $h > 0$ all of the curves

$$(R\cos t, R\sin t, ht)$$

have the same normal lines for all $R \in \mathbb{R}$.

DEFINITION 3.2.2. We say that two curves $\mathbf{q}$ and $\mathbf{q}^*$ are *Bertrand mates* if it is possible to find a common parametrization of both curves such that their principal normal lines agree at corresponding points.

THEOREM 3.2.3. *Let* $\mathbf{q}$ *and* $\mathbf{q}^*$ *be Bertrand mates with non-zero curvatures and torsion. Then either the curves agree or there are linear relationships*

$$a\kappa + b\tau = 1,\ a\kappa^* - b\tau^* = 1$$

between curvature and torsion. Conversely any curve with non-zero curvature and torsion where $a\kappa + b\tau = 1$ *for some constants* a, b *has a Bertrand mate.*

Proof. We'll use s, s^* for the arclength of the two curves. That two curves are Bertrand mates is equivalent to

$$\mathbf{N}(t) = \pm\mathbf{N}^*(t)$$

and

$$\mathbf{q}^*(t) = \mathbf{q}(t) + r(t)\mathbf{N}(t)$$

for some function $r(t)$.

The first condition implies

$$\frac{d}{dt}(\mathbf{T}\cdot\mathbf{T}^*) = \frac{ds}{dt}\kappa\mathbf{N}\cdot\mathbf{T}^* + \frac{ds^*}{dt}\kappa^*\mathbf{T}\cdot\mathbf{N}^* = 0,$$

which shows that there is an angle θ such that

$$\mathbf{T}^*(t) = \mathbf{T}(t)\cos\theta + \mathbf{B}(t)\sin\theta.$$

Differentiating the second condition implies

$$\frac{ds^*}{dt}\mathbf{T}^* = \frac{ds}{dt}\mathbf{T} + \frac{dr}{dt}\mathbf{N} + r\frac{ds}{dt}(-\kappa\mathbf{T} + \tau\mathbf{B}).$$

Here $\mathbf{N}$ is perpendicular to all of the other vectors so it follows that r is constant. Note that when $r = 0$ the curves are equal and that in general the distance between the curves is given by $|r|$. In any case we obtain the relationship

$$\frac{ds^*}{dt}(\mathbf{T}(t)\cos\theta + \mathbf{B}(t)\sin\theta) = \frac{ds^*}{dt}\mathbf{T}^* = \frac{ds}{dt}\mathbf{T} + r\frac{ds}{dt}(-\kappa\mathbf{T} + \tau\mathbf{B}),$$

which implies

$$\frac{ds^*}{ds}\cos\theta = (1 - r\kappa)$$

and

$$\frac{ds^*}{ds}\sin\theta = r\tau.$$

When $r \neq 0$ the fact that $\tau \neq 0$ implies

$$-(1 - r\kappa)\sin\theta + r\tau\cos\theta = 0,$$

which shows

$$\kappa r + \tau r\cot\theta = 1.$$

Switching the roles force us to change the sign of θ. Thus

$$\mathbf{T}(t) = \mathbf{T}^*(t)\cos\theta - \mathbf{B}^*(t)\sin\theta$$

and

$$\kappa^* r - \tau^* r\cot\theta = 1.$$

Conversely assume that we have a regular curve $\mathbf{q}(s)$ parametrized by arclength so that

$$\kappa r + \tau r\cot\theta = 1.$$

Inspired by our conclusions from the first part of the proof we define

$$\mathbf{q}^*(s) = \mathbf{q}(s) + r\mathbf{N}(s)$$

and note that

$$\frac{d\mathbf{q}^*}{ds} = \mathbf{T} + r(-\kappa\mathbf{T} + \tau\mathbf{B}) = \tau r(\cot\theta\mathbf{T} + \mathbf{B}).$$

Thus $\mathbf{T}^* = \pm(\cos\theta\mathbf{T} + \sin\theta\mathbf{B})$. This shows that

$$\frac{d\mathbf{T}^*}{ds} = \pm(\kappa\cos\theta - \tau\sin\theta)\mathbf{N}$$

and in particular that $\mathbf{N}^* = \pm\mathbf{N}$. □

Exercises. All curves will be regular and have positive curvature.

(1) A curve is *planar* if there is a vector $\mathbf{k}$ such that $\mathbf{q}(t) \cdot \mathbf{k}$ is constant.
  (a) Show that this is equivalent to saying that the tangent $\mathbf{T}$ is always perpendicular to $\mathbf{k}$ and implies that all derivatives $\frac{d^k\mathbf{q}}{dt^k}$, $k \geq 1$ are perpendicular to $\mathbf{k}$.
  (b) Show that a curve is planar if and only if τ vanishes.

(2) Consider solutions to the second order equation
$$\mathbf{a} = F(\mathbf{q}, \mathbf{v}).$$
Show that all solutions are planar if $F(\mathbf{q}, \mathbf{v}) \in \operatorname{span}\{\mathbf{q}, \mathbf{v}\}$ for all vectors $\mathbf{q}, \mathbf{v}$. This happens in particular when the force field F is radial, i.e., F is proportional to position $\mathbf{q}$.

(3) Let $\mathbf{q}(t)$ and $\mathbf{q}^*(t)$ be two regular curves that admit a common parametrization such that their tangent lines are parallel at corresponding points.
  (a) Show that their normals and binormals are also parallel.
  (b) Show that
$$\frac{\kappa^*}{\kappa} = \frac{ds}{ds^*} = \frac{\tau^*}{\tau}.$$

(4) (Lancret, 1806) A *generalized helix* is a curve such that $\mathbf{T} \cdot \mathbf{k}$ is constant for some fixed vector $\mathbf{k}$.
  (a) Show that this is equivalent to the normal $\mathbf{N}$ always being perpendicular to $\mathbf{k}$, i.e., the unit tangent is planar. Note that since the unit tangent traces a curve on the sphere it has to lie in the intersection of the unit sphere and a plane, i.e., a latitude, and must in particular be a circle.
  (b) Show that a curve is a generalized helix if and only if the ratio τ/κ is constant.
  (c) Show that this is equivalent to assuming that the curvature of the unit tangent $\mathbf{T}$ is constant.
  (d) Show that this is equivalent to the torsion of $\mathbf{T}$ vanishing.

(5) Show that a unit speed circular helix has constant Darboux vector, and conversely that any unit speed curve with constant Darboux vector is a helix.

(6) Let $\mathbf{q}(t) = (x(t), y(t), z(t))$ be a generalized helix that lies on the cylinder $x^2 + y^2 = 1$.
  (a) Show that as long as $(x(t), y(t))$ is not stationary, then the curve can be parametrized as
$$\mathbf{q}(\phi) = (\cos\phi, \sin\phi, z(\phi)).$$
  (b) Using that parametrization compute the normal component of the acceleration
$$\mathbf{a} - \frac{\mathbf{a} \cdot \mathbf{v}}{|\mathbf{v}|^2}\mathbf{v}$$
and show that this vector can only stay perpendicular to vectors $\mathbf{k} = (0, 0, c)$ and in this case only when $z'' = 0$.

(c) Show that $(x(t), y(t))$ is never stationary. Hint: First show that it can't be stationary everywhere as it can't be a line parallel to the z-axis.

(d) Conclude that the original curve is a circular helix.

(7) Let $\mathbf{q}(t) = (x(t), y(t), z(t))$ be a generalized helix that lies on the cone $x^2 + y^2 = z^2$ with $z > 0$. Show that the planar curve $(x(t), y(t))$ forms a constant angle with the radial lines and conclude that it is either a radial line or can be reparametrized as a logarithmic spiral

$$(x(\phi), y(\phi)) = ae^{b\phi}(\cos\phi, \sin\phi).$$

Hint: Look at the previous exercise, but the calculations are more involved.

(8) A curve is *spherical* if it lies on some sphere. Show that a curve is spherical if and only if its normal planes all pass through some fixed point.

(9) Assume we have a unit speed spherical curve. If the center of the sphere is $\mathbf{c}$ and the radius R, then the curve must satisfy

$$|\mathbf{q}(s) - \mathbf{c}|^2 = R^2.$$

(a) Show that if a spherical curve has nowhere vanishing curvature, then

$$\begin{aligned}(\mathbf{q} - \mathbf{c}) \cdot \mathbf{N} &= -\frac{1}{\kappa}, \\ \tau(\mathbf{q} - \mathbf{c}) \cdot \mathbf{B} &= \frac{d}{ds}\left(\frac{1}{\kappa}\right).\end{aligned}$$

(b) Show that if both curvature and torsion are nowhere vanishing, then

$$\frac{1}{\kappa^2} + \left(\frac{1}{\tau}\frac{d}{ds}\left(\frac{1}{\kappa}\right)\right)^2 = R^2$$

and

$$\frac{\tau}{\kappa} + \frac{d}{ds}\left(\frac{1}{\tau}\frac{d}{ds}\left(\frac{1}{\kappa}\right)\right) = 0.$$

(10) Show that a unit speed curve on a sphere of radius R satisfies

$$\kappa \geq \frac{1}{R}.$$

(11) Conversely, show that if a curve $\mathbf{q}$ with nowhere vanishing curvature and torsion satisfies

$$\frac{\tau}{\kappa} + \frac{d}{ds}\left(\frac{1}{\tau}\frac{d}{ds}\left(\frac{1}{\kappa}\right)\right) = 0,$$

then

$$\frac{1}{\kappa^2} + \left(\frac{1}{\tau}\frac{d}{ds}\left(\frac{1}{\kappa}\right)\right)^2 = R^2$$

for some constant R. Furthermore show that

$$\mathbf{c}(s) = \mathbf{q} + \frac{1}{\kappa}\mathbf{N} + \frac{1}{\tau}\frac{d}{ds}\left(\frac{1}{\kappa}\right)\mathbf{B}$$

is constant and conclude that $\mathbf{q}$ lies on the sphere with center c and radius R.

(12) Prove that a unit speed curve $\mathbf{q}$ with non-zero curvature and torsion lies on a sphere if there are constants a, b such that

$$\kappa\left(a\cos\left(\int\tau ds\right)+b\sin\left(\int\tau ds\right)\right)=1.$$

Hint: Show

$$\frac{1}{\tau}\frac{d}{ds}\left(\frac{1}{\kappa}\right)=-a\sin\left(\int\tau ds\right)+b\cos\left(\int\tau ds\right)$$

and

$$\frac{\tau}{\kappa}=-\frac{d}{ds}\left(-a\sin\left(\int\tau ds\right)+b\cos\left(\int\tau ds\right)\right)$$

and use the previous exercise.

(13) Show that if a curve with constant curvature lies on a sphere then it is part of a circle, i.e., it is forced to be planar.

(14) Show that

$$\mathbf{q}^*(s)=\mathbf{q}+\frac{1}{\kappa}\mathbf{N}+\frac{1}{\kappa}\cot\left(\int\tau ds\right)\mathbf{B}$$

defines an evolute for $\mathbf{q}$. Hint: See remark 1.3.10.

(15) Show that a planar curve has infinitely many Bertrand mates.

(16) Let $\mathbf{q}, \mathbf{q}^*$ be two Bertrand mates.

(a) (Schell) Show that

$$\tau\tau^*=\frac{\sin^2\theta}{r^2}.$$

(b) (Mannheim) Show that

$$(1-r\kappa)(1+r\kappa^*)=\cos^2\theta.$$

(17) Consider a curve $\mathbf{q}(s)$ parametrized by arclength with positive curvature and non-vanishing torsion such that

$$\kappa r+\tau r\cot\theta=1,$$

i.e., there is a Bertrand mate.

(a) Show that the Bertrand mate is uniquely determined by r.

(b) Show that if $\mathbf{q}$ has two different Bertrand mates then it must be a generalized helix.

(c) Show that if a generalized helix has a Bertrand mate, then its curvature and torsion are constant, consequently it is a circular helix.

(18) Investigate properties of a pair of curves that have the same normal planes at corresponding points, i.e., their tangent lines are parallel.

(19) Investigate properties of a pair of curves that have the same binormal lines at corresponding points.

3.3. Closed Space Curves

We start by studying spherical curves. In fact any regular space curve generates a natural spherical curve, the unit tangent. We studied this for planar curves in section 2.4 where the unit tangent became a curve on a circle. In that case the length of the unit tangent curve can be interpreted as an integral of the curvature and it also measures how much the curve turns. When the planar curve is closed this turning necessarily has to be a multiple of 2π.

A regular spherical curve $\mathbf{q}(t) : I \to S^2(1)$ has an alternate set of equations that describe its properties. Instead of the principal normal it has a signed normal that is tangent to the sphere. If we note that $\mathbf{q}$ is also normal to the sphere, then the signed normal can be defined as the vector

$$\mathbf{S} = \mathbf{q} \times \mathbf{T}.$$

This leads to the a new set of equations

$$\begin{aligned}
\frac{d\mathbf{T}}{dt} &= \frac{ds}{dt}\left(\kappa_g \mathbf{S} - \mathbf{q}\right), \\
\frac{d\mathbf{S}}{dt} &= \frac{ds}{dt}\left(-\kappa_g \mathbf{T}\right), \\
\frac{d\mathbf{q}}{dt} &= \frac{ds}{dt}\mathbf{T},
\end{aligned}$$

where the *geodesic curvature* κ_g is defined as $\kappa_g = \frac{d\mathbf{T}}{ds} \cdot \mathbf{S}$. It measures how far a curve is from being a great circle as those curves have the property that $\frac{d\mathbf{T}}{ds} \cdot \mathbf{S} = 0$. The last equation is obvious by now. The first then follows from our definition of κ_g and the second from the other two.

There is also a Crofton formula for spherical curves where we count intersections with oriented great circles. An oriented great circle is uniquely determined by its corresponding North pole if we think in terms of the right hand rule. Thus intersections with the oriented great circle given by $\mathbf{x}$ can be counted as

$$n_{\mathbf{q}}(\mathbf{x}) = \left|\{t \mid \mathbf{x} \cdot \mathbf{q}(t) = 0\}\right|$$

and Crofton's formula becomes

$$\frac{1}{4}\int_{S^2} n_{\mathbf{q}}(\mathbf{x})\, d\mathbf{x} = L(\mathbf{q}).$$

A similar proof works in this case, but we offer an alternate proof using the above equations.

Any point $\mathbf{q}(t)$ on the curve, will clearly intersect all great circles going through that point. These great circles are in turn given by the points along the great circle that is the equator for $\mathbf{q}(t)$. This equator can be parametrized by

$$\mathbf{x}(\theta, t) = \cos(\theta)\,\mathbf{T}(t) + \sin(\theta)\,\mathbf{S}(t).$$

Thus the surface integral becomes

$$\begin{aligned}
\int_a^b \int_0^{2\pi} \left|\frac{d\mathbf{x}}{ds} \times \frac{d\mathbf{x}}{dt}\right| d\theta dt &= \int_a^b \int_0^{2\pi} \left|(-\sin(\theta)\mathbf{T} + \cos(\theta)\mathbf{S}) \times ((\kappa_g \mathbf{S} - \mathbf{q})\cos(\theta) - \kappa_g \mathbf{T}\sin(\theta))\right| \frac{ds}{dt} d\theta dt \\
&= \int_a^b \int_0^{2\pi} \left|(-\sin(\theta)\mathbf{T} + \cos(\theta)\mathbf{S}) \times (-\mathbf{q}\cos(\theta))\right| \frac{ds}{dt} d\theta dt \\
&= \int_0^{L(\mathbf{q})} \int_0^{2\pi} \left|(-\sin(\theta)\mathbf{T} + \cos(\theta)\mathbf{S}) \times (-\mathbf{q}\cos(\theta))\right| d\theta ds \\
&= \int_0^{L(\mathbf{q})} \int_0^{2\pi} \left|\left(\sin(\theta)\cos(\theta)\mathbf{T}(t) \times \mathbf{q} - \cos^2(\theta)\mathbf{S} \times \mathbf{q}\right)\right| d\theta ds \\
&= \int_0^{L(\mathbf{q})} \int_0^{2\pi} \sqrt{\sin^2(\theta)\cos^2(\theta) + \cos^4(\theta)} d\theta ds \\
&= \int_0^{L(\mathbf{q})} \int_0^{2\pi} |\cos(\theta)| \, d\theta ds \\
&= \int_0^{L(\mathbf{q})} 4 ds \\
&= 4L(\mathbf{q}).
\end{aligned}$$

Theorem 3.3.1. *(Fenchel, 1929) If* $\mathbf{q}$ *is a closed space curve, then*

$$\int \kappa ds \geq 2\pi$$

with equality holding only for simple planar convex curves.

Proof. Note that the total curvature is the length of the unit tangent. If the unit tangent field lies in a hemisphere with pole $\mathbf{x}$, i.e., $\mathbf{T} \cdot \mathbf{x} \geq 0$ for all s, then after integration we obtain

$$(\mathbf{q}(L) - \mathbf{q}(0)) \cdot \mathbf{x} \geq 0.$$

However, $\mathbf{q}(L) = \mathbf{q}(0)$ as the curve is closed. So it follows that $\mathbf{T} \cdot \mathbf{x} = 0$ for all s, i.e., the unit tangent is always perpendicular to $\mathbf{x}$ and hence the curve is planar.

This shows that if the curve is not planar, then the unit tangent never lies in a hemisphere. This in turn implies that the unit tangent must intersect all great circles in at least two points. In fact if it does not intersect a certain great circle, then it must lie in an open hemisphere. If it intersects a great circle exactly once, then it must lie on one side of it and be tangent to the great circle. By moving the great circle slightly away from the point of tangency we obtain a new great circle that does not intersect the unit tangent, another contradiction. Having now shown that $\mathbf{T}$ intersects all great circles at least twice we have from Crofton's formula that

$$\int \kappa ds = L(\mathbf{T}) = \frac{1}{4} \int_{S^2} n_{\mathbf{T}}(\mathbf{x}) \, d\mathbf{x} \geq \frac{2}{4} \cdot 4\pi = 2\pi.$$

This shows that the total curvature must be $\geq 2\pi$. And that equality forces the curve to be planar. Finally we know that for planar curves the total absolute curvature is $> 2\pi$ unless the curve is convex. □

Definition 3.3.2. A simple closed curve $\mathbf{q}$ is called an *unknot* or said to be *unknotted* if there is a one-to-one map from the disc to $\mathbb{R}^3$ such that boundary of the disc is $\mathbf{q}$.

THEOREM 3.3.3. *(Fary, 1949 and Milnor, 1950) If a simple closed space curve is knotted, then*

$$\int \kappa ds \geq 4\pi.$$

PROOF. We assume that $\int \kappa ds < 4\pi$ and show that the curve is not knotted. Crofton's formula tells us that

$$\frac{1}{4}\int_{S^2} n_{\mathbf{T}}(\mathbf{x})\,d\mathbf{x} = \int \kappa ds < 4\pi.$$

As the sphere has area 4π this can only happen if we can find $\mathbf{x}$ such that $n_{\mathbf{T}}(\mathbf{x}) \leq 3$. Now observe that

$$\frac{d(\mathbf{q}\cdot\mathbf{x})}{ds} = \mathbf{T}(s)\cdot\mathbf{x}.$$

So the function $\mathbf{q}\cdot\mathbf{x}$ has at most three critical points. Since $\mathbf{q}$ is closed there will be a maximum and a minimum. The third critical point, should it exist, can consequently only be an inflection point. Assume that the minimum is obtained at $s=0$ and the maximum at $s_0 \in (0,L)$. The third critical point can be assumed to be in $(0,s_0)$. This implies that the function $\mathbf{q}(s)\cdot x$ is strictly increasing on $(0,s_0)$ and strictly decreasing on (s_0,L). For each $t \in (0,s_0)$ we can then find a unique $s(t) \in (s_0,L)$ such that $\mathbf{q}(t)\cdot\mathbf{x} = \mathbf{q}(s(t))\cdot\mathbf{x}$. Join the two points $\mathbf{q}(t)$ and $\mathbf{q}(s(t))$ by a segment. These segments will sweep out an area whose boundary is the curve and no two of the segments intersect as they belong to parallel planes orthogonal to $\mathbf{x}$. This shows that the curve is the unknot. $\square$

Exercises.

(1) Let $\mathbf{q}$ be a unit speed spherical curve
  (a) Show that

$$\begin{aligned}
\kappa^2 &= 1+\kappa_g^2,\\
\mathbf{N} &= \frac{1}{\kappa}\left(-\mathbf{q}+\kappa_g\mathbf{S}\right),\\
\mathbf{B} &= \frac{1}{\kappa}\left(\kappa_g\mathbf{q}+\mathbf{S}\right),\\
\tau &= \frac{1}{1+\kappa_g^2}\frac{d\kappa_g}{ds}.
\end{aligned}$$

  (b) Show that $\mathbf{q}$ is planar if and only if the curvature is constant.

(2) Show that for a regular spherical curve $\mathbf{q}(t)$

$$\kappa_g = \frac{\det\begin{bmatrix}\mathbf{q} & \frac{d\mathbf{q}}{dt} & \frac{d^2\mathbf{q}}{dt^2}\end{bmatrix}}{\left(\frac{ds}{dt}\right)^3}.$$

(3) (Jacobi) Let $\mathbf{q}(s):[0,L]\to\mathbb{R}^3$ be a closed unit speed curve with positive curvature and consider the unit normal $\mathbf{N}$ as a closed curve on S^2.
  (a) Show that if $s_{\mathbf{N}}$ denotes the arclength parameter of $\mathbf{N}$, then

$$\left(\frac{ds_{\mathbf{N}}}{ds}\right)^2 = \kappa_{\mathbf{q}}^2+\tau_{\mathbf{q}}^2,$$

  where $\kappa_{\mathbf{q}}$ and $\tau_{\mathbf{q}}$ are the curvature and torsion of $\mathbf{q}$.

(b) Show that the geodesic curvature κ_g of $\mathbf{N}$ is given by

$$\kappa_g = \frac{\kappa_{\mathbf{q}} \frac{d\tau_{\mathbf{q}}}{ds} - \tau_{\mathbf{q}} \frac{d\kappa_{\mathbf{q}}}{ds}}{\left(\kappa_{\mathbf{q}}^2 + \tau_{\mathbf{q}}^2\right)^{\frac{3}{2}}}.$$

(c) Show that

$$\int_0^L \kappa_g(s) \frac{ds_{\mathbf{N}}}{ds} ds = 0.$$

(4) Let $\mathbf{q}(t)$ be a regular closed space curve with positive curvature. Show that if its curvature is $\leq R^{-1}$, then its length is $\geq 2\pi R$.

(5) (Curvature characterization of great circles) Show that $\mathbf{q}(t) = a\cos(t) + b\sin(t)$ where a, b are orthonormal is a unit speed spherical curve whose geodesic curvature vanishes. Conversely show that any spherical unit speed curve whose geodesic curvature vanishes is of that form.

(6) Show that the curve

$$\mathbf{q}(t) = (\cos(t)\cos(at), \sin(t)\cos(at), \sin(at))$$

lies on the unit sphere. Compute its curvature.

(7) Show that a simple closed planar curve is unknotted. Show similarly that a simple closed spherical curve is unknotted. Hint: This relies on an improved version of the Jordan curve theorem.

(8) The trefoil curve is given by

$$\mathbf{q}(t) = ((a + R\cos(3t))\cos(2t), (a + R\cos(3t))\sin(2t), R\sin(3t)),$$

where $a > b > 0$ and $t \in [0, 2\pi]$. Sketch this curve (it lies on a torus which is created by rotating the circle in the x, z-plane of radius R centered at $(a, 0, 0)$ around the z-axis) and try to prove that it is knotted.

(9) Let $\mathbf{q}(s) : [0, L] \to S^2$ be a closed curve. For each $\mathbf{x} \in S^2$ let $\phi(\mathbf{x}) \in [0, \pi]$ be the largest spherical distance from $\mathbf{x}$ to the curve $\mathbf{q}(s)$, $s \in [0, L]$, i.e., $\phi(\mathbf{x}) = \max_{s \in [0,L]} \arccos(\mathbf{x} \cdot \mathbf{q}(s))$.

(a) Show that if $\phi(\mathbf{x}) \geq \frac{\pi}{2}$ for all $\mathbf{x}$ then

$$L \geq 2\pi.$$

Hint: Proceed as in the proof of Fenchel's theorem.

(b) Show that if $L < 2\pi$, then $\mathbf{q}$ lies in an open hemisphere. Hint: Use Crofton's formula.

(c) Show that if $\phi = \min_{\mathbf{x}} \phi(\mathbf{x}) < \frac{\pi}{2}$, then $L \geq 4\phi$. Hint: If $L < 4\phi$, then divide the curve into two arcs of equal length and let $\mathbf{x}$ be the midpoint on the shorter part of the great circle that goes through the points that divide the curve. On the other hand $\phi(\mathbf{x}) = \arccos(\mathbf{x} \cdot \mathbf{q}(s_0)) \geq \phi$. Use this to show that the arc that contains $\mathbf{q}(s_0)$ must have length $\geq 2\phi$. This contradicts that the two arcs divide the curve into pieces of equal length.

(d) (Segre, 1947) Apply the results to the binormal $\mathbf{B}$ of a closed unit speed curve to obtain results for the total absolute torsion $\int |\tau| \, ds$ by first showing that $L(\mathbf{B}) = \int |\tau| \, ds$.

(10) Show that if $\mathbf{C}(\sigma)$ is a unit speed curve on the unit sphere, then for all r, θ the curve

$$\mathbf{q} = r \int \mathbf{C} d\sigma + r \cot\theta \int \mathbf{C} \times \frac{d\mathbf{C}}{d\sigma} d\sigma$$

has a Bertrand mate. Hint: Start by establishing the formulas

$$\begin{aligned} \mathbf{T} &= \sin\theta \mathbf{C} + \cos\theta \mathbf{C} \times \frac{d\mathbf{C}}{d\sigma}, \\ \mathbf{N} &= \pm \frac{d\mathbf{C}}{d\sigma}, \\ \mathbf{B} &= \pm \left(-\cos\theta \mathbf{C} + \sin\theta \mathbf{C} \times \frac{d\mathbf{C}}{d\sigma} \right). \end{aligned}$$

Conversely show that any curve that has a Bertrand mate can be written in this way.

(11) Let $\mathbf{q}(s) : [0, L] \to S^2$ be a closed unit speed spherical curve. Show that

$$\int_0^L \frac{\tau}{\kappa} ds = 0.$$

Hint: Use that for a spherical curve $\frac{\tau}{\kappa} = \frac{df}{ds}$ for a suitable function f.

(12) Let $\mathbf{q}(s) : [0, L] \to S^2$ be a unit speed spherical curve and write

$$\mathbf{q} = \alpha \mathbf{T} + \beta \mathbf{N} + \gamma \mathbf{B}.$$

(a) Show that

$$\alpha = 0, \ \beta = -\frac{1}{\kappa}, \ \frac{d\beta}{ds} = \tau\gamma.$$

(b) When $\kappa > 1$ show that

$$\tau = \frac{df}{ds}$$

for a suitable function $f(s)$ that only depends on κ and κ'.

(c) When $\kappa > 1$ and $\mathbf{q}$ is closed show that

$$\int_0^L \tau ds = 0.$$

(d) When $\kappa(s) = 1$ for $s = 0, L$ and $\kappa > 1$ for $s \in (0, L)$ show that

$$\int_0^L \tau ds = 0.$$

Note this does not rely on $\mathbf{q}$ being closed.

(e) Show that if $\kappa = 1$ at only finitely many points and $\mathbf{q}$ is closed then

$$\int_0^L \tau ds = 0.$$

This result holds for all closed spherical curves. Segre has also shown that a closed space curve with $\int_0^L \tau ds = 0$ must be spherical.

CHAPTER 4

Basic Surface Theory

4.1. Surfaces

DEFINITION 4.1.1. A *parametrized surface* is defined as a map $\mathbf{q}(u,v) : U \subset \mathbb{R}^2 \to \mathbb{R}^3$ where $\frac{\partial \mathbf{q}}{\partial u}$ and $\frac{\partial \mathbf{q}}{\partial v}$ are linearly independent.

For parametrized surfaces we generally do not worry about self-intersections or other topological pathologies. For example one can parametrize all but the North and South pole of a sphere $S^2(R) = \left\{ q \in \mathbb{R}^3 \mid |q| = R > 0 \right\}$ using latitudes and meridians:

$$\mathbf{q}(\mu,\phi) = R \begin{bmatrix} \cos\mu\cos\phi \\ \sin\mu\cos\phi \\ \sin\phi \end{bmatrix},$$

where $\phi \in \left(-\frac{\pi}{2}, \frac{\pi}{2}\right)$ denotes the latitude and μ the meridian/longitude. This is a valid parametrization of a surface as long as $\cos\phi \neq 0$. This parametrization is called the *equirectangular* parametrization and is the most common way of coordinatizing Earth and the sky. Curiously, it predates Cartesian coordinates by about 1500 years and is very likely the oldest parametrization of a surface that is still in use.

DEFINITION 4.1.2. A *reparametrization of a parametrized surface* $\mathbf{q}(u,v) : U \subset \mathbb{R}^2 \to \mathbb{R}^3$ is a parametrized surface $\mathbf{q}(s,t) : O \subset \mathbb{R}^2 \to \mathbb{R}^3$ such that the parameters are smooth functions of each other on their respective domains: $(u,v) = (u(s,t), v(s,t))$ for all $(s,t) \in O$, $(s,t) = (s(u,v), t(u,v))$ for $u, v \in U$, and finally that with these changes we still obtain the same surface $\mathbf{q}(u,v) = \mathbf{q}(s,t)$.

DEFINITION 4.1.3. A map $F : O \to U$ between open sets $O, U \subset \mathbb{R}^2$ is called a *diffeomorphism* if it is one-to-one, onto and both F and the inverse map $F^{-1} : U \to O$ are smooth. Thus a reparametrization comes from a diffeomorphism between the domains.

When we wish to avoid self-intersections, then we resort to the more restrictive class of surfaces that come from the next two general constructions. For curves this corresponds to the notion of being simple and in that case we could have used the approach we shall take for surfaces.

The first construction is to use a particularly nice way of parametrizing surfaces without self-intersections or other nasty topological problems. These are the three different types of parametrizations where the surface is represented as a smooth graph:

$$\begin{aligned} \mathbf{q}(u,v) &= (u, v, f(u,v)), \\ \mathbf{q}(u,v) &= (u, f(u,v), v), \\ \mathbf{q}(u,v) &= (f(u,v), u, v). \end{aligned}$$

They are also known as *Monge patches*.

EXAMPLE 4.1.4. The western hemisphere on $S^2(1)$ can be parametrized using the y, z coordinates

$$\mathbf{q}(u,v) = \begin{bmatrix} -\sqrt{u^2+v^2} \\ u \\ v \end{bmatrix},$$

where $(u,v) \in U = \left\{u^2+v^2<1\right\}$. Using latitudes/meridians the parametrization is instead

$$\mathbf{q}(\mu,\phi) = \begin{bmatrix} \cos\mu\cos\phi \\ \sin\mu\cos\phi \\ \sin\phi \end{bmatrix},$$

with $(\mu,\phi) \in \left(\frac{\pi}{2},\frac{3\pi}{2}\right) \times \left(-\frac{\pi}{2},\frac{\pi}{2}\right)$. Setting these two expressions equal to each other tells us that

$$\begin{bmatrix} u \\ v \end{bmatrix} = G(\mu,\phi) = \begin{bmatrix} \sin\mu\cos\phi \\ \sin\phi \end{bmatrix}.$$

This map is smooth and it is not hard to check that as a map from $\left(\frac{\pi}{2},\frac{3\pi}{2}\right)\times\left(-\frac{\pi}{2},\frac{\pi}{2}\right)$ to U it is one-to-one and onto. The differential is

$$DG = \frac{\partial(u,v)}{\partial(\mu,\phi)} = \begin{bmatrix} \frac{\partial u}{\partial\mu} & \frac{\partial u}{\partial\phi} \\ \frac{\partial v}{\partial\mu} & \frac{\partial v}{\partial\phi} \end{bmatrix} = \begin{bmatrix} \cos\mu\cos\phi & -\sin\mu\sin\phi \\ 0 & \cos\phi \end{bmatrix}.$$

The determinant is $\cos\mu\cos^2\phi$ which is always negative on our domain. The inverse function theorem then guarantees us that G is indeed a diffeomorphism. In this case it is also possible to construct the inverse using inverse trigonometric functions.

THEOREM 4.1.5. *Let $\mathbf{q}(u,v) : U \subset \mathbb{R}^2 \to \mathbb{R}^3$ be a parametrized surface. For every $(u_0,v_0) \in U$ there exists a neighborhood $(u_0,v_0) \in V \in U$ such that the smaller parametrized surface $\mathbf{q}(u,v) : V \to \mathbb{R}^3$ can be represented as a Monge patch.*

PROOF. By assumption the matrix

$$\begin{bmatrix} \frac{\partial\mathbf{q}}{\partial u} & \frac{\partial\mathbf{q}}{\partial v} \end{bmatrix} = \begin{bmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \\ \frac{\partial z}{\partial u} & \frac{\partial z}{\partial v} \end{bmatrix}$$

always has rank 2. Assume for the sake of argument that at (u_0,v_0) the middle row is a linear combination of the other two rows. Then the matrix

$$\begin{bmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial z}{\partial u} & \frac{\partial z}{\partial v} \end{bmatrix}$$

is nonsingular at (u_0,v_0). Thus the map $(x,z) = (x(u,v), z(u,v)) : U \to \mathbb{R}^2$ has nonsingular differential at (u_0,v_0). The Inverse Function Theorem then tells us that there must exist neighborhoods $(u_0,v_0) \in V \subset U$ and $(x(u_0,v_0), x(u_0,v_0)) \in O \subset \mathbb{R}^2$ such that function $(x,z) = (x(u,v), z(u,v)) : V \to O$ can be smoothly inverted, i.e., there is a smooth inverse $(u,v) = (u(x,z), v(x,z)) : O \to V$ that allows us to smoothly solve for (u,v) in terms of (x,z). This gives us the desired reparametrization to a Monge patch

$$\begin{bmatrix} x(u,v) \\ y(u,v) \\ z(u,v) \end{bmatrix} = \mathbf{q}(u,v) = \mathbf{q}(x,z) = \begin{bmatrix} x \\ y(u(x,z), v(x,z)) \\ z \end{bmatrix}.$$

$\square$

DEFINITION 4.1.6. A *surface* is defined as a subset $M \subset \mathbb{R}^3$ where for all $q \in M$, there is exists an open set $O \subset \mathbb{R}^3$ such that $O \cap M$ can be represented as a Monge patch, i.e., it is locally a smooth graph over one of the three coordinate planes.

A parametrization $\mathbf{q}(u,v) : U \subset \mathbb{R}^2 \to \mathbb{R}^3$ is called a *coordinate system* if the map is one-to-one and the image $\mathbf{q}(U)$ is a surface.

EXAMPLE 4.1.7. Despite the above theorem not all parametrized surfaces are surfaces in this restrictive sense. Let $\mathbf{q}(u) = (x(u), y(u))$ be a regular planar curve, then we obtain a parametrized surface $\mathbf{q}(u,v) = (x(u), y(u), v)$. However, this might not be a surface if the planar curve looks like a figure 8. We could also take something like a figure 6 but parametrize it so that the loop gets arbitrarily close without intersecting. In the latter case we simply parametrize the figure 6 using an open interval $(0,1)$.

The second construction comes from considering level sets. A level set is a set of the form

$$\{(x,y,z) \in O \mid F(x,y,x) = c\},$$

where c is some fixed constant and $O \subset \mathbb{R}^3$ is some open set.

EXAMPLE 4.1.8. For example

$$x^2 + y^2 + z^2 = R^2$$

describes the sphere as a level set. Depending on where we are on the sphere different parametrizations are possible. At points where, say, $y < 0$ we can use

$$\mathbf{q}(u,v) = \left(u, -\sqrt{R^2 - u^2 - v^2}, v\right)$$

This will in fact parametrize all points where $y < 0$ if we use all (u,v) with $u^2+v^2 < R^2$.

We also have a general theorem that can be used

THEOREM 4.1.9. *Let $F : O \to \mathbb{R}$ be a smooth function and $c \in \mathbb{R}$ a constant. The level set*

$$M = \{(x,y,z) \in O \mid F(x,y,x) = c\}$$

is a smooth surface if it is not empty and for all $q \in M$ the gradient

$$\nabla F(q) = \begin{bmatrix} \frac{\partial F}{\partial x}(q) \\ \frac{\partial F}{\partial y}(q) \\ \frac{\partial F}{\partial z}(q) \end{bmatrix} \neq 0.$$

PROOF. Fix $q = (x_0, y_0, z_0) \in M$ and assume for the sake of argument that $\frac{\partial F}{\partial y}(q) \neq 0$. The implicit function theorem tells us that there are neighborhoods $q \in O_1 \in O$ and $(x_0, z_0) \in U \in \mathbb{R}^2$ as well as a smooth function $f(u,v) : U \to \mathbb{R}$ such that for all $(u,v) \in U$ we have $(u, f(u,v), v) \in O_1$ and

$$M \cap O_1 = \{(x,y,z) \in O_1 \mid F(x,y,x) = c\} = \{(u, f(u,v), v) \mid (u,v) \in U\}.$$

Thus $M \cap O_1$ can be written as a graph over the (x,z)-plane. $\square$

Exercises.

(1) A *generalized cylinder* is determined by a regular curve $\alpha(t)$ and a vector X that is never tangent to the curve. It consists of the lines that are parallel to the vector and pass through the curve.
   (a) Show that
   $$\mathbf{q}(s,t) = \alpha(t) + sX$$
   is a natural parametrization and show that it gives a parametrized surface.
   (b) Show that we can reconstruct the cylinder so that the curve lies in the plane perpendicular to the vector X. Hint: Try the case where $X = (0,0,1)$ and the plane is the (x,y)-plane and make sure your new parametrization is a valid parametrization precisely when the old parametrization was valid.
   (c) What happens when the curve is given by two equations and you also want the surface to be given by an equation? Hint: If a planar curve in the (x,y)-plane given by $F(x,y) = c$, then the generalized cylinder is also given by $F(x,y) = c$.

(2) A *generalized cone* is generated by a regular curve $\alpha(t)$ and a point p not on the curve. It consists of the lines that pass through the point and the curve.
   (a) Show that
   $$\mathbf{q}(s,t) = s(\alpha(t) - p) + p$$
   is a natural parametrization and determine when/where it yields a parametrized surface.
   (b) Show that we can replace $\alpha(t)$ by a curve $\delta(t)$ that lies on a unit sphere centered at the vertex p of the cone.

(3) Let $F(x,y,z)$ be homogeneous of degree n, i.e., $F(\lambda x, \lambda y, \lambda z) = \lambda^n F(x,y,z)$. Show that the set $\{(x,y,z) \neq 0 \mid F(x,y,z) = 0\}$ defines a cone.

(4) A *ruled surface* is given by a parametrization of the form
   $$\mathbf{q}(s,t) = \alpha(t) + sX(t).$$
   It is evidently a surface that is a union of lines (rulers) and generalizes the constructions in the previous exercises. Give conditions on α, X and the parameter s that guarantee we get a parametrized surface. A special case occurs when X is tangent to α. These are also called *tangent developables*.

(5) A *surface of revolution* is determined by a planar regular curve and a line in the same plane. The surface is generated by rotating the curve around the line.
   (a) Show that for a curve $(r(t), z(t))$ in the (x,z)- plane that is rotated around the z-axis the parametrization is
   $$\mathbf{q}(t,\mu) = (r(t)\cos\mu, r(t)\sin\mu, z(t))$$
   and show that it is a parametrized surface.
   (b) Show that the equation for the surface is $F\left(\sqrt{x^2+y^2}, z\right) = c$ when the curve is given by $F(x,z) = c$ with $x > 0$.

(6) Many classical surfaces are of the form
   $$F(x,y,z) = ax^2 + by^2 + cz^2 + dx + ey + fz + g = 0.$$

These are called *quadratic surfaces* if one of a, b, or $c \neq 0$.

(a) Give conditions on the coefficients such that it generates a surface ($g = 0$ takes special care).
(b) Under what conditions does it become a surface of revolution around the z-axis?
(c) Show that when the equation does not depend on one of the coordinates, then we obtain a generalized cylinder.
(d) When, say $c = 0$, but $abf \neq 0$ we obtain a *paraboloid*. It is *elliptic* when a, b have the same sign and otherwise *hyperbolic*. Draw pictures of these two situations.
(e) When $abc \neq 0$ show that it can be rewritten in the form

$$F(x, y, z) = a(x - x_0)^2 + b(y - y_0)^2 + c(z - z_0)^2 + h = 0.$$

(f) When all three a, b, c have the same sign show that it is either empty or an *ellipsoid.*
(g) When not all of a, b, c have the same sign and $h \neq 0$ we obtain a *hyperboloid*. Show that it might be connected or have two components (called sheets) depending of the signs of all four constants.
(h) When not all of a, b, c have the same sign and $h = 0$ we obtain a cone.
(i) Given constants $a_x,\ a_y,\ a_z$ determine when

$$\begin{aligned}
\mathbf{q}(u, v) &= (a_x \sin u \sin v, a_y \sin u \cos v, a_z \cos u), \\
\mathbf{q}(u, v) &= (a_x \sinh u \sin v, a_y \sinh u \cos v, a_z \cosh u), \\
\mathbf{q}(u, v) &= (a_x \sinh u \sinh v, a_y \sinh u \cosh v, a_z \sinh u), \\
\mathbf{q}(u, v) &= \left(a_x u \cos v, a_y u \sin v, a_z u^2\right), \\
\mathbf{q}(u, v) &= \left(a_x u \cosh v, a_y u \sinh v, a_z u^2\right),
\end{aligned}$$

yield parameterizations and identify them with the appropriate quadratics.

(7) Let $\mathbf{q}(z, \mu) = \left(\sqrt{1 - z^2} \cos \mu, \sqrt{1 - z^2} \sin \mu, z\right)$ with $-1 < z < 1$ and $-\pi < \mu < \pi$. Show that $\mathbf{q}$ defines a surface. What is the surface?
(8) Consider a regular curve $\alpha(t)$ with non-vanishing curvature and construct the *tube* of radius R around it

$$\mathbf{q}(t, \phi) = \alpha(t) + R(\mathbf{N}_\alpha \cos \phi + \mathbf{B}_\alpha \sin \phi),$$

where $\mathbf{N}_\alpha,\ \mathbf{B}_\alpha$ are the normal and binormal to the curve.

(a) Show that this defines a parametrized surface as long as $\kappa_\alpha < R^{-1}$.
(b) Show by example that this surface might intersect itself if there is a cord of length $< 2R$ that is normal to the curve at both end points.
(c) Show that when α is a circle, then we obtain a surface of revolution that looks like a torus.

(9) For which R is the level set

$$z\left(x^2 + y^2\right) - xy = R$$

a surface? When $R = 0$ show that we get a surface as long as $(x, y) \neq (0, 0)$. This is called *Plücker's conoid*. Show that it is in fact a ruled surface where all of the lines are of the form $(0, 0, c) + t(a, b, 0)$.

(10) Show that
$$\left(x^2+y^2+z^2+R^2-r^2\right)^2=4R^2\left(x^2+y^2\right)$$
defines a surface when $R>r>0$. Show that it is rotationally symmetric and a torus.

(11) The *helicoid* is given by the equation
$$\tan\frac{z}{h}=\frac{y}{x}$$
where $h\neq 0$ is a fixed constant.
  (a) Show that this defines a surface for suitable (x,y,z).
  (b) Show that the surface can be parametrized by
$$\mathbf{q}(r,\theta)=(r\cos\theta,r\sin\theta,h\theta)$$
  and determine for which r,θ this defines a parametrized surface. Note that for fixed r we obtain helices.

(12) *Enneper's surface* is defined by the parametrization
$$\mathbf{q}(u,v)=\left(u-\frac{1}{3}u^3+uv^2,v-\frac{1}{3}v^3+vu^2,u^2-v^2\right).$$
  (a) For which u,v does this define a parametrization?
  (b) Show that Enneper's surface satisfies the equation
$$\left(\frac{y^2-x^2}{2z}+\frac{2z^2}{9}+\frac{2}{3}\right)^3=6\left(\frac{y^2-x^2}{4z}-\frac{1}{4}\left(x^2+y^2+\frac{8}{9}z^2\right)+\frac{2}{9}\right)^2.$$

4.2. Tangent Spaces and Maps

Definition 4.2.1. The *tangent space* at $p\in M$ of a (parametrized) surface is defined as
$$T_pM=\operatorname{span}\left\{\frac{\partial\mathbf{q}}{\partial u},\frac{\partial\mathbf{q}}{\partial v}\right\},$$
and *normal space*
$$N_pM=\left(T_pM\right)^{\perp}.$$

Remark 4.2.2. For a parametrized surface with self-intersections this is a bit ambivalent as the tangent in that case depends on the parameter values (u,v) and not just the point $p=\mathbf{q}(u,v)$. This is just as for curves where the tangent line at a point really is the tangent line at a point with respect to a specific parameter value.

Proposition 4.2.3. *Both tangent and normal spaces are subspaces that do not change under reparametrization.*

Proof. This would seem intuitively clear, just as with curves, where the tangent line does not depend on parametrizations. For curves it boils down to the simple fact that velocities for different parametrizations are proportional and hence define the same tangent lines. With surfaces something similar happens, but it is a bit more involved. Suppose we have two different parametrizations of the same surface:
$$\mathbf{q}(s,t)=\mathbf{q}(u,v).$$

This tells us that the parameters are functions of each other

$$\begin{aligned} u &= u(s,t),\ v = v(s,t), \\ s &= s(u,v),\ t = t(u,v). \end{aligned}$$

The chain rule then gives us

$$\frac{\partial \mathbf{q}}{\partial u} = \frac{\partial \mathbf{q}}{\partial s}\frac{\partial s}{\partial u} + \frac{\partial \mathbf{q}}{\partial t}\frac{\partial t}{\partial u} \in \operatorname{span}\left\{\frac{\partial \mathbf{q}}{\partial s}, \frac{\partial \mathbf{q}}{\partial t}\right\}.$$

Similarly

$$\frac{\partial \mathbf{q}}{\partial v} \in \operatorname{span}\left\{\frac{\partial \mathbf{q}}{\partial s}, \frac{\partial \mathbf{q}}{\partial t}\right\}$$

In the other direction we clearly also get

$$\frac{\partial \mathbf{q}}{\partial s}, \frac{\partial \mathbf{q}}{\partial t} \in \operatorname{span}\left\{\frac{\partial \mathbf{q}}{\partial u}, \frac{\partial \mathbf{q}}{\partial v}\right\}.$$

This shows that at a fixed point p on a surface the tangent space does not depend on how the surface is parametrized. The normal space is then also well defined. □

Note that the chain rule shows in matrix notation that

$$\begin{aligned} \begin{bmatrix} \frac{\partial \mathbf{q}}{\partial u} & \frac{\partial \mathbf{q}}{\partial v} \end{bmatrix} &= \begin{bmatrix} \frac{\partial \mathbf{q}}{\partial s} & \frac{\partial \mathbf{q}}{\partial t} \end{bmatrix} \begin{bmatrix} \frac{\partial s}{\partial u} & \frac{\partial s}{\partial v} \\ \frac{\partial t}{\partial u} & \frac{\partial t}{\partial v} \end{bmatrix}, \\ \begin{bmatrix} \frac{\partial \mathbf{q}}{\partial s} & \frac{\partial \mathbf{q}}{\partial t} \end{bmatrix} &= \begin{bmatrix} \frac{\partial \mathbf{q}}{\partial u} & \frac{\partial \mathbf{q}}{\partial v} \end{bmatrix} \begin{bmatrix} \frac{\partial u}{\partial s} & \frac{\partial u}{\partial t} \\ \frac{\partial v}{\partial s} & \frac{\partial v}{\partial t} \end{bmatrix} \end{aligned}$$

with

$$\begin{bmatrix} \frac{\partial s}{\partial u} & \frac{\partial s}{\partial v} \\ \frac{\partial t}{\partial u} & \frac{\partial t}{\partial v} \end{bmatrix}^{-1} = \begin{bmatrix} \frac{\partial u}{\partial s} & \frac{\partial u}{\partial t} \\ \frac{\partial v}{\partial s} & \frac{\partial v}{\partial t} \end{bmatrix}.$$

A better way of defining the tangent space that also shows that it is independent of parameterizations comes from the next result.

Proposition 4.2.4. *The tangent space at $q = \mathbf{q}(u_0, v_0)$ for a (parametrized) surface is given by*

$$T_q M = \left\{\mathbf{v} \in \mathbb{R}^3 \mid \mathbf{v} = \frac{d\mathbf{q}}{dt}(0) \text{ for a smooth curve } \mathbf{q}(t) : I \to M \text{ with } \mathbf{q}(0) = q\right\}.$$

Proof. Any curve $\mathbf{q}(t)$ on the surface that passes through q at $t = 0$ can be written as

$$\mathbf{q}(t) = \mathbf{q}(u(t), v(t))$$

for smooth functions $u(t)$ and $v(t)$ with $u(0) = u_0$ and $v(0) = v_0$ as long as t is sufficiently small. This is because the parametrization is locally one-to-one. If we write the curve this way, then

$$\frac{d\mathbf{q}}{dt} = \frac{\partial \mathbf{q}}{\partial u}\frac{du}{dt} + \frac{\partial \mathbf{q}}{\partial v}\frac{dv}{dt}.$$

Showing that velocities of curves on the surface are always tangent vectors. Conversely by using $u(t) = at + u_0$ and $v(t) = bt + v_0$ we obtain all possible linear combinations of tangent vectors as

$$\frac{d\mathbf{q}}{dt}(0) = \frac{\partial \mathbf{q}}{\partial u}a + \frac{\partial \mathbf{q}}{\partial v}b.$$

□

COROLLARY 4.2.5. *Let $M = \{(x, y, z) \in O \mid F(x, y, x) = c\}$ be a smooth surface as in theorem 4.1.9. The normal is given by*

$$\mathbf{N}(q) = \frac{\nabla F(q)}{|\nabla F(q)|} = \frac{1}{\left(\left(\frac{\partial F}{\partial x}(q)\right)^2 + \left(\frac{\partial F}{\partial y}(q)\right) + \left(\frac{\partial F}{\partial z}(q)\right)\right)^{1/2}} \begin{bmatrix} \frac{\partial F}{\partial x}(q) \\ \frac{\partial F}{\partial y}(q) \\ \frac{\partial F}{\partial z}(q) \end{bmatrix}.$$

PROOF. We saw in proposition 4.2.4 that any tangent vector in $T_q M$ can be represented as a velocity vector $\dot{\mathbf{q}}(0)$. Since $\mathbf{q}(t) \in M$ it follows that $F(\mathbf{q}(t)) = c$ for all t. Then chain rule then implies that

$$0 = \nabla F(\mathbf{q}(0)) \cdot \dot{\mathbf{q}}(0) = \nabla F(q) \cdot \dot{\mathbf{q}}(0).$$

This shows that the gradient is perpendicular to all tangent vectors and hence a normal vector. This shows the claim as $\frac{\nabla F(q)}{|\nabla F(q)|}$ is a unit vector. □

EXAMPLE 4.2.6. The sphere of radius R centered at the origin has a unit normal given by the unit radial vector at $q = (x, y, z) \in S^2(R)$

$$\mathbf{N} = \frac{1}{R} \begin{bmatrix} x \\ y \\ z \end{bmatrix}.$$

The basis for the tangent space with respect to the meridian/latitude parametrization is

$$\frac{\partial \mathbf{q}}{\partial \mu} = R \begin{bmatrix} -\sin\mu\cos\phi \\ \cos\mu\cos\phi \\ 0 \end{bmatrix}, \quad \frac{\partial \mathbf{q}}{\partial \phi} = R \begin{bmatrix} -\cos\mu\sin\phi \\ -\sin\mu\sin\phi \\ \cos\phi \end{bmatrix}.$$

It is often useful to find coordinates suited to a particular situation. However, unlike for curves, it isn't always possible to parametrize a surface such that the coordinate curves are unit speed and orthogonal to each other. But there is one general construction we can do.

THEOREM 4.2.7. *Assume that we have linearly independent tangent vector fields X, Y defined on a surface M. Then it is possible to find a parametrization $\mathbf{q}(u, v)$ in a neighborhood of any point such that $\frac{\partial \mathbf{q}}{\partial u}$ is proportional to X and $\frac{\partial \mathbf{q}}{\partial v}$ is proportional to Y.*

PROOF. The vector fields have integral curves forming a net on the surface. Apparently the goal is to reparametrize the curves in this net in some fashion. The difficulty lies in ensuring that the levels where u is constant correspond to the v-curves, and vice versa. We proceed as with the classical construction of Cartesian coordinates. Select a point p and let the u-axis be the integral curve for X through p, similarly set the v-axis be the integral curve for Y through p. Both of these curves retain the parametrizations that make them integral curves for X and Y. Thus p will naturally correspond to $(u, v) = (0, 0)$. We now wish to assign (u, v) coordinates to a point q near p. There are also unique integral curves for X and Y through q. These will be our way of projecting onto the chosen axes and will in this way yield the desired coordinates. Specifically $u(q)$ is the parameter where the integral curve for Y through q intersects the u-axis, and similarly with $v(q)$. In general integral curves can intersect axes in several places or might not intersect them at all. However, a continuity argument offers some justification when we consider that the axes themselves are the proper integral curves for the qs that lie on these axes and so when q sufficiently close to both axes it should have a well defined set of coordinates. We also note that as the projection happens along integral curves we have ensured that coordinate curves are simply reparametrizations of integral curves. To completely justify this construction we need to know quite a bit about the existence, uniqueness and smoothness of solutions to differential equations and the inverse function theorem. $\square$

REMARK 4.2.8. Note that this proof gives us a little more information. Specifically, we obtain a parametrization where the parameter curves through $(0,0)$ are the integral curves for X and Y.

DEFINITION 4.2.9. A *map* between surfaces $F : M_1 \to M_2$ is simply an assignment of points in the first surface to points in the second. The map is *smooth* if around every point $q \in M_1$ we can find a parametrization $\mathbf{q}_1(u,v)$ where $q = \mathbf{q}_1(u_0,v_0)$ such that the composition $F \circ \mathbf{q}_1 : U \to \mathbb{R}^3$ is a smooth map as a map from the space of parameters to the ambient space that contains the target M_2.

We can also define maps between parametrized surfaces in a similar way. Clearly parametrizations are themselves smooth maps. It is also often the case that the compositions $F \circ \mathbf{q}_1$ are themselves parametrizations.

EXAMPLE 4.2.10. Two classical examples of maps are the *Archimedes* and *Mercator projections* from the sphere to the cylinder of the same radius placed to touch the sphere at the equator. We give the formulas for the unit sphere and note that neither map is defined at the poles.

The Archimedes map is simply a horizontal projection that preserves the z-coordinate

$$A\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} \frac{x}{\sqrt{x^2+y^2}} \\ \frac{y}{\sqrt{x^2+y^2}} \\ z \end{bmatrix}.$$

In the meridian/latitude parametrization it looks particularly nice:

$$A\begin{bmatrix} \cos\mu\cos\phi \\ \sin\mu\cos\phi \\ \sin\phi \end{bmatrix} = \begin{bmatrix} \cos\mu \\ \sin\mu \\ \sin\phi \end{bmatrix}.$$

Note that what is here referred to as the Archimedes map is often called the Lambert projection. However, Archimedes was the first to discover that the areas of the sphere and cylinder are equal. This will be discussed in greater detail in section 4.5.

The Mercator projection (1569) differs in that the z-coordinate is not preserved:

$$M\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} \frac{x}{\sqrt{x^2+y^2}} \\ \frac{y}{\sqrt{x^2+y^2}} \\ \frac{1}{2}\log\frac{1+z}{1-z} \end{bmatrix}$$

or

$$M\begin{bmatrix} \cos\mu\cos\phi \\ \sin\mu\cos\phi \\ \sin\phi \end{bmatrix} = \begin{bmatrix} \cos\mu \\ \sin\mu \\ \frac{1}{2}\log\frac{1+\sin\phi}{1-\sin\phi} \end{bmatrix}.$$

Both of these maps really are maps in the traditional sense that they can be used to picture the Earth on a flat piece of paper by cutting the cylinder vertically and unfolding it. This unfolding is done along a meridian. For Eurocentric people it is along the date line. In the Americas one also often sees maps cut along a meridian that is just East of India.

DEFINITION 4.2.11. The differential of a smooth map $F : M_1 \to M_2$ at $q \in M_1$ is the map

$$DF_q : T_q M_1 \to T_{F(q)} M_2$$

defined by

$$DF_q(\mathbf{v}) = \frac{dF \circ \mathbf{q}}{dt}(0)$$

if $\mathbf{q}(t)$ is a curve (in M_1) with $q = \mathbf{q}(0)$ and $\mathbf{v} = \frac{d\mathbf{q}}{dt}(0)$.

PROPOSITION 4.2.12. *When* $\mathbf{v} = \frac{d\mathbf{q}}{dt}(0) = \frac{\partial \mathbf{q}}{\partial u}\mathbf{v}^u + \frac{\partial \mathbf{q}}{\partial v}\mathbf{v}^v$ *we have*

$$DF_q(\mathbf{v}) = \begin{bmatrix} \frac{\partial F \circ \mathbf{q}}{\partial u} & \frac{\partial F \circ \mathbf{q}}{\partial v} \end{bmatrix} \begin{bmatrix} \mathbf{v}^u \\ \mathbf{v}^v \end{bmatrix}.$$

In particular, the differential is a linear map and is completely determined by the two partial derivatives $\frac{\partial F \circ \mathbf{q}}{\partial u}, \frac{\partial F \circ \mathbf{q}}{\partial v}$.

PROOF. This follows from the chain rule:

$$\begin{aligned} \frac{dF \circ \mathbf{q}}{dt}(t) &= \frac{dF(\mathbf{q}(t))}{dt} \\ &= \frac{dF(\mathbf{q}(u(t), v(t)))}{dt} \\ &= \frac{\partial F \circ \mathbf{q}}{\partial u}\frac{du}{dt} + \frac{\partial F \circ \mathbf{q}}{\partial v}\frac{dv}{dt} \\ &= \begin{bmatrix} \frac{\partial F \circ \mathbf{q}}{\partial u} & \frac{\partial F \circ \mathbf{q}}{\partial v} \end{bmatrix} \begin{bmatrix} \frac{du}{dt} \\ \frac{dv}{dt} \end{bmatrix}. \end{aligned}$$

□

EXAMPLE 4.2.13. The Archimedes map satisfies

$$\frac{\partial (A \circ \mathbf{q})}{\partial \mu} = \begin{bmatrix} -\sin\mu \\ \cos\mu \\ 0 \end{bmatrix}, \quad \frac{\partial (A \circ \mathbf{q})}{\partial \phi} = \begin{bmatrix} 0 \\ 0 \\ -\cos\phi \end{bmatrix}$$

and the Mercator map

$$\frac{\partial (M \circ \mathbf{q})}{\partial \mu} = \begin{bmatrix} -\sin\mu \\ \cos\mu \\ 0 \end{bmatrix}, \quad \frac{\partial (M \circ \mathbf{q})}{\partial \phi} = \begin{bmatrix} 0 \\ 0 \\ \frac{1}{\cos\phi} \end{bmatrix}.$$

DEFINITION 4.2.14. We say that a surface M is *orientable* if we can select a smooth normal field. Thus we require a smooth function

$$\mathbf{N} : M \to S^2(1) \subset \mathbb{R}^3$$

such that for all $q \in M$ the vector $\mathbf{N}(q)$ is perpendicular to the tangent space $T_q M$. The map $\mathbf{N} : M \to S^2(1)$ is called the *Gauss map*.

PROPOSITION 4.2.15. *A surface which is given as a level set is orientable.*

PROOF. Form corollary 4.2.5 we know that the normal can be given by

$$\mathbf{N} = \frac{\nabla F}{|\nabla F|}$$

if $M = \{q \in O \mid F(q) = c\}$. □

EXAMPLE 4.2.16. A parametrized surface $\mathbf{q}(u, v) : U \to \mathbb{R}^3$ always has a natural map $\mathbf{N}(u, v) : U \to \mathbb{R}^3$ defined by

$$\mathbf{N}(u, v) = \frac{\frac{\partial \mathbf{q}}{\partial u} \times \frac{\partial \mathbf{q}}{\partial v}}{\left|\frac{\partial \mathbf{q}}{\partial u} \times \frac{\partial \mathbf{q}}{\partial v}\right|}$$

that gives a unit normal vector at each point. However, it is possible (as well shall see in the exercises) that there are parameter values that give the same points on the surface without giving the same normal vectors.

DEFINITION 4.2.17. The parameters u, v on a parameterized surface $\mathbf{q}(u, v)$ define two differentials du and dv. These are not mysterious infinitesimals, but linear functions on tangent vectors to the surface that compute the coefficients of the vector with respect to the basis $\frac{\partial \mathbf{q}}{\partial u}, \frac{\partial \mathbf{q}}{\partial v}$. Thus

$$\begin{aligned} du(\mathbf{v}) &= du\left(\frac{\partial \mathbf{q}}{\partial u}\mathbf{v}^u + \frac{\partial \mathbf{q}}{\partial v}\mathbf{v}^v\right) = \mathbf{v}^u, \\ dv(\mathbf{v}) &= dv\left(\frac{\partial \mathbf{q}}{\partial u}\mathbf{v}^u + \frac{\partial \mathbf{q}}{\partial v}\mathbf{v}^v\right) = \mathbf{v}^v, \end{aligned}$$

and

$$\mathbf{v} = \begin{bmatrix} \frac{\partial \mathbf{q}}{\partial u} & \frac{\partial \mathbf{q}}{\partial v} \end{bmatrix} \begin{bmatrix} du \\ dv \end{bmatrix} (\mathbf{v}) = \begin{bmatrix} \frac{\partial \mathbf{q}}{\partial u} & \frac{\partial \mathbf{q}}{\partial v} \end{bmatrix} \begin{bmatrix} \mathbf{v}^u \\ \mathbf{v}^v \end{bmatrix}.$$

From the chain rule we obtain the very natural transformation laws for differentials

$$\begin{aligned} du &= \frac{\partial u}{\partial s}ds + \frac{\partial u}{\partial t}dt, \\ dv &= \frac{\partial v}{\partial s}ds + \frac{\partial v}{\partial t}dt, \end{aligned}$$

or

$$\begin{bmatrix} du \\ dv \end{bmatrix} = \begin{bmatrix} \frac{\partial u}{\partial s} & \frac{\partial u}{\partial t} \\ \frac{\partial v}{\partial s} & \frac{\partial v}{\partial t} \end{bmatrix} \begin{bmatrix} ds \\ dt \end{bmatrix}.$$

Exercises.

(1) Show that the ruled surface

$$\mathbf{q}(t, \phi) = (\cos\phi, \sin\phi, 0) + t\left(\sin\frac{\phi}{2}\cos\phi, \sin\frac{\phi}{2}\sin\phi, \cos\frac{\phi}{2}\right)$$

defines a parametrized surface. It is called the *Möbius band*. Show that it is not orientable by showing that when $t = 0$ and $\phi = \pm\pi$ we obtain the same point and tangent space on the surface, but the normals

$$\mathbf{N}(t, \phi) = \frac{\frac{\partial \mathbf{q}}{\partial t} \times \frac{\partial \mathbf{q}}{\partial \phi}}{\left|\frac{\partial \mathbf{q}}{\partial t} \times \frac{\partial \mathbf{q}}{\partial \phi}\right|}$$

are not the same.

(2) Show that $\mathbf{q}(t,\phi) = t(\cos\phi, \sin\phi, 1)$ defines a parametrization for $(t,\phi) \in (0,\infty)\times\mathbb{R}$. Show that the corresponding surface is $x^2+y^2-z^2=0$, $z>0$. Show that this parametrization is not one-to-one. Find a different parametrization of the entire surface that is one-to-one.

(3) The *inversion in the unit sphere or circle* is defined as

$$F(q) = \frac{q}{|q|^2}.$$

(a) Show that this is a diffeomorphism of $\mathbb{R}^n - 0$ to it self with the property that $q \cdot F(q) = 1$.

(b) Show that F preserves the unit sphere, but reverses the unit normal directions.

(c) Let M be a surface. Show that $M^* = F(M)$ defines another surface. Show that $DF : T_qM \to T_{q^*}M^*$ satisfies

$$DF(v) = \frac{v - 2(q\cdot v)q}{|q|^2}.$$

(d) Show that if a normal $\mathbf{N}$ is a unit normal to M then a unit normal to M^* is given by

$$\mathbf{N}^* = -\mathbf{N} + q\frac{2q\cdot\mathbf{N}}{|q|^2}.$$

(4) A *perspective projection* is defined as a radial projection along lines emanating from a fixed point $\mathbf{c}\in\mathbb{R}^n$ to a hyper-plane $H\subset\mathbb{R}^n$.

(a) Let $\mathbf{c} = (0,0,c) \in \mathbb{R}^3$ and H be the (x,y)-plane. Show that the projection is given by $(x,y,z) \mapsto \left(\frac{x}{z-c}, \frac{y}{z-c}, 0\right)$.

(b) Let $\mathbf{c} = (0,0,0) \in \mathbb{R}^3$ and H be the $\{z=1\}$-plane. Show that the projection is given by $(x,y,z) \mapsto \left(\frac{x}{z}, \frac{y}{z}, 1\right)$.

(c) Let $\mathbf{c} = (0,0,1) \in \mathbb{R}^3$ and H be the $\{z=-1\}$-plane. Show that the projection is given by $(x,y,z) \mapsto \left(\frac{x}{2-2z}, \frac{y}{2-2z}, -1\right)$.

(5) Consider the two maps $\mathbf{q}^{\pm} : \mathbb{R}^n \to \mathbb{R}^n\times\mathbb{R} = \mathbb{R}^{n+1}$

$$\mathbf{q}^{\pm}(q) = (q,0) + \frac{1-|q|^2}{1+|q|^2}(q,\mp 1).$$

These two maps are inverses of perspective projections to the unit sphere. There are also called *stereographic projections.*

(a) Show that these maps are one-to-one, map in to the unit sphere, and that together they cover the unit sphere.

(b) Show that they are the inverse maps of the projections from $(0,\mp 1)\in\mathbb{R}^n\times\mathbb{R}$ to the $\mathbb{R}^n\subset\mathbb{R}^n\times\mathbb{R}$ plane where the last coordinate vanishes when these projections are restricted to the unit-sphere.

(c) Show that $\mathbf{q}^+\left(\frac{q}{|q|^2}\right) = \mathbf{q}^-(q)$ and $\mathbf{q}^+(q) = \mathbf{q}^-\left(\frac{q}{|q|^2}\right)$.

(6) Consider the two surfaces M_1 and M_2 defined by the parametrizations:

$$\begin{aligned}
\mathbf{q}_1(t,\phi) &= (\sinh\phi\cos t, \sinh\phi\sin t, t) \\
&= (0,0,u) + \sinh\phi(\cos u, \sin u, 0), \\
\mathbf{q}_2(t,\phi) &= (\cosh t\cos\phi, \cosh t\sin\phi, t).
\end{aligned}$$

(a) Show that $\mathbf{q}_1 : \mathbb{R} \times \mathbb{R} \to M_1$ is a one-to-one parametrization of a helicoid (see section 4.1 exercise 11.

(b) Show that $\mathbf{q}_2$ is a parametrization that is not one-to-one. Show that M_2 is rotationally symmetric (see section 4.1 exercise 5) and can also be described by the equation

$$x^2 + y^2 = \cosh^2 z.$$

Show further that this equation defines a surface. It is called the *catenoid*.

(c) Define a map $F : M_1 \to M_2$ by $F \circ \mathbf{q}_1 (t, \phi) = \mathbf{q}_2 (t, \theta)$. Show that this map is smooth, not one-to-one, but is locally a diffeomorphism.

4.3. The Abstract Framework

This section can be skipped.

As with curves, parametrized surfaces can have intersections and other nasty complications. Nevertheless, it is often easier to develop formulas for parametrized surfaces.

For a parametrized surface $\mathbf{q}(u, v)$ we have the velocities of the coordinate vector fields

$$\frac{\partial \mathbf{q}}{\partial u}, \frac{\partial \mathbf{q}}{\partial v}.$$

While these can be normalized to be unit vectors we can't guarantee that they are orthogonal. Nor can we necessarily find parameters that make the coordinate fields orthonormal. We shall see that there are geometric obstructions to finding such parametrizations.

Before discussing general surfaces it might be instructive to see what happens if $\mathbf{q}(u, v) : U \to \mathbb{R}^2$ is simply a reparametrization of the plane. Thus $\frac{\partial \mathbf{q}}{\partial u}, \frac{\partial \mathbf{q}}{\partial v}$ form a basis at each point $\mathbf{q}$. Taking partial derivatives of these fields give us

$$\begin{aligned}
\frac{\partial}{\partial u} \begin{bmatrix} \frac{\partial \mathbf{q}}{\partial u} & \frac{\partial \mathbf{q}}{\partial v} \end{bmatrix} &= \begin{bmatrix} \frac{\partial^2 \mathbf{q}}{\partial u^2} & \frac{\partial^2 \mathbf{q}}{\partial u \partial v} \end{bmatrix} = \begin{bmatrix} \frac{\partial \mathbf{q}}{\partial u} & \frac{\partial \mathbf{q}}{\partial v} \end{bmatrix} [\Gamma_u], \\
\frac{\partial}{\partial v} \begin{bmatrix} \frac{\partial \mathbf{q}}{\partial u} & \frac{\partial \mathbf{q}}{\partial v} \end{bmatrix} &= \begin{bmatrix} \frac{\partial \mathbf{q}}{\partial v \partial u} & \frac{\partial^2 \mathbf{q}}{\partial v^2} \end{bmatrix} = \begin{bmatrix} \frac{\partial \mathbf{q}}{\partial u} & \frac{\partial \mathbf{q}}{\partial v} \end{bmatrix} [\Gamma_v]
\end{aligned}$$

or in condensed form

$$\frac{\partial}{\partial w} \begin{bmatrix} \frac{\partial \mathbf{q}}{\partial u} & \frac{\partial \mathbf{q}}{\partial v} \end{bmatrix} = \begin{bmatrix} \frac{\partial^2 \mathbf{q}}{\partial w \partial u} & \frac{\partial^2 \mathbf{q}}{\partial w \partial v} \end{bmatrix} = \begin{bmatrix} \frac{\partial \mathbf{q}}{\partial u} & \frac{\partial \mathbf{q}}{\partial v} \end{bmatrix} [\Gamma_w], \ w = u, v.$$

The matrices $[\Gamma_w]$ tell us how the tangent fields change with respect to themselves. A good example comes from considering polar coordinates $\mathbf{q}(r, \theta) = (r \cos \theta, r \sin \theta)$:

$$\begin{aligned}
\frac{\partial \mathbf{q}}{\partial r} &= \begin{bmatrix} \cos \theta \\ \sin \theta \end{bmatrix}, \frac{\partial \mathbf{q}}{\partial \theta} = \begin{bmatrix} -r \sin \theta \\ r \cos \theta \end{bmatrix}, \\
\frac{\partial^2 \mathbf{q}}{\partial r \partial \theta} &= \frac{\partial^2 \mathbf{q}}{\partial \theta \partial r} = \begin{bmatrix} -\sin \theta \\ \cos \theta \end{bmatrix}, \frac{\partial^2 \mathbf{q}}{\partial r^2} = 0, \frac{\partial^2 \mathbf{q}}{\partial \theta^2} = \begin{bmatrix} -r \cos \theta \\ -r \sin \theta \end{bmatrix}
\end{aligned}$$

so

$$\begin{aligned}
\frac{\partial}{\partial r} \begin{bmatrix} \frac{\partial \mathbf{q}}{\partial r} & \frac{\partial \mathbf{q}}{\partial \theta} \end{bmatrix} &= \begin{bmatrix} \frac{\partial^2 \mathbf{q}}{\partial r \partial r} & \frac{\partial^2 \mathbf{q}}{\partial r \partial \theta} \end{bmatrix} = \begin{bmatrix} \frac{\partial \mathbf{q}}{\partial r} & \frac{\partial \mathbf{q}}{\partial \theta} \end{bmatrix} \begin{bmatrix} 0 & 0 \\ 0 & \frac{1}{r} \end{bmatrix}, \\
\frac{\partial}{\partial \theta} \begin{bmatrix} \frac{\partial \mathbf{q}}{\partial r} & \frac{\partial \mathbf{q}}{\partial \theta} \end{bmatrix} &= \begin{bmatrix} \frac{\partial^2 \mathbf{q}}{\partial \theta \partial r} & \frac{\partial^2 \mathbf{q}}{\partial \theta \partial \theta} \end{bmatrix} = \begin{bmatrix} \frac{\partial \mathbf{q}}{\partial r} & \frac{\partial \mathbf{q}}{\partial \theta} \end{bmatrix} \begin{bmatrix} 0 & -r \\ \frac{1}{r} & 0 \end{bmatrix}
\end{aligned}$$

and

$$\begin{aligned} [\Gamma_r] &= \begin{bmatrix} 0 & 0 \\ 0 & \frac{1}{r} \end{bmatrix}, \\ [\Gamma_\theta] &= \begin{bmatrix} 0 & -r \\ \frac{1}{r} & 0 \end{bmatrix}. \end{aligned}$$

The key is that only Cartesian coordinates have the property that its coordinate fields are constant. When using general coordinates we are naturally forced to find these quantities. To see why this is, consider a curve $\mathbf{q}(t) = (r(t)\cos\theta(t), r(t)\sin\theta(t))$ in the plane. Its velocity is naturally given by

$$\dot{\mathbf{q}} = \dot{r}\frac{\partial \mathbf{q}}{\partial r} + \dot{\theta}\frac{\partial \mathbf{q}}{\partial \theta}.$$

If we wish to calculate its acceleration, then we must compute the derivatives of the coordinate fields. This involves the chain rule as well as the formulas just developed

$$\begin{aligned} \ddot{\mathbf{q}} &= \ddot{r}\frac{\partial \mathbf{q}}{\partial r} + \ddot{\theta}\frac{\partial \mathbf{q}}{\partial \theta} + \dot{r}\frac{d}{dt}\frac{\partial \mathbf{q}}{\partial r} + \dot{\theta}\frac{d}{dt}\frac{\partial \mathbf{q}}{\partial \theta} \\ &= \ddot{r}\frac{\partial \mathbf{q}}{\partial r} + \ddot{\theta}\frac{\partial \mathbf{q}}{\partial \theta} + \dot{r}\left(\frac{dr}{dt}\frac{\partial}{\partial r} + \frac{d\theta}{dt}\frac{\partial}{\partial \theta}\right)\frac{\partial \mathbf{q}}{\partial r} + \dot{\theta}\left(\frac{dr}{dt}\frac{\partial}{\partial r} + \frac{d\theta}{dt}\frac{\partial}{\partial \theta}\right)\frac{\partial \mathbf{q}}{\partial \theta} \\ &= \ddot{r}\frac{\partial \mathbf{q}}{\partial r} + \ddot{\theta}\frac{\partial \mathbf{q}}{\partial \theta} + \dot{r}^2\frac{\partial^2 \mathbf{q}}{\partial r^2} + 2\dot{r}\dot{\theta}\frac{\partial^2 \mathbf{q}}{\partial r \partial \theta} + \dot{\theta}^2\frac{\partial^2 \mathbf{q}}{\partial \theta^2} \\ &= \ddot{r}\frac{\partial \mathbf{q}}{\partial r} + \ddot{\theta}\frac{\partial \mathbf{q}}{\partial \theta} + 2\dot{r}\dot{\theta}\frac{1}{r}\frac{\partial \mathbf{q}}{\partial \theta} - \dot{\theta}^2 r\frac{\partial \mathbf{q}}{\partial r} \\ &= \left(\ddot{r} - r\dot{\theta}^2\right)\frac{\partial \mathbf{q}}{\partial r} + \left(\ddot{\theta} + \frac{2\dot{r}\dot{\theta}}{r}\right)\frac{\partial \mathbf{q}}{\partial \theta}. \end{aligned}$$

Note that $r\dot{\theta}^2$ corresponds to the centrifugal force that you feel when forced to move in a circle.

The term $\ddot{\theta} + \frac{2\dot{r}\dot{\theta}}{r}$ is related to Kepler's second law under a central force field. In that context

$$\ddot{\theta} + \frac{2\dot{r}\dot{\theta}}{r} = 0$$

as the force and hence acceleration is radial. This in turn implies that $r^2\dot{\theta}$ is constant as Kepler's law states.

The general goal will be to develop a similar set of ideas for surfaces and in addition to find other ways of calculating $[\Gamma_w]$ that depend on the geometry of the tangent fields.

Before generalizing we make another rather startling observation. Taking one more derivative we obtain

$$\begin{aligned} \frac{\partial^2}{\partial w_2 \partial w_1}\begin{bmatrix} \frac{\partial \mathbf{q}}{\partial u} & \frac{\partial \mathbf{q}}{\partial v} \end{bmatrix} &= \frac{\partial}{\partial w_2}\left(\begin{bmatrix} \frac{\partial \mathbf{q}}{\partial u} & \frac{\partial \mathbf{q}}{\partial v} \end{bmatrix}[\Gamma_{w_1}]\right) \\ &= \left(\frac{\partial}{\partial w_2}\begin{bmatrix} \frac{\partial \mathbf{q}}{\partial u} & \frac{\partial \mathbf{q}}{\partial v} \end{bmatrix}\right)[\Gamma_{w_1}] + \begin{bmatrix} \frac{\partial \mathbf{q}}{\partial u} & \frac{\partial \mathbf{q}}{\partial v} \end{bmatrix}\left[\frac{\partial \Gamma_{w_1}}{\partial w_2}\right] \\ &= \begin{bmatrix} \frac{\partial \mathbf{q}}{\partial u} & \frac{\partial \mathbf{q}}{\partial v} \end{bmatrix}[\Gamma_{w_2}][\Gamma_{w_1}] + \begin{bmatrix} \frac{\partial \mathbf{q}}{\partial u} & \frac{\partial \mathbf{q}}{\partial v} \end{bmatrix}\left[\frac{\partial \Gamma_{w_1}}{\partial w_2}\right] \\ &= \begin{bmatrix} \frac{\partial \mathbf{q}}{\partial u} & \frac{\partial \mathbf{q}}{\partial v} \end{bmatrix}\left([\Gamma_{w_2}][\Gamma_{w_1}] + \left[\frac{\partial \Gamma_{w_1}}{\partial w_2}\right]\right). \end{aligned}$$

Switching the order of the derivatives cannot not change the outcome,

$$\frac{\partial^2}{\partial w_1 \partial w_2}\left[\begin{array}{cc}\frac{\partial \mathbf{q}}{\partial u} & \frac{\partial \mathbf{q}}{\partial v}\end{array}\right] = \left[\begin{array}{cc}\frac{\partial \mathbf{q}}{\partial u} & \frac{\partial \mathbf{q}}{\partial v}\end{array}\right]\left(\left[\Gamma_{w_1}\right]\left[\Gamma_{w_2}\right] + \left[\frac{\partial \Gamma_{w_2}}{\partial w_1}\right]\right).$$

But the answer does look different when we use $w_1 = u$ and $w_2 = v$. Therefore, we can conclude that

$$\left[\Gamma_v\right]\left[\Gamma_u\right] + \left[\frac{\partial \Gamma_u}{\partial v}\right] = \left[\Gamma_u\right]\left[\Gamma_v\right] + \left[\frac{\partial \Gamma_v}{\partial u}\right]$$

or

$$\left[\frac{\partial \Gamma_v}{\partial u}\right] - \left[\frac{\partial \Gamma_u}{\partial v}\right] + \left[\Gamma_u\right]\left[\Gamma_v\right] - \left[\Gamma_v\right]\left[\Gamma_u\right] = 0.$$

For polar coordinates this can be verified directly:

$$\begin{aligned}
\left[\frac{\partial \Gamma_r}{\partial \theta}\right] - \left[\frac{\partial \Gamma_\theta}{\partial r}\right] &= 0 - \left[\begin{array}{cc} 0 & -1 \\ -\frac{1}{r^2} & 0 \end{array}\right] = \left[\begin{array}{cc} 0 & 1 \\ \frac{1}{r^2} & 0 \end{array}\right], \\
\left[\Gamma_r\right]\left[\Gamma_\theta\right] - \left[\Gamma_\theta\right]\left[\Gamma_r\right] &= \left[\begin{array}{cc} 0 & 0 \\ 0 & \frac{1}{r} \end{array}\right]\left[\begin{array}{cc} 0 & -r \\ \frac{1}{r} & 0 \end{array}\right] - \left[\begin{array}{cc} 0 & -r \\ \frac{1}{r} & 0 \end{array}\right]\left[\begin{array}{cc} 0 & 0 \\ 0 & \frac{1}{r} \end{array}\right] \\
&= \left[\begin{array}{cc} 0 & 1 \\ \frac{1}{r^2} & 0 \end{array}\right].
\end{aligned}$$

This means that the two matrices of functions $\left[\Gamma_u\right], \left[\Gamma_v\right]$ have some nontrivial relations between them that are not evident from the definition.

For a surface $\mathbf{q}(u, v)$ in $\mathbb{R}^3$ we add to the tangent vectors the unit normal

$$\mathbf{N}(u, v) = \frac{\frac{\partial \mathbf{q}}{\partial u} \times \frac{\partial \mathbf{q}}{\partial v}}{\left|\frac{\partial \mathbf{q}}{\partial u} \times \frac{\partial \mathbf{q}}{\partial v}\right|}$$

in order to get a basis. While $\mathbf{N}$ does depend on the parametrizations we note that as it is normal to a plane in $\mathbb{R}^3$ there are in fact only two choices $\pm\mathbf{N}$, just as with planar curves.

This means that we consider frames $\left[\begin{array}{ccc}\frac{\partial \mathbf{q}}{\partial u} & \frac{\partial \mathbf{q}}{\partial v} & \mathbf{N}\end{array}\right]$ and derivatives of such frames

$$\frac{\partial}{\partial w}\left[\begin{array}{ccc}\frac{\partial \mathbf{q}}{\partial u} & \frac{\partial \mathbf{q}}{\partial v} & \mathbf{N}\end{array}\right] = \left[\begin{array}{ccc}\frac{\partial^2 \mathbf{q}}{\partial w \partial u} & \frac{\partial^2 \mathbf{q}}{\partial w \partial v} & \frac{\partial \mathbf{N}}{\partial w}\end{array}\right] = \left[\begin{array}{ccc}\frac{\partial \mathbf{q}}{\partial u} & \frac{\partial \mathbf{q}}{\partial v} & \mathbf{N}\end{array}\right]\left[D_w\right],$$

where w can be either u or v.

The entries of D_w are divided up into parts or blocks. The principal 2×2 block consisting of what appears in the first two rows and columns depends only on tangential information. This block corresponds to the $\left[\Gamma_w\right]$ that we defined in the plane using general coordinates. The remaining parts, consisting of the third row and third column, depend on normal information. Since $\mathbf{N}$ is a unit vector the 33 entry actually vanishes:

$$0 = \frac{\partial \left|\mathbf{N}\right|^2}{\partial w} = 2\mathbf{N} \cdot \frac{\partial \mathbf{N}}{\partial w}.$$

Showing that $\frac{\partial \mathbf{N}}{\partial w}$ lies in the tangent space and hence does not have a normal component.

As before we have

$$\frac{\partial^2}{\partial w_1 \partial w_2}\left[\begin{array}{ccc}\frac{\partial \mathbf{q}}{\partial u} & \frac{\partial \mathbf{q}}{\partial v} & \mathbf{N}\end{array}\right] = \frac{\partial^2}{\partial w_2 \partial w_1}\left[\begin{array}{ccc}\frac{\partial \mathbf{q}}{\partial u} & \frac{\partial \mathbf{q}}{\partial v} & \mathbf{N}\end{array}\right].$$

In particular,

$$[D_u]\,[D_v] + \left[\frac{\partial D_v}{\partial u}\right] = [D_v]\,[D_u] + \left[\frac{\partial D_u}{\partial v}\right]$$

or

$$\left[\frac{\partial D_v}{\partial u}\right] - \left[\frac{\partial D_u}{\partial v}\right] + [D_u]\,[D_v] - [D_v]\,[D_u] = 0.$$

As we shall see, other interesting features emerge when we try to restrict attention to the tangential and normal parts of these matrices.

Elie Cartan developed an approach that uses orthonormal bases. Thus he chose an orthonormal frame E_1, E_2, E_3 along part of a surface with the property that $E_3 = \mathbf{N}$ is normal to the surface. Consequently, E_1, E_2 form an orthonormal basis for the tangent space. The goal is again to take derivatives. For that purpose we can still use parameters

$$\frac{\partial}{\partial w}\begin{bmatrix} E_1 & E_2 & E_3 \end{bmatrix} = \begin{bmatrix} \frac{\partial E_1}{\partial w} & \frac{\partial E_2}{\partial w} & \frac{\partial E_3}{\partial w} \end{bmatrix} = \begin{bmatrix} E_1 & E_2 & E_3 \end{bmatrix}[D_w].$$

The first observation is that $[D_w]$ is skew-symmetric since we used an orthonormal basis:

$$\begin{bmatrix} E_1 & E_2 & E_3 \end{bmatrix}^t \begin{bmatrix} E_1 & E_2 & E_3 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

so

$$\begin{aligned} 0 &= \frac{\partial}{\partial w}\left(\begin{bmatrix} E_1 & E_2 & E_3 \end{bmatrix}^t \begin{bmatrix} E_1 & E_2 & E_3 \end{bmatrix}\right) \\ &= \left(\frac{\partial}{\partial w}\begin{bmatrix} E_1 & E_2 & E_3 \end{bmatrix}\right)^t \begin{bmatrix} E_1 & E_2 & E_3 \end{bmatrix} \\ &\quad + \begin{bmatrix} E_1 & E_2 & E_3 \end{bmatrix}^t \frac{\partial}{\partial w}\begin{bmatrix} E_1 & E_2 & E_3 \end{bmatrix} \\ &= \left(\begin{bmatrix} E_1 & E_2 & E_3 \end{bmatrix}[D_w]\right)^t \begin{bmatrix} E_1 & E_2 & E_3 \end{bmatrix} \\ &\quad + \begin{bmatrix} E_1 & E_2 & E_3 \end{bmatrix}^t \begin{bmatrix} E_1 & E_2 & E_3 \end{bmatrix}[D_w] \\ &= [D_w]^t + [D_w]. \end{aligned}$$

In particular, there will only be 3 entries to sort out. This is a significant reduction from what we had to deal with above. What is more, the entries can easily be found by computing the dot products

$$E_i \cdot \frac{\partial E_j}{\partial w}.$$

This is also in sharp contrast to what happens in the above situation as we shall see. Taking one more derivative will again yield a formula

$$\left[\frac{\partial D_{w_2}}{\partial w_1}\right] - \left[\frac{\partial D_{w_1}}{\partial w_2}\right] = [D_{w_2}]\,[D_{w_1}] - [D_{w_1}]\,[D_{w_2}],$$

where both sides are skew symmetric.

Given the simplicity of using orthonormal frames it is perhaps puzzling why one would bother developing the more cumbersome approach that uses coordinate fields. The answer lies, as with curves, in the unfortunate fact that it is often easier to find coordinate fields than orthonormal bases that are easy to work with.

Monge patches are prime examples. For specific examples and many theoretical developments, however, Cartan's approach has many advantages.

4.4. The First Fundamental Form

Let $\mathbf{q}(u,v) : U \to \mathbb{R}^3$ be a parametrized surface. At each point of this surface we have a basis

$$\begin{aligned} & \frac{\partial \mathbf{q}}{\partial u}(u,v), \\ & \frac{\partial \mathbf{q}}{\partial v}(u,v), \\ \mathbf{N}(u,v) &= \frac{\frac{\partial \mathbf{q}}{\partial u} \times \frac{\partial \mathbf{q}}{\partial v}}{\left|\frac{\partial \mathbf{q}}{\partial u} \times \frac{\partial \mathbf{q}}{\partial v}\right|}. \end{aligned}$$

These vectors are again parametrized by u, v. The first two vectors are tangent to the surface and give us an unnormalized version of the tangent vector for a curve, while the third is the normal and is naturally normalized just as the normal vector is for a curve. One of the issues that make surface theory more difficult than curve theory is that there is no canonical parametrization along the lines of the arclength parametrization for curves.

The *first fundamental form* is the symmetric positive definite form that comes from the matrix

$$\begin{aligned} [\mathrm{I}] &= \begin{bmatrix} \frac{\partial \mathbf{q}}{\partial u} & \frac{\partial \mathbf{q}}{\partial v} \end{bmatrix}^t \begin{bmatrix} \frac{\partial \mathbf{q}}{\partial u} & \frac{\partial \mathbf{q}}{\partial v} \end{bmatrix} \\ &= \begin{bmatrix} \frac{\partial \mathbf{q}}{\partial u} \cdot \frac{\partial \mathbf{q}}{\partial u} & \frac{\partial \mathbf{q}}{\partial u} \cdot \frac{\partial \mathbf{q}}{\partial v} \\ \frac{\partial \mathbf{q}}{\partial v} \cdot \frac{\partial \mathbf{q}}{\partial u} & \frac{\partial \mathbf{q}}{\partial v} \cdot \frac{\partial \mathbf{q}}{\partial v} \end{bmatrix} \\ &= \begin{bmatrix} g_{uu} & g_{uv} \\ g_{vu} & g_{vv} \end{bmatrix}. \end{aligned}$$

For a curve the analogous term would simply be the square of the speed

$$\left(\frac{d\mathbf{q}}{dt}\right)^t \frac{d\mathbf{q}}{dt} = \frac{d\mathbf{q}}{dt} \cdot \frac{d\mathbf{q}}{dt}.$$

The first fundamental form dictates how one computes dot products of vectors tangent to the surface assuming they are expanded according to the basis $\frac{\partial \mathbf{q}}{\partial u}, \frac{\partial \mathbf{q}}{\partial v}$. If

$$\begin{aligned} X &= X^u \frac{\partial \mathbf{q}}{\partial u} + X^v \frac{\partial \mathbf{q}}{\partial v} = \begin{bmatrix} \frac{\partial \mathbf{q}}{\partial u} & \frac{\partial \mathbf{q}}{\partial v} \end{bmatrix} \begin{bmatrix} X^u \\ X^v \end{bmatrix}, \\ Y &= Y^u \frac{\partial \mathbf{q}}{\partial u} + Y^v \frac{\partial \mathbf{q}}{\partial v} = \begin{bmatrix} \frac{\partial \mathbf{q}}{\partial u} & \frac{\partial \mathbf{q}}{\partial v} \end{bmatrix} \begin{bmatrix} Y^u \\ Y^v \end{bmatrix}, \end{aligned}$$

then

$$\begin{aligned}
\mathrm{I}(X,Y) &= \begin{bmatrix} X^u & X^v \end{bmatrix} \begin{bmatrix} g_{uu} & g_{uv} \\ g_{vu} & g_{vv} \end{bmatrix} \begin{bmatrix} Y^u \\ Y^v \end{bmatrix} \\
&= \begin{bmatrix} X^u & X^v \end{bmatrix} \begin{bmatrix} \frac{\partial \mathbf{q}}{\partial u} & \frac{\partial \mathbf{q}}{\partial v} \end{bmatrix}^t \begin{bmatrix} \frac{\partial \mathbf{q}}{\partial u} & \frac{\partial \mathbf{q}}{\partial v} \end{bmatrix} \begin{bmatrix} Y^u \\ Y^v \end{bmatrix} \\
&= \left(\begin{bmatrix} \frac{\partial \mathbf{q}}{\partial u} & \frac{\partial \mathbf{q}}{\partial v} \end{bmatrix} \begin{bmatrix} X^u \\ X^v \end{bmatrix} \right)^t \left(\begin{bmatrix} \frac{\partial \mathbf{q}}{\partial u} & \frac{\partial \mathbf{q}}{\partial v} \end{bmatrix} \begin{bmatrix} Y^u \\ Y^v \end{bmatrix} \right) \\
&= X^t Y \\
&= X \cdot Y.
\end{aligned}$$

In particular, we see that while the *metric coefficients* $g_{w_1 w_2}$ depend on our parametrization. The dot product $\mathrm{I}(X,Y)$ of two tangent vectors remains the same if we change parameters. Note that I stands for the bilinear form $\mathrm{I}(X,Y)$ which does not depend on parametrizations, while [I] is the matrix representation for a fixed parametrization.

Our first observation is that the normalization factor $\left| \frac{\partial \mathbf{q}}{\partial u} \times \frac{\partial \mathbf{q}}{\partial v} \right|$ can be computed from $[\mathrm{I}]$.

Definition 4.4.1. The area form of a parametrized surface is given by

$$\sqrt{\det [\mathrm{I}]}.$$

The next lemma shows that this is given by the area of the parallelogram spanned by $\frac{\partial \mathbf{q}}{\partial u}$, $\frac{\partial \mathbf{q}}{\partial v}$.

Lemma 4.4.2. *We have*

$$\left| \frac{\partial \mathbf{q}}{\partial u} \times \frac{\partial \mathbf{q}}{\partial v} \right|^2 = \det [\mathrm{I}] = g_{uu} g_{vv} - (g_{uv})^2 .$$

Proof. This is simply the observation that both sides of the equation are formulas for the square of the area of the parallelogram spanned by $\frac{\partial \mathbf{q}}{\partial u}$, $\frac{\partial \mathbf{q}}{\partial v}$, i.e.,

$$\left| \frac{\partial \mathbf{q}}{\partial u} \times \frac{\partial \mathbf{q}}{\partial v} \right|^2 = \left| \frac{\partial \mathbf{q}}{\partial u} \right|^2 \left| \frac{\partial \mathbf{q}}{\partial v} \right|^2 - \left(\frac{\partial \mathbf{q}}{\partial u} \cdot \frac{\partial \mathbf{q}}{\partial v} \right)^2 .$$

□

The inverse

$$[\mathrm{I}]^{-1} = \begin{bmatrix} g_{uu} & g_{uv} \\ g_{vu} & g_{vv} \end{bmatrix}^{-1} = \begin{bmatrix} g^{uu} & g^{uv} \\ g^{vu} & g^{vv} \end{bmatrix}$$

can be used to find the expansion of a tangent vector by computing its dots products with the basis:

Proposition 4.4.3. *If $X \in T_p M$, then*

$$\begin{aligned}
X &= \left(g^{uu} \left(X \cdot \frac{\partial \mathbf{q}}{\partial u} \right) + g^{uv} \left(X \cdot \frac{\partial \mathbf{q}}{\partial v} \right) \right) \frac{\partial \mathbf{q}}{\partial u} \\
&\quad + \left(g^{vu} \left(X \cdot \frac{\partial \mathbf{q}}{\partial u} \right) + g^{vv} \left(X \cdot \frac{\partial \mathbf{q}}{\partial v} \right) \right) \frac{\partial \mathbf{q}}{\partial v} \\
&= \begin{bmatrix} \frac{\partial \mathbf{q}}{\partial u} & \frac{\partial \mathbf{q}}{\partial v} \end{bmatrix} [\mathrm{I}]^{-1} \begin{bmatrix} \frac{\partial \mathbf{q}}{\partial u} & \frac{\partial \mathbf{q}}{\partial v} \end{bmatrix}^t X.
\end{aligned}$$

More generally for any $Z \in \mathbb{R}^3$

$$
\begin{aligned}
Z &= \left(g^{uu}\left(Z \cdot \frac{\partial \mathbf{q}}{\partial u}\right) + g^{uv}\left(Z \cdot \frac{\partial \mathbf{q}}{\partial v}\right)\right)\frac{\partial \mathbf{q}}{\partial u} \\
&\quad + \left(g^{vu}\left(Z \cdot \frac{\partial \mathbf{q}}{\partial u}\right) + g^{vv}\left(Z \cdot \frac{\partial \mathbf{q}}{\partial v}\right)\right)\frac{\partial \mathbf{q}}{\partial v} + (Z \cdot \mathbf{N})\,\mathbf{N} \\
&= \begin{bmatrix} \frac{\partial \mathbf{q}}{\partial u} & \frac{\partial \mathbf{q}}{\partial v} \end{bmatrix} [\mathrm{I}]^{-1} \begin{bmatrix} \frac{\partial \mathbf{q}}{\partial u} & \frac{\partial \mathbf{q}}{\partial v} \end{bmatrix}^t Z + (Z \cdot \mathbf{N})\,\mathbf{N}.
\end{aligned}
$$

PROOF. This formula works for $X \in T_pM$ by writing

$$
X = \begin{bmatrix} \frac{\partial \mathbf{q}}{\partial u} & \frac{\partial \mathbf{q}}{\partial v} \end{bmatrix} \begin{bmatrix} X^u \\ X^v \end{bmatrix} = X^u \frac{\partial \mathbf{q}}{\partial u} + X^v \frac{\partial \mathbf{q}}{\partial v}
$$

and then observing that

$$
\begin{aligned}
\begin{bmatrix} \frac{\partial \mathbf{q}}{\partial u} & \frac{\partial \mathbf{q}}{\partial v} \end{bmatrix} [\mathrm{I}]^{-1} \begin{bmatrix} \frac{\partial \mathbf{q}}{\partial u} & \frac{\partial \mathbf{q}}{\partial v} \end{bmatrix}^t X &= \begin{bmatrix} \frac{\partial \mathbf{q}}{\partial u} & \frac{\partial \mathbf{q}}{\partial v} \end{bmatrix} [\mathrm{I}]^{-1} \begin{bmatrix} \frac{\partial \mathbf{q}}{\partial u} & \frac{\partial \mathbf{q}}{\partial v} \end{bmatrix}^t \begin{bmatrix} \frac{\partial \mathbf{q}}{\partial u} & \frac{\partial \mathbf{q}}{\partial v} \end{bmatrix} \begin{bmatrix} X^u \\ X^v \end{bmatrix} \\
&= \begin{bmatrix} \frac{\partial \mathbf{q}}{\partial u} & \frac{\partial \mathbf{q}}{\partial v} \end{bmatrix} [\mathrm{I}]^{-1} [\mathrm{I}] \begin{bmatrix} X^u \\ X^v \end{bmatrix} \\
&= \begin{bmatrix} \frac{\partial \mathbf{q}}{\partial u} & \frac{\partial \mathbf{q}}{\partial v} \end{bmatrix} \begin{bmatrix} X^u \\ X^v \end{bmatrix} \\
&= X.
\end{aligned}
$$

Note that the operation

$$
\begin{bmatrix} \frac{\partial \mathbf{q}}{\partial u} & \frac{\partial \mathbf{q}}{\partial v} \end{bmatrix} [\mathrm{I}]^{-1} \begin{bmatrix} \frac{\partial \mathbf{q}}{\partial u} & \frac{\partial \mathbf{q}}{\partial v} \end{bmatrix}^t
$$

can be applied to any vector in $\mathbb{R}^3$ and that its kernel is spanned by $\mathbf{N}$. In fact, it orthogonally projects the vector to a vector in the tangent space. For a general vector $Z \in \mathbb{R}^3$ the result then easily follows by decomposing it

$$
\begin{aligned}
Z &= X + (Z \cdot \mathbf{N})\,\mathbf{N}, \\
X &= Z - (Z \cdot \mathbf{N})\,\mathbf{N}
\end{aligned}
$$

and then using that we know what happens to X. □

Defining the gradient of a function is another important use of the first fundamental form as well as its inverse. Let $f(u, v)$ be viewed as a function on the surface $\mathbf{q}(u, v)$. Our definition of the gradient should definitely be so that it conforms with the chain rule for a curve $c(t) = \mathbf{q}(u(t), v(t))$. Thus on one hand we want

$$
\begin{aligned}
\frac{d(f \circ c)}{dt} &= \nabla f \cdot \dot{c} \\
&= \begin{bmatrix} (\nabla f)^u & (\nabla f)^v \end{bmatrix} [\mathrm{I}] \begin{bmatrix} \frac{du}{dt} \\ \frac{dv}{dt} \end{bmatrix}
\end{aligned}
$$

while the chain rule also dictates

$$
\frac{d(f \circ c)}{dt} = \begin{bmatrix} \frac{\partial f}{\partial u} & \frac{\partial f}{\partial v} \end{bmatrix} \begin{bmatrix} \frac{du}{dt} \\ \frac{dv}{dt} \end{bmatrix}.
$$

Thus

$$
\begin{bmatrix} (\nabla f)^u & (\nabla f)^v \end{bmatrix} = \begin{bmatrix} \frac{\partial f}{\partial u} & \frac{\partial f}{\partial v} \end{bmatrix} [\mathrm{I}]^{-1}
$$

or

$$\begin{aligned}
\nabla f &= \begin{bmatrix} \frac{\partial \mathbf{q}}{\partial u} & \frac{\partial \mathbf{q}}{\partial v} \end{bmatrix} \begin{bmatrix} (\nabla f)^u \\ (\nabla f)^v \end{bmatrix} \\
&= \begin{bmatrix} \frac{\partial \mathbf{q}}{\partial u} & \frac{\partial \mathbf{q}}{\partial v} \end{bmatrix} \left(\begin{bmatrix} \frac{\partial f}{\partial u} & \frac{\partial f}{\partial v} \end{bmatrix} [\mathrm{I}]^{-1} \right)^t \\
&= \begin{bmatrix} \frac{\partial \mathbf{q}}{\partial u} & \frac{\partial \mathbf{q}}{\partial v} \end{bmatrix} [\mathrm{I}]^{-1} \begin{bmatrix} \frac{\partial f}{\partial u} & \frac{\partial f}{\partial v} \end{bmatrix}^t \\
&= \left(g^{uu} \frac{\partial f}{\partial u} + g^{uv} \frac{\partial f}{\partial v} \right) \frac{\partial \mathbf{q}}{\partial u} + \left(g^{vu} \frac{\partial f}{\partial u} + g^{vv} \frac{\partial f}{\partial v} \right) \frac{\partial \mathbf{q}}{\partial v}.
\end{aligned}$$

In particular, we see that changing coordinates changes the gradient in such a way that it isn't simply the vector corresponding to the partial derivatives! The other nice feature is that we now have a concept of the gradient that gives a vector field independently of parametrizations. The defining equation

$$\frac{d\,(f \circ c)}{dt} = \nabla f \cdot \dot{c} = \mathrm{I}\,(\nabla f, \dot{c})$$

gives an implicit definition of ∇f that makes sense without reference to parametrizations of the surface.

Exercises.

(1) For a surface of revolution $\mathbf{q}\,(t, \mu) = (r\,(t) \cos \mu, r\,(t) \sin \mu, z\,(t))$ (see section 4.1 exercise 5) show that the first fundamental form is given by

$$\begin{bmatrix} g_{tt} & g_{t\mu} \\ g_{\mu t} & g_{\mu\mu} \end{bmatrix} = \begin{bmatrix} \dot{r}^2 + \dot{z}^2 & 0 \\ 0 & r^2 \end{bmatrix}.$$

A special and important case of this occurs when $z = 0$ and $r = t$ as that corresponds to polar coordinates in the plane.

(2) Assume that we have a cone (see section 4.1 exercise 2) given by

$$\mathbf{q}\,(r, \phi) = r\delta\,(\phi)\,,$$

where δ is a space curve with $|\delta| = 1$ and $\left|\frac{d\delta}{d\phi}\right| = 1$. Show that the first fundamental form is given by

$$\begin{bmatrix} g_{rr} & g_{r\phi} \\ g_{\phi r} & g_{\phi\phi} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & r^2 \end{bmatrix}$$

and compare this to polar coordinates in the plane.

(3) Assume that we have a generalized cylinder (see section 4.1 exercise 1) given by

$$\mathbf{q}\,(s, t) = (x\,(s)\,, y\,(s)\,, t)\,,$$

where $(x\,(s)\,, y\,(s))$ is unit speed. Show that the first fundamental form is given by

$$\begin{bmatrix} g_{ss} & g_{st} \\ g_{ts} & g_{tt} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

(4) Assume that we have a ruled surface (see section 4.1 exercise 4) given by

$$\mathbf{q}\,(s, t) = \alpha\,(t) + sX\,(t)\,,$$

where α is a space curve and X is a unit vector field based along this curve. Show that the first fundamental form is given by

$$\begin{bmatrix} g_{ss} & g_{st} \\ g_{ts} & g_{tt} \end{bmatrix} = \begin{bmatrix} 1 & \frac{d\alpha}{ds} \cdot X \\ \frac{d\alpha}{ds} \cdot X & \left|\frac{d\alpha}{dt} + s\frac{dX}{dt}\right|^2 \end{bmatrix}.$$

(5) Show that if we have a parametrized surface $\mathbf{q}(r,\theta)$ such that the first fundamental form is given by

$$\begin{bmatrix} g_{rr} & g_{r\theta} \\ g_{\theta r} & g_{\theta\theta} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & r^2 \end{bmatrix},$$

then we can locally reparametrize the surface to $\mathbf{q}(u,v)$ where the new first fundamental form is

$$\begin{bmatrix} g_{uu} & g_{uv} \\ g_{vu} & g_{vv} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

Hint: Let $u = r\cos\theta$ and $v = r\sin\theta$.

(6) Let $\alpha(s)$ be a unit speed curve with non-zero curvature, binormal $\mathbf{B}_\alpha$ and torsion τ. Show that the first fundamental form for the ruled surface

$$\mathbf{q}(s,t) = \alpha(s) + t\mathbf{B}_\alpha(s)$$

is given by

$$\begin{bmatrix} g_{ss} & g_{st} \\ g_{ts} & g_{tt} \end{bmatrix} = \begin{bmatrix} 1 + t^2\tau^2 & 0 \\ 0 & 1 \end{bmatrix}.$$

(7) Compute the first fundamental form of the Möbius band

$$\mathbf{q}(t,\phi) = (\cos\phi, \sin\phi, 0) + t\left(\sin\frac{\phi}{2}\cos\phi, \sin\frac{\phi}{2}\sin\phi, \cos\frac{\phi}{2}\right).$$

(8) Assume a surface has a parametrization $\mathbf{q}(s,\mu)$ where

$$\begin{bmatrix} g_{ss} & g_{s\mu} \\ g_{\mu s} & g_{\mu\mu} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & r^2 \end{bmatrix},$$

where $r(s)$ is only a function of s.

(a) Show that if $0 < \frac{dr}{ds} < 1$, then there is a function $z(s)$ so that $(r(s), 0, z(s))$ is a unit speed curve.

(b) Conclude that there is a surface of revolution with the same first fundamental form.

(9) Assume a surface has a parametrization $\mathbf{q}(u,v)$ where

$$\begin{bmatrix} g_{uu} & g_{uv} \\ g_{vu} & g_{vv} \end{bmatrix} = \begin{bmatrix} r^2 & 0 \\ 0 & r^2 \end{bmatrix},$$

where $r(u) > 0$ is only a function of u. Show that there is a reparametrization $u = u(s)$ such that the first fundamental form becomes

$$\begin{bmatrix} g_{ss} & g_{sv} \\ g_{vs} & g_{vv} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & r^2 \end{bmatrix}.$$

(10) For a parametrized surface $\mathbf{q}(u,v)$ show that

$$\mathbf{N} \times \frac{\partial \mathbf{q}}{\partial u} = \frac{g_{uu}\frac{\partial \mathbf{q}}{\partial v} - g_{uv}\frac{\partial \mathbf{q}}{\partial u}}{\left|\frac{\partial \mathbf{q}}{\partial u} \times \frac{\partial \mathbf{q}}{\partial v}\right|},$$

$$\mathbf{N} \times \frac{\partial \mathbf{q}}{\partial v} = \frac{g_{uv} \frac{\partial \mathbf{q}}{\partial v} - g_{vv} \frac{\partial \mathbf{q}}{\partial u}}{\left| \frac{\partial \mathbf{q}}{\partial u} \times \frac{\partial \mathbf{q}}{\partial v} \right|}.$$

(11) If we have a parametrization where

$$[\mathrm{I}] = \begin{bmatrix} 1 & 0 \\ 0 & g_{vv} \end{bmatrix},$$

then the coordinate function $f(u,v) = u$ has

$$\nabla u = \frac{\partial \mathbf{q}}{\partial u}.$$

(12) Show that it is always possible to find an orthogonal parametrization, i.e., g_{uv} vanishes. Hint: Use theorem 4.2.7.

(13) Show that if

$$\frac{\partial g_{uu}}{\partial v} = \frac{\partial g_{vv}}{\partial u} = g_{uv} = 0,$$

then we can reparametrize u and v separately, i.e., $u = u(s)$ and $v = v(t)$, in such a way that we obtain *Cartesian coordinates*:

$$\begin{aligned} g_{ss} &= g_{tt} = 1, \\ g_{st} &= 0. \end{aligned}$$

(14) Show that if

$$\frac{\partial^2 \mathbf{q}}{\partial u \partial v} = 0,$$

then

$$\mathbf{q}(u,v) = F(u) + G(v),$$

and conclude that

$$\frac{\partial g_{uu}}{\partial v} = \frac{\partial g_{vv}}{\partial u} = 0.$$

Give an example where $g_{uv} \neq 0$.

(15) Consider a unit speed curve $\alpha(s)$ with non-vanishing curvature and the tube (see section 4.1 exercise 8) of radius R around it

$$\mathbf{q}(s,\phi) = \alpha(s) + R(\mathbf{N}_\alpha \cos\phi + \mathbf{B}_\alpha \sin\phi),$$

where $\mathbf{T}_\alpha, \mathbf{N}_\alpha, \mathbf{B}_\alpha$ are the unit tangent, normal, and binormal to the curve.

(a) Show that $\mathbf{T}_\alpha$ and $-\mathbf{N}_\alpha \sin\phi + \mathbf{B}_\alpha \cos\phi$ are an orthonormal basis for the tangent space and that the normal to the tube is $\mathbf{N} = -(\mathbf{N}_\alpha \cos\phi + \mathbf{B}_\alpha \sin\phi)$.

(b) Show that

$$\begin{bmatrix} g_{ss} & g_{s\phi} \\ g_{\phi s} & g_{\phi\phi} \end{bmatrix} = \begin{bmatrix} (1-\kappa R)^2 + (\tau R)^2 & \tau R^2 \\ \tau R^2 & R^2 \end{bmatrix}.$$

4.5. Special Maps and Parametrizations

Definition 4.5.1. We call a map $F : M_1 \to M_2$ between surfaces an *isometry* if its differential preserves the first fundamental form

$$\mathrm{I}^{g_1}(X,Y) = \mathrm{I}^{g_2}(DF(X), DF(Y)).$$

We call the map *area preserving* if it preserves the areas of parallelograms spanned by vectors.

We call the map *conformal* if it preserves angles between vectors.

When the first surface is given as a parametrized surface these conditions can be quickly checked.

Proposition 4.5.2. *Let $\mathbf{q} : U \to M_1$ be a parametrization and $F : M_1 \to M_2$ a map. If $F \circ \mathbf{q}$ is also a parametrization, then the map is an isometry if*

$$[\mathrm{I}^{g_1}] = [\mathrm{I}^{g_2}],$$

area preserving if

$$\det[\mathrm{I}^{g_1}] = \det[\mathrm{I}^{g_2}],$$

and conformal if

$$[\mathrm{I}^{g_1}] = \lambda^2 [\mathrm{I}^{g_2}]$$

for some non-zero function λ.

Proof. Note that it is not necessary to first check that $F\circ\mathbf{q}$ is also a parametrization as that will be a consequence of any one of the three conditions if we define

$$[\mathrm{I}^{g_2}] = \begin{bmatrix} \frac{\partial F\circ\mathbf{q}}{\partial u}\cdot\frac{\partial F\circ\mathbf{q}}{\partial u} & \frac{\partial F\circ\mathbf{q}}{\partial u}\cdot\frac{\partial F\circ\mathbf{q}}{\partial v} \\ \frac{\partial F\circ\mathbf{q}}{\partial v}\cdot\frac{\partial F\circ\mathbf{q}}{\partial u} & \frac{\partial F\circ\mathbf{q}}{\partial v}\cdot\frac{\partial F\circ\mathbf{q}}{\partial v} \end{bmatrix}$$

and observe that $\frac{\partial F\circ\mathbf{q}}{\partial v}, \frac{\partial F\circ\mathbf{q}}{\partial u}$ are linearly independent if and only if the matrix $[\mathrm{I}^{g_2}]$ has nonzero determinant.

Next note that the chain rule implies that

$$DF\left(\frac{\partial\mathbf{q}}{\partial u}\right) = \frac{\partial F\circ\mathbf{q}}{\partial u},\ DF\left(\frac{\partial\mathbf{q}}{\partial v}\right) = \frac{\partial F\circ\mathbf{q}}{\partial v}$$

So all three conditions are necessarily true if the map is an isometry, area preserving, or conformal respectively. More generally, we see that

$$DF(X) = DF\left(X^u\frac{\partial\mathbf{q}}{\partial u} + X^v\frac{\partial\mathbf{q}}{\partial v}\right) = X^u\frac{\partial F\circ\mathbf{q}}{\partial u} + X^v\frac{\partial F\circ\mathbf{q}}{\partial v}.$$

So if $[\mathrm{I}^{g_1}] = [\mathrm{I}^{g_2}]$, then it follows from a direct but mildly long calculation that $\mathrm{I}^{g_1}(X,Y) = \mathrm{I}^{g_2}(DF(X), DF(Y))$.

Similarly, if

$$\left|\frac{\partial\mathbf{q}}{\partial u}\times\frac{\partial\mathbf{q}}{\partial v}\right|^2 = \det[\mathrm{I}^{g_1}] = \det[\mathrm{I}^{g_2}] = \left|\frac{\partial F\circ\mathbf{q}}{\partial u}\times\frac{\partial F\circ\mathbf{q}}{\partial v}\right|^2,$$

then it also follows from a somewhat tedious calculation that

$$|X\times Y|^2 = |DF(X)\times DF(Y)|^2.$$

Finally, as angles are given by

$$\cos\angle(X,Y) = \frac{X\cdot Y}{|X|\,|Y|}$$

the last statement follows by a similar calculation from the assumption that $[\mathrm{I}^{g_1}] = \lambda^2 [\mathrm{I}^{g_2}]$.

The last statement can also be rephrased without the use of λ by checking that

$$\frac{\frac{\partial F\circ\mathbf{q}}{\partial u}\cdot\frac{\partial F\circ\mathbf{q}}{\partial u}}{\frac{\partial\mathbf{q}}{\partial u}\cdot\frac{\partial\mathbf{q}}{\partial u}} = \frac{\frac{\partial F\circ\mathbf{q}}{\partial v}\cdot\frac{\partial F\circ\mathbf{q}}{\partial v}}{\frac{\partial\mathbf{q}}{\partial v}\cdot\frac{\partial\mathbf{q}}{\partial v}}$$

and

$$\frac{\partial F\circ\mathbf{q}}{\partial v}\cdot\frac{\partial F\circ\mathbf{q}}{\partial u} = \frac{\frac{\partial F\circ\mathbf{q}}{\partial u}\cdot\frac{\partial F\circ\mathbf{q}}{\partial u}}{\frac{\partial\mathbf{q}}{\partial u}\cdot\frac{\partial\mathbf{q}}{\partial u}}\frac{\partial\mathbf{q}}{\partial v}\cdot\frac{\partial\mathbf{q}}{\partial u}$$

□

DEFINITION 4.5.3. In case the map is a parametrization $\mathbf{q} : U \to M$ then we always use the *Cartesian metric* on U given by

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

So the parametrization is an isometry or *Cartesian* when

$$[\mathrm{I}] = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix},$$

area preserving when

$$\det[\mathrm{I}] = 1,$$

and conformal or *isothermal* when

$$g_{uu} = g_{vv},$$

$$g_{uv} = 0.$$

EXAMPLE 4.5.4. It follows from proposition 4.5.2 and example 4.2.13 that the Archimedes map is area preserving and the Mercator map is conformal.

DEFINITION 4.5.5. The *area* of a parametrized surface $\mathbf{q}(u, v) : U \to M$ over a region $R \subset U$ where $\mathbf{q}$ is one-to-one is defined by the integral

$$\mathrm{Area}(\mathbf{q}(R)) = \int_R \sqrt{\det[\mathrm{I}]}\,du\,dv.$$

PROPOSITION 4.5.6. *The area is independent under reparametrization.*

PROOF. Assume we have a different parametrization $\mathbf{q}(s, t) : V \to M$ and a new region $T \subset V$ with $\mathbf{q}(R) = \mathbf{q}(T)$ and the property that the reparametrization

$(u(s,t), v(s,t)) : T \to R$ is a diffeomorphism. Then

$$\begin{aligned}
\operatorname{Area}(\mathbf{q}(R)) &= \int_R \sqrt{\det[\mathrm{I}]}\,dudv \\
&= \int_R \sqrt{\det\left(\begin{bmatrix} \frac{\partial \mathbf{q}}{\partial u} & \frac{\partial \mathbf{q}}{\partial v} \end{bmatrix}^t \begin{bmatrix} \frac{\partial \mathbf{q}}{\partial u} & \frac{\partial \mathbf{q}}{\partial v} \end{bmatrix}\right)}\,dudv \\
&= \int_R \sqrt{\det\left(\left(\begin{bmatrix} \frac{\partial \mathbf{q}}{\partial s} & \frac{\partial \mathbf{q}}{\partial t} \end{bmatrix}\begin{bmatrix} \frac{\partial s}{\partial u} & \frac{\partial s}{\partial v} \\ \frac{\partial t}{\partial u} & \frac{\partial t}{\partial v} \end{bmatrix}\right)^t \begin{bmatrix} \frac{\partial \mathbf{q}}{\partial s} & \frac{\partial \mathbf{q}}{\partial t} \end{bmatrix}\begin{bmatrix} \frac{\partial s}{\partial u} & \frac{\partial s}{\partial v} \\ \frac{\partial t}{\partial u} & \frac{\partial t}{\partial v} \end{bmatrix}\right)}\,dudv \\
&= \int_R \sqrt{\det\left(\begin{bmatrix} \frac{\partial s}{\partial u} & \frac{\partial s}{\partial v} \\ \frac{\partial t}{\partial u} & \frac{\partial t}{\partial v} \end{bmatrix}^t \begin{bmatrix} \frac{\partial \mathbf{q}}{\partial s} & \frac{\partial \mathbf{q}}{\partial t} \end{bmatrix}^t \begin{bmatrix} \frac{\partial \mathbf{q}}{\partial s} & \frac{\partial \mathbf{q}}{\partial t} \end{bmatrix}\begin{bmatrix} \frac{\partial s}{\partial u} & \frac{\partial s}{\partial v} \\ \frac{\partial t}{\partial u} & \frac{\partial t}{\partial v} \end{bmatrix}\right)}\,dudv \\
&= \int_R \sqrt{\det\begin{bmatrix} \frac{\partial s}{\partial u} & \frac{\partial s}{\partial v} \\ \frac{\partial t}{\partial u} & \frac{\partial t}{\partial v} \end{bmatrix}^t}\det\sqrt{\begin{bmatrix} \frac{\partial \mathbf{q}}{\partial s} & \frac{\partial \mathbf{q}}{\partial t} \end{bmatrix}^t \begin{bmatrix} \frac{\partial \mathbf{q}}{\partial s} & \frac{\partial \mathbf{q}}{\partial t} \end{bmatrix}}\sqrt{\det\begin{bmatrix} \frac{\partial s}{\partial u} & \frac{\partial s}{\partial v} \\ \frac{\partial t}{\partial u} & \frac{\partial t}{\partial v} \end{bmatrix}}\,dudv \\
&= \int_R \det\sqrt{\begin{bmatrix} \frac{\partial \mathbf{q}}{\partial s} & \frac{\partial \mathbf{q}}{\partial t} \end{bmatrix}^t \begin{bmatrix} \frac{\partial \mathbf{q}}{\partial s} & \frac{\partial \mathbf{q}}{\partial t} \end{bmatrix}}\left|\det\begin{bmatrix} \frac{\partial s}{\partial u} & \frac{\partial s}{\partial v} \\ \frac{\partial t}{\partial u} & \frac{\partial t}{\partial v} \end{bmatrix}\right|\,dudv \\
&= \int_R \det\sqrt{\begin{bmatrix} \frac{\partial \mathbf{q}}{\partial s} & \frac{\partial \mathbf{q}}{\partial t} \end{bmatrix}^t \begin{bmatrix} \frac{\partial \mathbf{q}}{\partial s} & \frac{\partial \mathbf{q}}{\partial t} \end{bmatrix}}\,dsdt,
\end{aligned}$$

where the last equality follows from the change of variables formula for integrals. □

Exercises.

(1) Check if the parameterization $\mathbf{q}(t,\phi) = t(\cos\phi, \sin\phi, 1)$ for the cone is an isometry, area preserving, or conformal? Can the surface be reparametrized to have any of these properties? Hint: See section 4.4 exercise 2.

(2) Show that the following two parametrizations of the unit sphere are area preserving:

(a) (Lambert, 1772)

$$\mathbf{q}(\mu, z) = \begin{bmatrix} \sqrt{1-z^2}\cos\mu \\ \sqrt{1-z^2}\sin\mu \\ z \end{bmatrix}, \ |\mu| < \pi, \ |z| < 1.$$

(b) (Sinusoidal projection, Cossin, 1570)

$$\mathbf{q}(s,t) = \begin{bmatrix} \cos s\cos\left(\frac{t}{\cos s}\right) \\ \cos s\sin\left(\frac{t}{\cos s}\right) \\ \sin s \end{bmatrix}, \ |s| < \frac{\pi}{2}, \ t < \pi\cos s.$$

(c) Relate the Lambert parametrization to the Archimedes map.

(3) (Stabius-Werner, c. 1500, Sylvanus, 1511, Bonne, c. 1780) Show that the Bonne parametrizations

$$\mathbf{q}(r,\theta) = \begin{bmatrix} \cos(r-r_0)\cos\left(\frac{r(\theta-\pi/2)}{\cos(r-r_0)}\right) \\ \cos(r-r_0)\sin\left(\frac{r(\theta-\pi/2)}{\cos(r-r_0)}\right) \\ \sin(r-r_0) \end{bmatrix},$$

have the property that $\det[\mathrm{I}] = r^2$. Conclude that they are area preserving when (r,θ) correspond to polar coordinates

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

For $r_0 = 0$ this is a sinusoidal projection, for $r_0 = \pi/2$ the Stabius-Werner projection, and for $0 < r_0 < \pi/2$ the Sylvanus projection. The planar shape of these maps is bordered on the outside by an implicitly given curve

$$|r| \left|\theta - \tfrac{\pi}{2}\right| = \cos\left(r - r_0\right)$$

as $r \to \pi/2$ this looks a heart shaped region.

(4) Show that the inversion map

$$F\left(q\right) = \frac{q}{\left|q\right|^2}.$$

is a conformal map of $\mathbb{R}^n - 0$ to it self. Hint: See section 4.2 exercise 3.

(5) Show that the inverse stereographic projections $\mathbf{q}^{\pm} : \mathbb{R}^n \to \mathbb{R}^n \times \mathbb{R} = \mathbb{R}^{n+1}$ defined by

$$\mathbf{q}^{\pm}\left(q\right) = \left(q, 0\right) + \frac{1 - \left|q\right|^2}{1 + \left|q\right|^2}\left(q, \mp 1\right)$$

are conformal parametrizations of the unit sphere. Hint: See section 4.2 exercise 5. More specifically when $n = 2$ it is given by

$$\mathbf{q}^{\pm}\left(u, v\right) = \left(\frac{2u}{u^2 + v^2 + 1}, \frac{2v}{u^2 + v^2 + 1}, \frac{u^2 + v^2 \mp 1}{u^2 + v^2 + 1}\right).$$

(6) Consider the map $F : H \to \mathbb{R}^2$ defined by

$$\begin{aligned} F\left(x, y\right) &= \frac{1}{x^2 + \left(y + 1\right)^2}\left(2x, 2\left(y + 1\right)\right) + \left(0, -1\right) \\ &= \frac{1}{x^2 + \left(y + 1\right)^2}\left(2x, 1 - x^2 - y^2\right), \end{aligned}$$

where $H = \{(x, y) \mid y > 0\}$.

(a) Show that F is one-to-one and that the image is $D = \left\{(x, y) \mid x^2 + y^2 < 1\right\}$. Hint: Show that

$$\left|F\left(x, y\right)\right|^2 = 1 - \frac{4y}{x^2 + \left(y + 1\right)^2}.$$

(b) Show that the inverse is given by

$$\begin{aligned} F^{-1}\left(u, v\right) &= \frac{1}{u^2 + \left(v + 1\right)^2}\left(2u, 2\left(v + 1\right)\right) + \left(0, -1\right) \\ &= \frac{1}{u^2 + \left(v + 1\right)^2}\left(2u, 1 - u^2 - v^2\right). \end{aligned}$$

(c) Show that F and F^{-1} are conformal.

(d) Show that F can be interpreted as an inversion in the circle of radius $\sqrt{2}$ centered at $(0, -1)$.

(7) Show that Enneper's surface

$$\mathbf{q}\left(u, v\right) = \left(u - \frac{1}{3}u^3 + uv^2, v - \frac{1}{3}v^3 + vu^2, u^2 - v^2\right)$$

defines a conformal parametrization.

(8) Consider a map $F : S^2 \to P$, where $P = \{z = 1\}$ is the plane tangent to the North Pole, that takes each meridian to the radial line that is tangent to the meridian at the North Pole. Sometimes the map might just be defined on part of the sphere such as the upper hemisphere.
(a) Show that such a map has a parametrization of the form

$$F\left(\begin{bmatrix} \cos\mu\cos\phi \\ \sin\mu\cos\phi \\ \sin\phi \end{bmatrix}\right) = \begin{bmatrix} r(\phi)\cos\mu \\ r(\phi)\sin\mu \\ 1 \end{bmatrix}$$

for some function r, where $r\left(\frac{\pi}{2}\right) = 0$.
(b) Show that when $r = \sqrt{2(1-\sin\phi)}$, then we obtain an area preserving map on the upper hemisphere.
(c) Show that when the map projects a point on the upper hemisphere along the radial line through the origin, then $r = \cot\phi$. Show that this map takes all great circles (not just meridians) to straight lines. This is also called the Beltrami projection and is an example of a perspective projection (see section 4.2 exercise 4).
(d) Show that the inverse of the Beltrami projection from (c) onto the upper hemisphere is given by

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

(9) Show that a map $F : M \to M^*$ that is both conformal and area preserving is an isometry.

(10) Consider a ruled surface

$$\mathbf{q}(s,t) = \alpha(s) + tX(s)$$

where α is unit speed and X is a unit field. Show that it is conformal if and only if it is Cartesian (in which case X is constant and normal to α for all s.) Hint: See section 4.2 exercise 4.

(11) Find a conformal map from a surface of revolution $\mathbf{q}_1(r,\mu) = (r\cos\mu, r\sin\mu, z_1(r))$ to a circular cylinder $\mathbf{q}_2(r,\mu) = (\cos\mu, \sin r\mu, z_2(r))$.

(12) Reparametrize the curve $(r(u), z(u))$ so that the new parametrization

$$\mathbf{q}(t,\mu) = (r(t)\cos\mu, r(t)\sin\mu, z(t))$$

is conformal.

(13) Find an area preserving map from a surface of revolution $\mathbf{q}_1(r,\mu) = (r\cos\mu, r\sin\mu, z_1(r))$ to a circular cylinder $\mathbf{q}_2(r,\mu) = (\cos\mu, \sin\mu, z_2(r))$.

(14) Reparametrize the curve $(r(u), z(u))$ so that the new parametrization

$$\mathbf{q}(t,\mu) = (r(t)\cos\mu, r(t)\sin\mu, z(t))$$

is area preserving.

(15) Show that a Monge patch $z = f(x,y)$ is area preserving if and only if f is constant.

(16) Show that a Monge patch $z = f(x,y)$ is conformal if and only if f is constant.

(17) Show that the equation

$$ax + by + cz = d$$

defines a surface if and only if $(a,b,c) \neq (0,0,0)$. Show that this surface has a parametrization that is Cartesian.

(18) The *conoid* is a special type of ruled surface where α is a straight line and X always lies in a fixed plane. The simplest case is when α is the z-axis and X lies in the (x,y)-plane

$$\begin{aligned} \mathbf{q}(s,t) &= (tx(s), ty(s), z(s)) \\ &= (0,0,z(s)) + t(x(s), y(s), 0) \end{aligned}$$

(a) Compute its first fundamental form when $|X| = 1$.
(b) Show that this parametrization is conformal (or area preserving) if and only if the surface is a plane.
(c) Show that this surface is a helicoid when both X and z have constant speed.
(d) Show that such a helicoid can be reparametrized using $t = t(v)$ so as to obtain either a conformal or an area preserving parametrization.

(19) Consider the two parametrized surfaces given by

$$\begin{aligned} \mathbf{q}_1(\phi,u) &= (\sinh\phi\cos u, \sinh\phi\sin u, u) \\ &= (0,0,u) + \sinh\phi(\cos u, \sin u, 0) \\ \mathbf{q}_2(t,\mu) &= (\cosh t\cos\mu, \cosh t\sin\mu, t) \end{aligned}$$

Compute the first fundamental forms for both surfaces and construct a local isometry from the first surface to the second. (The first surface is a ruled surface with a one-to-one parametrization called the helicoid, the second surface is a surface of revolution called the catenoid.) Hint: See section 4.2 exercise 6.

(20) (Girard, 1626) A hemisphere on the unit sphere S^2 is the part that lies on one side of a great circle. A lune is the intersection of two hemispheres. It has two antipodal vertices. A spherical triangle is the region bounded by three hemispheres.
(a) Show that the area of a hemisphere is 2π.
(b) Use the Archimedes map to show that the area of a lune where the great circles meet at an angle of α is 2α.
(c) If $A(H)$ denotes the area of a region on S^2 use a Venn type diagram to show that

$$\begin{aligned} A(H_1 \cup H_2 \cup H_3) &= A(H_1) + A(H_2) + A(H_3) \\ &\quad - A(H_1 \cap H_3) - A(H_2 \cap H_3) - A(H_1 \cap H_2) \\ &\quad + A(H_1 \cap H_2 \cap H_3). \end{aligned}$$

(d) Let H_1, H_2, H_3 be hemispheres and $H_i' = S^2 - H_i$ the complementary hemispheres. Show that

$$H_1' \cap H_2' \cap H_3' = S^2 - H_1 \cup H_2 \cup H_3.$$

And further show that the spherical triangle $H_1 \cap H_2 \cap H_3$ is congruent to the spherical triangle $H_1' \cap H_2' \cap H_3'$ via the antipodal map.
(e) Show that the area A of a spherical triangle is given by

$$A = \alpha + \beta + \gamma - \pi,$$

where α, β, γ are the interior angles at the vertices of the triangle.

4.6. The Gauss Formulas

We are now going to compute the partial derivatives of our basis in both the u and v directions. Since these derivatives might not be tangential we get a formula that looks like

$$\frac{\partial^2 \mathbf{q}}{\partial w_1 \partial w_2} = \begin{bmatrix} \frac{\partial \mathbf{q}}{\partial u} & \frac{\partial \mathbf{q}}{\partial v} \end{bmatrix} [\mathrm{I}]^{-1} \begin{bmatrix} \frac{\partial \mathbf{q}}{\partial u} & \frac{\partial \mathbf{q}}{\partial v} \end{bmatrix}^t \frac{\partial^2 \mathbf{q}}{\partial w_1 \partial w_2} + \left(\frac{\partial^2 \mathbf{q}}{\partial w_1 \partial w_2} \cdot \mathbf{N} \right) \mathbf{N}.$$

The goal here and in the next two chapters is to first understand the normal part of this formula

$$\frac{\partial^2 \mathbf{q}}{\partial w_1 \partial w_2} \cdot \mathbf{N}$$

and then the tangential part

$$\Gamma_{w_1 w_2} = \begin{bmatrix} \frac{\partial \mathbf{q}}{\partial u} & \frac{\partial \mathbf{q}}{\partial v} \end{bmatrix} [\mathrm{I}]^{-1} \begin{bmatrix} \frac{\partial \mathbf{q}}{\partial u} & \frac{\partial \mathbf{q}}{\partial v} \end{bmatrix}^t \frac{\partial^2 \mathbf{q}}{\partial w_1 \partial w_2}.$$

Definition 4.6.1. The *Christoffel symbols of the first kind* are defined as

$$\Gamma_{w_1 w_2 w} = \frac{\partial^2 \mathbf{q}}{\partial w_1 \partial w_2} \cdot \frac{\partial \mathbf{q}}{\partial w}.$$

Definition 4.6.2. The *second fundamental form* with respect to the normal $\mathbf{N}$ is defined as

$$\begin{aligned} \mathrm{II}^{\mathbf{N}}(X,Y) &= \begin{bmatrix} X^u & X^v \end{bmatrix} [\mathrm{II}] \begin{bmatrix} Y^u \\ Y^v \end{bmatrix} \\ &= \begin{bmatrix} X^u & X^v \end{bmatrix} \begin{bmatrix} L_{uu} & L_{uv} \\ L_{vu} & L_{vv} \end{bmatrix} \begin{bmatrix} Y^u \\ Y^v \end{bmatrix}, \end{aligned}$$

where

$$L_{w_1 w_2} = \frac{\partial^2 \mathbf{q}}{\partial w_1 \partial w_2} \cdot \mathbf{N}.$$

The superscript $\mathbf{N}$ refers to the choice of normal and is usually suppressed since there are only two choices for the normal $\pm\mathbf{N}$. This also tells us that $\mathbf{N}\mathrm{II}$ is independent of the normal.

To further simplify expressions we also need to do the appropriate multiplication with $g^{w_4 w_5}$.

Definition 4.6.3. The *Christoffel symbols of the second kind* are defined as

$$\begin{aligned} \Gamma^w_{w_1 w_2} &= g^{wu}\Gamma_{w_1 w_2 u} + g^{wv}\Gamma_{w_1 w_2 v}, \\ \begin{bmatrix} \Gamma^u_{w_1 w_2} \\ \Gamma^v_{w_1 w_2} \end{bmatrix} &= \begin{bmatrix} g^{uu} & g^{uv} \\ g^{vu} & g^{vv} \end{bmatrix} \begin{bmatrix} \Gamma_{w_1 w_2 u} \\ \Gamma_{w_1 w_2 v} \end{bmatrix} \\ &= [\mathrm{I}]^{-1} \begin{bmatrix} \frac{\partial \mathbf{q}}{\partial u} & \frac{\partial \mathbf{q}}{\partial v} \end{bmatrix}^t \frac{\partial^2 \mathbf{q}}{\partial w_1 \partial w_2}. \end{aligned}$$

This now gives us the tangential component as

$$\begin{aligned} \Gamma_{w_1 w_2} &= \begin{bmatrix} \frac{\partial \mathbf{q}}{\partial u} & \frac{\partial \mathbf{q}}{\partial v} \end{bmatrix} [\mathrm{I}]^{-1} \begin{bmatrix} \frac{\partial \mathbf{q}}{\partial u} & \frac{\partial \mathbf{q}}{\partial v} \end{bmatrix}^t \frac{\partial^2 \mathbf{q}}{\partial w_1 \partial w_2} \\ &= \Gamma^u_{w_1 w_2} \frac{\partial \mathbf{q}}{\partial u} + \Gamma^v_{w_1 w_2} \frac{\partial \mathbf{q}}{\partial v}. \end{aligned}$$

The second derivatives of $\mathbf{q}(u,v)$ can now be expressed as follows in terms of the Christoffel symbols of the second kind and the second fundamental form. These are often called the *Gauss formulas*:

$$\begin{aligned}
\frac{\partial^2 \mathbf{q}}{\partial u^2} &= \Gamma_{uu}^{u}\frac{\partial \mathbf{q}}{\partial u} + \Gamma_{uu}^{v}\frac{\partial \mathbf{q}}{\partial v} + L_{uu}\mathbf{N}, \\
\frac{\partial^2 \mathbf{q}}{\partial u \partial v} &= \Gamma_{uv}^{u}\frac{\partial \mathbf{q}}{\partial u} + \Gamma_{uv}^{v}\frac{\partial \mathbf{q}}{\partial v} + L_{uv}\mathbf{N} = \frac{\partial^2 \mathbf{q}}{\partial v \partial u}, \\
\frac{\partial^2 \mathbf{q}}{\partial v^2} &= \Gamma_{vv}^{u}\frac{\partial \mathbf{q}}{\partial u} + \Gamma_{vv}^{v}\frac{\partial \mathbf{q}}{\partial v} + L_{vv}\mathbf{N},
\end{aligned}$$

or

$$\frac{\partial^2 \mathbf{q}}{\partial w_1 \partial w_2} = \Gamma_{w_1 w_2}^{u}\frac{\partial \mathbf{q}}{\partial u} + \Gamma_{w_1 w_2}^{v}\frac{\partial \mathbf{q}}{\partial v} + L_{w_1 w_2}\mathbf{N},$$

or

$$\frac{\partial}{\partial w}\left[\begin{array}{cc} \frac{\partial \mathbf{q}}{\partial u} & \frac{\partial \mathbf{q}}{\partial v} \end{array}\right] = \left[\begin{array}{ccc} \frac{\partial \mathbf{q}}{\partial u} & \frac{\partial \mathbf{q}}{\partial v} & \mathbf{N} \end{array}\right]\left[\begin{array}{cc} \Gamma_{wu}^{u} & \Gamma_{wv}^{u} \\ \Gamma_{wu}^{v} & \Gamma_{wv}^{v} \\ L_{wu} & L_{wv} \end{array}\right].$$

This means that we have introduced notation for the first two columns in $[D_w]$ from section 4.3. The last column is related to the last row and will also be examined in the next chapter.

As we shall see, and indeed already saw in section 4.3 when considering polar coordinates in the plane, these formulas are important for defining accelerations of curves. They are however also important for giving a proper definition of the Hessian or second derivative matrix of a function on a surface. This will be explored in an exercise later.

The task of calculating the second fundamental form is fairly straightforward, but will be postponed until the next chapter. Calculating the Christoffel symbols is more complicated and is delayed until we've gotten used to the second fundamental form.

Exercises.

(1) Show that
$$\frac{\partial \mathbf{N}}{\partial w}$$
is always tangent to the surface.

(2) Show that
$$\frac{\partial^2 \mathbf{q}}{\partial w_1 \partial w_2} \cdot \mathbf{N} = -\frac{\partial \mathbf{q}}{\partial w_2} \cdot \frac{\partial \mathbf{N}}{\partial w_1}.$$
This shows that the derivatives of the normal can be computed knowing the first and second fundamental forms.

(3) Show that [II] vanishes if and only if the normal vector is constant. Show in turn that this happens if and only if the surface is part of a plane.

(4) Show that when $\mathbf{q}(u,v)$ is a Cartesian parametrization, i.e.,
$$[\mathrm{I}] = \left[\begin{array}{cc} 1 & 0 \\ 0 & 1 \end{array}\right],$$
then the Christoffel symbols vanish. Hint: This is not obvious since we don't know $\frac{\partial^2 \mathbf{q}}{\partial w_1 \partial w_2}$.

CHAPTER 5

Extrinsic Geometry

The goal of *extrinsic geometry* is to study the shape of curves and surfaces through understanding how the tangent lines or space vary in relation to an extrinsic space such as Euclidean space. When we studied curves this was essentially all we did. For surfaces there is also a rich *intrinsic geometry* that only addresses concepts that can be calculated via the first fundamental form. This will be studied in chapter 6 and 7.

5.1. Curves on Surfaces

In this section we offer a geometric construction that allows us to show that the second fundamental form is, like the first fundamental form, defined independently of parametrizations. This will also be done more algebraically in section 5.2.

The key observation is that if we have a surface M and a point $p \in M$, then the tangent space T_pM and normal space $N_pM = (T_pM)^{\perp}$ are defined independently of our parametrizations (see proposition 4.2.3). Therefore, if we have a vector Z in Euclidean space then its projection onto both the tangent space and the normal space are also independently defined.

Consider a curve $\mathbf{q}(t)$ on the surface. We know that the velocity $\dot{\mathbf{q}}$ and acceleration $\ddot{\mathbf{q}}$ can be calculated without reference to parametrizations of the surface. This means that the projections of $\ddot{\mathbf{q}}$ onto the normal space, $\ddot{\mathbf{q}}^{\mathrm{II}} = (\ddot{\mathbf{q}} \cdot \mathbf{N})\,\mathbf{N}$, and the tangent space, $\ddot{\mathbf{q}}^{\mathrm{I}} = \dot{\mathbf{q}} - (\ddot{\mathbf{q}} \cdot \mathbf{N})\,\mathbf{N}$, can be computed without reference to parametrizations. This shows that tangential and normal accelerations are well-defined.

Theorem 5.1.1. *(Euler, 1760 and Meusnier, 1776) The normal component of the acceleration satisfies*

$$(\ddot{\mathbf{q}} \cdot \mathbf{N})\,\mathbf{N} = \ddot{\mathbf{q}}^{\mathrm{II}} = \mathbf{N}\mathrm{II}\,(\dot{\mathbf{q}}, \dot{\mathbf{q}})$$

In particular, two curves with the same velocity at a point have the same normal acceleration components.

Proof. We have to show that

$$\ddot{\mathbf{q}} \cdot \mathbf{N} = \mathrm{II}\,(\dot{\mathbf{q}}, \dot{\mathbf{q}})$$

To do so we select a parametrization and write $\mathbf{q}(t) = \mathbf{q}(u(t), v(t))$, then

$$\dot{\mathbf{q}} = \begin{bmatrix} \frac{\partial \mathbf{q}}{\partial u} & \frac{\partial \mathbf{q}}{\partial v} \end{bmatrix} \begin{bmatrix} \dot{u} \\ \dot{v} \end{bmatrix}$$

and

$$
\begin{aligned}
\ddot{\mathbf{q}} &= \left(\frac{d}{dt}\begin{bmatrix} \frac{\partial \mathbf{q}}{\partial u} & \frac{\partial \mathbf{q}}{\partial v} \end{bmatrix}\right)\begin{bmatrix} \dot{u} \\ \dot{v} \end{bmatrix} + \begin{bmatrix} \frac{\partial \mathbf{q}}{\partial u} & \frac{\partial \mathbf{q}}{\partial v} \end{bmatrix}\begin{bmatrix} \ddot{u} \\ \ddot{v} \end{bmatrix} \\
&= \begin{bmatrix} \dot{u} & \dot{v} \end{bmatrix}\begin{bmatrix} \frac{\partial^2 \mathbf{q}}{\partial u^2} & \frac{\partial^2 \mathbf{q}}{\partial v \partial u} \\ \frac{\partial^2 \mathbf{q}}{\partial u \partial v} & \frac{\partial^2 \mathbf{q}}{\partial v^2} \end{bmatrix}\begin{bmatrix} \dot{u} \\ \dot{v} \end{bmatrix} + \begin{bmatrix} \frac{\partial \mathbf{q}}{\partial u} & \frac{\partial \mathbf{q}}{\partial v} \end{bmatrix}\begin{bmatrix} \ddot{u} \\ \ddot{v} \end{bmatrix}.
\end{aligned}
$$

Taking inner products with the normal will eliminate the second term as it is a tangent vector so we obtain

$$
\begin{aligned}
\ddot{\mathbf{q}} \cdot \mathbf{N} &= \begin{bmatrix} \dot{u} & \dot{v} \end{bmatrix}\begin{bmatrix} \frac{\partial^2 \mathbf{q}}{\partial u^2} \cdot \mathbf{N} & \frac{\partial^2 \mathbf{q}}{\partial v \partial u} \cdot \mathbf{N} \\ \frac{\partial^2 \mathbf{q}}{\partial u \partial v} \cdot \mathbf{N} & \frac{\partial^2 \mathbf{q}}{\partial v^2} \cdot \mathbf{N} \end{bmatrix}\begin{bmatrix} \dot{u} \\ \dot{v} \end{bmatrix} \\
&= \begin{bmatrix} \dot{u} & \dot{v} \end{bmatrix}\begin{bmatrix} L_{uu} & L_{vu} \\ L_{uv} & L_{vv} \end{bmatrix}\begin{bmatrix} \dot{u} \\ \dot{v} \end{bmatrix} \\
&= \mathrm{II}\left(\dot{\mathbf{q}}, \dot{\mathbf{q}}\right).
\end{aligned}
$$

□

Based on the velocity characterization of the tangent space in proposition 4.2.4 we now obtain the following definition of the second fundamental form.

DEFINITION 5.1.2. We can now redefine $\mathbf{N}\mathrm{II}(Z, Z)$ by selecting a curve on the surface with $\dot{\mathbf{q}} = Z$ and then using that $\mathbf{N}\mathrm{II}(\dot{\mathbf{q}}, \dot{\mathbf{q}})$ is the normal component of $\ddot{\mathbf{q}}$.

To compute $\mathbf{N}\mathrm{II}(X, Y)$ we can use *polarization*:

$$
\mathbf{N}\mathrm{II}(X, Y) = \frac{1}{2}\left(\mathbf{N}\mathrm{II}(X + Y, X + Y) - \mathbf{N}\mathrm{II}(X, X) - \mathbf{N}\mathrm{II}(Y, Y)\right).
$$

As with space curves we define the unit tangent $\mathbf{T}$ for a regular curve $\mathbf{q}(t)$ on a surface $\mathbf{q}(u, v)$. However, we use the normal $\mathbf{N}$ to the surface instead of the principal normal component of the acceleration in relation to the unit tangent. From these two vectors we can define the normal to the curve tangent to the surface as

$$
\mathbf{S} = \mathbf{N} \times \mathbf{T}.
$$

In this way curve theory on surfaces is closer to the theory of planar curves as we can think of $\mathbf{S}$ as the signed normal to the curve in the surface (see also section 3.3 for the special case of curves on spheres). Using an arclength parameter s we define the *normal curvature*

$$
\kappa_n = \mathrm{II}(\mathbf{T}, \mathbf{T}),
$$

the *geodesic curvature*

$$
\kappa_g = \mathbf{S} \cdot \frac{d\mathbf{T}}{ds}
$$

and the *geodesic torsion*

$$
\tau_g = \mathbf{N} \cdot \frac{d\mathbf{S}}{ds}
$$

Note that the geodesic curvature of curves on the sphere from section 3.3 is consistent with the above definition.

EXAMPLE 5.1.3. A plane always has vanishing second fundamental form as its normal is constant. This means that any curve in this plane has vanishing normal curvature and geodesic torsion. The geodesic curvature is the signed curvature $\kappa_\pm$.

EXAMPLE 5.1.4. A sphere of radius R centered at $\mathbf{c}$ is given by the equation

$$F(x,y,z) = |\mathbf{q}-\mathbf{c}|^2 = R^2 > 0.$$

The gradient is

$$\nabla F = 2(\mathbf{q}-\mathbf{c}) = 2(x-a, y-b, z-c),$$

which cannot vanish unless $\mathbf{q}=\mathbf{c}$. This shows that the sphere is a surface and also computes the two normals

$$\mathbf{N} = \pm\frac{1}{R}(\mathbf{q}-\mathbf{c}).$$

as $|\mathbf{q}-\mathbf{c}| = R$ (see theorem 4.1.9 and corollary 4.2.5). The $+$ gives us an outward pointing normal. If we have a parametrization $\mathbf{q}(u,v)$, then this relationship between $\mathbf{q}$ and $\mathbf{N}$ implies that $\frac{\partial \mathbf{q}}{\partial w} = \pm\frac{1}{R}\frac{\partial \mathbf{N}}{\partial w}$. This in turn implies that $\mathrm{II} = \pm\frac{1}{R}\mathrm{I}$.

This shows that the normal curvature of any curve on the sphere is $\pm\frac{1}{R}$.

We shall show below that only planes and spheres have the property that the normal curvature is the same for all curves on a surface. Another interesting consequence is an important theorem first noted by Euler and later in greater generality by Gauss that it is not possible to draws maps of the Earth with the property that all distances and angles are preserved.

THEOREM 5.1.5. *(Euler, 1775) A sphere does not admit a Cartesian parametrization.*

PROOF. Assume that $\mathbf{q}(u,v)$ is a Cartesian parametrization of part of a sphere of radius $R>0$. We start by showing that any line $(u(t), v(t))$ with zero acceleration becomes part of a great circle $\mathbf{q}(t) = \mathbf{q}(u(t), v(t))$ on the sphere. Great circles are characterized as curves with acceleration normal to the sphere, i.e., the tangential acceleration vanishes $\ddot{\mathbf{q}}^{\mathrm{I}} = 0$ (see section 3.3 exercise 5 or exercise 7 in this section). By assumption $\ddot{u}, \ddot{v} = 0$ so

$$\ddot{\mathbf{q}} = \begin{bmatrix} \dot{u} & \dot{v} \end{bmatrix} \begin{bmatrix} \frac{\partial^2 \mathbf{q}}{\partial u^2} & \frac{\partial^2 \mathbf{q}}{\partial v \partial u} \\ \frac{\partial^2 \mathbf{q}}{\partial u \partial v} & \frac{\partial^2 \mathbf{q}}{\partial v^2} \end{bmatrix} \begin{bmatrix} \dot{u} \\ \dot{v} \end{bmatrix}.$$

Thus $\ddot{\mathbf{q}}^{\mathrm{I}} = 0$ precisely when

$$\begin{aligned} \frac{\partial^2 \mathbf{q}}{\partial u^2} \cdot \frac{\partial \mathbf{q}}{\partial w} &= 0, \\ \frac{\partial^2 \mathbf{q}}{\partial v \partial u} \cdot \frac{\partial \mathbf{q}}{\partial w} &= 0, \\ \frac{\partial^2 \mathbf{q}}{\partial v^2} \cdot \frac{\partial \mathbf{q}}{\partial w} &= 0. \end{aligned}$$

Note that as $w = u, v$ there are 6 identities. Using that $\frac{\partial \mathbf{q}}{\partial u}, \frac{\partial \mathbf{q}}{\partial v}$ are unit vectors we obtain

$$0 = \frac{\partial}{\partial w}\left(\frac{\partial \mathbf{q}}{\partial u} \cdot \frac{\partial \mathbf{q}}{\partial u}\right) = 2\frac{\partial^2 \mathbf{q}}{\partial w \partial u} \cdot \frac{\partial \mathbf{q}}{\partial u} = \frac{\partial^2 \mathbf{q}}{\partial u \partial w} \cdot \frac{\partial \mathbf{q}}{\partial u}$$

and similarly

$$0 = \frac{\partial}{\partial w}\left(\frac{\partial \mathbf{q}}{\partial v} \cdot \frac{\partial \mathbf{q}}{\partial v}\right) = 2\frac{\partial^2 \mathbf{q}}{\partial w \partial v} \cdot \frac{\partial \mathbf{q}}{\partial v} = \frac{\partial^2 \mathbf{q}}{\partial v \partial w} \cdot \frac{\partial \mathbf{q}}{\partial v}.$$

This shows that four of the identities hold. Next we use that $\frac{\partial \mathbf{q}}{\partial u}, \frac{\partial \mathbf{q}}{\partial v}$ are perpendicular to conclude

$$0 = \frac{\partial}{\partial w}\left(\frac{\partial \mathbf{q}}{\partial u} \cdot \frac{\partial \mathbf{q}}{\partial v}\right) = \frac{\partial^2 \mathbf{q}}{\partial w \partial u} \cdot \frac{\partial \mathbf{q}}{\partial v} + \frac{\partial \mathbf{q}}{\partial u} \cdot \frac{\partial^2 \mathbf{q}}{\partial w \partial v}.$$

Depending on whether $w = u$ or v the second or first term on the left vanishes from what we just did. Thus the remaining term also vanishes. This takes care of the last two identities.

To finish the proof it remains to observe that if we select a small triangle in the u, v plane, then it is mapped to a congruent spherical triangle whose sides are parts of great circles. This, however, violates the spherical law of cosines or Girard's theorem (see section 4.5 exercise 20). To give a self contained argument here select an equilateral triangle in the plane with side lengths ϵ. Then we obtain an equilateral triangle on the sphere with side lengths ϵ and interior angles $\frac{\pi}{3}$. As the sides are parts of great circles we can check explicitly if this is possible. Let the vertices be $\mathbf{q}_i$, $i = 1, 2, 3$, then $\mathbf{q}_i \cdot \mathbf{q}_j = \cos\epsilon$ when $i \neq j$. The unit directions of the great circles at $\mathbf{q}_1$ are given by

$$\begin{aligned}
\mathbf{v}_{12} &= \frac{\mathbf{q}_2 - (\mathbf{q}_2 \cdot \mathbf{q}_1)\,\mathbf{q}_1}{\sqrt{1 - (\mathbf{q}_2 \cdot \mathbf{q}_1)^2}} = \frac{\mathbf{q}_2 - \cos\epsilon \mathbf{q}_1}{\sin\epsilon}, \\
\mathbf{v}_{13} &= \frac{\mathbf{q}_3 - \cos\epsilon \mathbf{q}_1}{\sin\epsilon}
\end{aligned}$$

and thus

$$\begin{aligned}
\frac{1}{2} &= \mathbf{v}_{12} \cdot \mathbf{v}_{13} \\
&= \left(\frac{\mathbf{q}_2 - \cos\epsilon \mathbf{q}_1}{\sin\epsilon}\right) \cdot \left(\frac{\mathbf{q}_3 - \cos\epsilon \mathbf{q}_1}{\sin\epsilon}\right) \\
&= \frac{\cos\epsilon - 2\cos^2\epsilon + \cos^2\epsilon}{\sin^2\epsilon} \\
&= \frac{\cos\epsilon - \cos^2\epsilon}{\sin^2\epsilon} \\
&= \frac{1}{2} - \frac{1}{8}\epsilon^2 + \cdots \\
&< \frac{1}{2}.
\end{aligned}$$

So we have arrived at a contradiction. □

Finally we prove what has come to be known as the Gauss-Bonnet theorem, but for now only in the case when the surface is a sphere. This result can also be used to show that the sphere doesn't have Cartesian coordinates.

Theorem 5.1.6. *Let $\mathbf{q}(s)$ be a simple closed unit speed curve on the unit sphere. Let A be the area enclosed by the curve, then*

$$A = 2\pi \pm \int \kappa_g ds.$$

Remark 5.1.7. Note that the closed curve divides the sphere in to two regions. Together these areas add up to 4π and differ from each other by $2\int \kappa_g ds$.

Remark 5.1.8. Much of what we do will be valid in the more general context of a surface of revolution where the (r, z)-curve is parametrized by arclength (see section4.1 exercise 5 for notation). The one change is that we'll no longer be calculating the area of the region, but rather the integral of a quantity that is called the Gauss curvature. More precisely, if the closed curve bounds a region Ω, then

$$-\int\int_{\Omega}\frac{\frac{d^2r}{d\phi^2}}{r}\sqrt{\det[\mathrm{I}]}d\mu d\phi = 2\pi \pm \int \kappa_g ds.$$

Proof. We parametrize the unit sphere as a surface of revolution

$$\mathbf{q}(\mu,\phi) = \begin{bmatrix} \cos\mu\cos\phi \\ \sin\mu\cos\phi \\ \sin\phi \end{bmatrix} = \begin{bmatrix} r(\phi)\cos\mu \\ r(\phi)\sin\mu \\ z(\phi) \end{bmatrix}$$

We also restrict the domain to be $(\mu,\phi) \in R = (0, 2\pi) \times \left(-\frac{\pi}{2}, \frac{\pi}{2}\right)$ and consider a unit speed curve

$$\mathbf{q}(s) = \mathbf{q}(\mu(s), \phi(s)),$$

where $(\mu(s), \phi(s)) \in R$ for $s \in [0, L]$. The area under consideration will then be the area bounded by the curve inside this region. We shall further assume that the curve runs counter clockwise in this region so that $\mathbf{S}$ points inwards. Thus the rotation of the curve is 2π if we think of it as a planar curve on the rectangle.

The two vectors

$$E_\mu = \begin{bmatrix} -\sin\mu \\ \cos\mu \\ 0 \end{bmatrix} = \frac{1}{\cos\phi}\frac{\partial\mathbf{q}}{\partial\mu} = \frac{1}{r}\frac{\partial\mathbf{q}}{\partial\mu},$$

$$E_\phi = \frac{\partial\mathbf{q}}{\partial\phi} = \begin{bmatrix} -\sin\phi\cos\mu \\ -\sin\phi\sin\mu \\ \cos\phi \end{bmatrix}$$

form an orthonormal basis for the tangent space at every point. Therefore, the unit tangent can be written as

$$\mathbf{T} = \frac{d\mathbf{q}}{ds} = \cos(\theta(s))E_\mu + \sin(\theta(s))E_\phi$$

and

$$\mathbf{S} = -\sin\theta E_\mu + \cos\theta E_\phi$$

The geodesic curvature can then be computed as

$$\begin{aligned}
\kappa_g &= \mathbf{S}\cdot\frac{d\mathbf{T}}{ds} \\
&= \mathbf{S}\cdot\left(\frac{d\theta}{ds}\mathbf{S}+\cos\theta\frac{dE_\mu}{ds}+\sin\theta\frac{dE_\phi}{ds}\right) \\
&= \frac{d\theta}{ds}-\sin^2\theta E_\mu\cdot\frac{dE_\phi}{ds}+\cos^2\theta E_\phi\cdot\frac{dE_\mu}{ds} \\
&= \frac{d\theta}{ds}+E_\phi\cdot\frac{dE_\mu}{ds} \\
&= \frac{d\theta}{ds}+\begin{bmatrix}-\sin\phi\cos\mu\\-\sin\phi\sin\mu\\\cos\phi\end{bmatrix}\cdot\begin{bmatrix}-\cos\mu\\-\sin\mu\\0\end{bmatrix}\frac{d\mu}{ds} \\
&= \frac{d\theta}{ds}+\sin\phi\frac{d\mu}{ds} \\
&= \frac{d\theta}{ds}-\frac{dr}{d\phi}\frac{d\mu}{ds}.
\end{aligned}$$

We can use Green's theorem to conclude that

$$\int\frac{dr}{d\phi}\frac{d\mu}{ds}ds=-\int\int\frac{d^2r}{d\phi^2}d\mu d\phi.$$

Thus yielding

$$\begin{aligned}
\int\frac{d\theta}{ds}ds-\int\kappa_g ds &= \int\frac{dr}{d\phi}\frac{d\mu}{ds}ds \\
&= \int\frac{dr}{d\phi}d\mu \\
&= -\int\int\frac{d^2r}{d\phi^2}d\mu d\phi \\
&= \int\int r\,d\mu d\phi \\
&= \int\int\sqrt{\det\left[\mathrm{I}\right]}d\mu d\phi \\
&= A.
\end{aligned}$$

Finally we have to evaluate

$$\int\frac{d\theta}{ds}ds$$

For a simple planar curve we know that it is 2π when the curve runs counterclockwise. In fact this is also true in this context. To see this we notice, as in the planar case, that it must be a multiple of 2π. We can continuously deform the the sphere to become a cylinder of height π through surfaces of revolution

$$\mathbf{q}_\epsilon(\mu,\phi)=\begin{bmatrix}(1-\epsilon+\epsilon r)\cos\mu\\(1-\epsilon+\epsilon r)\sin\mu\\(1-\epsilon)\phi+\epsilon z\end{bmatrix}=\begin{bmatrix}r_\epsilon\cos\mu\\r_\epsilon\sin\mu\\z_\epsilon\end{bmatrix}.$$

The first fundamental forms will be

$$[\mathrm{I}_\epsilon]=\begin{bmatrix}r_\epsilon^2 & 0\\0 & \left(\frac{dr_\epsilon}{d\phi}\right)^2+\left(\frac{dz_\epsilon}{d\phi}\right)^2\end{bmatrix},$$

which at $\epsilon = 1$ gives us the sphere and at $\epsilon = 0$ the Euclidean metric on R. The angle in these metrics is calculated via

$$\begin{aligned}
\cos\theta_\epsilon &= \frac{\mathrm{I}_\epsilon\left(\frac{d\mathbf{q}}{dt}, \begin{bmatrix} 1 \\ 0 \end{bmatrix}\right)}{\left|\frac{d\mathbf{q}}{dt}\right|_\epsilon \left|\begin{bmatrix} 1 \\ 0 \end{bmatrix}\right|_\epsilon} \\
&= \frac{r_\epsilon \frac{d\mu}{dt}}{\sqrt{r_\epsilon^2\left(\frac{d\mu}{dt}\right)^2 + \left(\left(\frac{dr_\epsilon}{d\phi}\right)^2 + \left(\frac{dz_\epsilon}{d\phi}\right)^2\right)\left(\frac{d\mu}{d\phi}\right)^2}}.
\end{aligned}$$

This clearly varies continuously with ϵ so

$$\int \frac{d\theta_\epsilon}{dt} dt$$

will also vary continuously as claimed. However as it is always a multiple of 2π it must stay constant. Finally, for the Euclidean metric it is 2π so this is also the value in general. □

Exercises.

(1) A curve $\mathbf{q}(t)$ on a surface is called an *asymptotic curve* if $\mathrm{II}(\dot{\mathbf{q}}, \dot{\mathbf{q}}) = 0$, i.e., κ_n vanishes. Show that the binormal to the curve is normal to the surface.

(2) Let $\alpha(s)$ be a unit speed curve with non-vanishing curvature. Show that α is an asymptotic curve on the ruled surface

$$\mathbf{q}(s,t) = \alpha(s) + t\mathbf{N}_\alpha(s),$$

where $\mathbf{N}_\alpha$ is the normal to α as a space curve.

(3) Let $\mathbf{q}(s)$ be a unit speed curve on a surface with normal $\mathbf{N}$. Show that $\kappa_g = 0$ if and only if

$$\det[\dot{\mathbf{q}}, \ddot{\mathbf{q}}, \mathbf{N}] = 0.$$

(4) Show that latitudes on a sphere have constant κ_g.

(5) Let $\mathbf{q}(s)$ be a unit speed curve on a surface with normal $\mathbf{N}$. Show that

$$\frac{d}{ds}\begin{bmatrix} \mathbf{T} & \mathbf{S} & \mathbf{N} \end{bmatrix} = \begin{bmatrix} \mathbf{T} & \mathbf{S} & \mathbf{N} \end{bmatrix}\begin{bmatrix} 0 & -\kappa_g & -\kappa_n \\ \kappa_g & 0 & -\tau_g \\ \kappa_n & \tau_g & 0 \end{bmatrix}.$$

(6) Let $\mathbf{q}(s)$ be a unit speed curve on a surface with normal $\mathbf{N}$. Show that the space curvature κ is related to the geodesic and normal curvatures as follows

$$\kappa^2 = \kappa_g^2 + \kappa_n^2$$

and that the torsion is given by

$$\tau = \tau_g + \frac{\kappa_g\dot{\kappa}_n - \kappa_n\dot{\kappa}_g}{\kappa_g^2 + \kappa_n^2}$$

Hint: Start by showing that

$$\ddot{\mathbf{q}} = \kappa_n\mathbf{N} + \kappa_g\mathbf{S}.$$

(7) For a curve on the unit sphere show that

(a) $\tau_g = 0$.

(b) $\kappa_g = 0$ if and only if it is a great circle.
(c) κ_g is constant if and only if it is a circle.

(8) Let $\mathbf{q}(t)$ be a regular curve on a surface, with $\mathbf{N}$ being the normal to the surface. Show that

$$\kappa_n = \frac{\mathrm{II}(\dot{\mathbf{q}}, \dot{\mathbf{q}})}{\mathrm{I}(\dot{\mathbf{q}}, \dot{\mathbf{q}})}, \ \kappa_g = \frac{\det(\dot{\mathbf{q}}, \ddot{\mathbf{q}}, \mathbf{N})}{(\mathrm{I}(\dot{\mathbf{q}}, \dot{\mathbf{q}}))^{3/2}}.$$

(9) Let $\mathbf{q}(u,v)$ be a parametrization such that $g_{uu} = 1$ and $g_{uv} = 0$. Prove that the u-curves are unit speed with acceleration that is perpendicular to the surface. The u-curves are given by $\mathbf{q}(u) = \mathbf{q}(u,v)$ where v is fixed.

(10) Consider a surface of revolution

$$\mathbf{q}(s,\theta) = (r(s)\cos(\theta), r(s)\sin(\theta), z(s)),$$

where $(r(s), 0, z(s))$ is unit speed.
(a) Compute the second fundamental form.
(b) Compute $\kappa_g,\ \kappa_n,\ \tau_g$ for the meridians $\mathbf{q}(s) = \mathbf{q}(s,\theta)$. Conclude that their acceleration is perpendicular to the surface
(c) Compute $\kappa_g,\ \kappa_n,\ \tau_g$ for the latitudes $\mathbf{q}(\theta) = \mathbf{q}(s,\theta)$.

(11) Let M be a surface with normal $\mathbf{N}$ and $X, Y \in T_pM$. Show that if $\mathbf{q}(t)$ and a curve with velocity X at $t = 0$ and $Y(t)$ is an extension of the vector Y to a vector field along $\mathbf{q}$, then

$$\mathrm{II}(X,Y) = \mathbf{N} \cdot \frac{dY}{dt}(0).$$

(12) Let M be a surface given by an equation $F(x,y,z) = R$.
(a) If $\mathbf{q}(t)$ is a curve on M show that

$$\begin{aligned}
\ddot{\mathbf{q}} \cdot \nabla F &= -\dot{\mathbf{q}}^t \begin{bmatrix} \frac{\partial^2 F}{\partial x^2} & \frac{\partial^2 F}{\partial y \partial x} & \frac{\partial^2 F}{\partial z \partial x} \\ \frac{\partial^2 F}{\partial x \partial y} & \frac{\partial^2 F}{\partial y^2} & \frac{\partial^2 F}{\partial z \partial y} \\ \frac{\partial^2 F}{\partial x \partial z} & \frac{\partial^2 F}{\partial y \partial z} & \frac{\partial^2 F}{\partial z^2} \end{bmatrix} \dot{\mathbf{q}} \\
&= -\dot{\mathbf{q}}^t \left[\frac{\partial \nabla F}{\partial(x,y,z)} \right] \dot{\mathbf{q}}
\end{aligned}$$

(b) Show that

$$\mathrm{II}(X,Y) = -\frac{X^t \left[\frac{\partial \nabla F}{\partial(x,y,z)} \right] Y}{|\nabla F|}$$

5.2. The Gauss and Weingarten Maps and Equations

In this section we will complete the collection of Gauss equations. What was missing from those equations were the partial derivates of the normal. These extra equations are also known as the *Weingarten equations*. To better understand them we introduce the *Weingarten map*. As we shall see, the matrix of this map is related to the second fundamental form in the same way the Christoffel symbols of the second kind are related to the symbols of the first kind.

We start by defining things abstractly and then present the matrix versions afterwards.

DEFINITION 5.2.1. The *Gauss map* for a surface M with normal $\mathbf{N}$ is the map $\mathbf{N} : M \to S^2(1)$ that takes each point to the chosen normal at that point. The *Weingarten map* at a point $p \in M$ is the linear map $L : T_pM \to T_pM$ defined as the negative of the differential of $\mathbf{N}$:

$$L = -D\mathbf{N}.$$

REMARK 5.2.2. The definition of the Weingarten map requires some explanation as the differential should be a linear map

$$D\mathbf{N} : T_pM \to T_{\mathbf{N}(p)}S^2(1)$$

However, the normal vector to any point $\mathbf{x} \in S^2(1)$ is simply $\mathbf{N} = \pm\mathbf{x}$. As the tangent space is the orthogonal complement to the normal vector it follows that

$$T_pM = T_{\mathbf{N}(p)}S^2(1).$$

For a parametrized surface this tells us.

PROPOSITION 5.2.3. *(The Weingarten Equations) For a parametrized surface* $\mathbf{q}(u, v)$ *we have*

$$\begin{aligned} -\frac{\partial \mathbf{N}}{\partial u} &= L\left(\frac{\partial \mathbf{q}}{\partial u}\right), \\ -\frac{\partial \mathbf{N}}{\partial v} &= L\left(\frac{\partial \mathbf{q}}{\partial v}\right). \end{aligned}$$

More generally for a curve $\mathbf{q}(t)$ *on the surface*

$$-\frac{d\mathbf{N}\circ\mathbf{q}}{dt} = L\left(\frac{d\mathbf{q}}{dt}\right).$$

PROOF. The equations simply follow from the chain rule and the first two are special cases of the last. If we write the curve $\mathbf{q}(t) = \mathbf{q}(u(t), v(t))$, then

$$\begin{aligned} L\left(\frac{d\mathbf{q}}{dt}\right) &= -D\mathbf{N}\left(\frac{d\mathbf{q}}{dt}\right) \\ &= -\frac{d\mathbf{N}\circ\mathbf{q}}{dt} \\ &= -\left(\frac{\partial \mathbf{N}}{\partial u}\frac{du}{dt} + \frac{\partial \mathbf{N}}{\partial v}\frac{dv}{dt}\right). \end{aligned}$$

This proves the claim □

Next we show that the Weingarten map L is a self-adjoint map with respect to the first fundamental form.

PROPOSITION 5.2.4. *The Weingarten map is abstractly related to the second fundamental through the first fundamental form by the formula:*

$$\mathrm{I}(L(X), Y) = \mathrm{II}(X, Y) = \mathrm{I}(X, L(Y))$$

In particular, L *is self-adjoint as* II *is symmetric.*

PROOF. Since the second fundamental form is symmetric $\mathrm{II}(X,Y) = \mathrm{II}(Y,X)$ it follows that we only need to show that $\mathrm{I}(L(X),Y) = \mathrm{II}(X,Y)$ as we then obtain

$$\begin{aligned} \mathrm{I}(X, L(Y)) &= \mathrm{I}(L(Y), X) \\ &= \mathrm{II}(Y, X) \\ &= \mathrm{II}(X, Y). \end{aligned}$$

Next we observe that it suffices to prove that

$$\mathrm{II}(X,X) = \mathrm{I}(L(X), X)$$

as polarization implies

$$\begin{aligned} \mathrm{II}(X,Y) &= \frac{1}{2}\left(\mathrm{II}(X+Y, X+Y) - \mathrm{II}(X,X) - \mathrm{II}(Y,Y)\right) \\ &= \frac{1}{2}\left(\mathrm{I}(L(X+Y), X+Y) - \mathrm{I}(L(X), X) - \mathrm{I}(L(Y), Y)\right) \\ &= \mathrm{I}(L(X), Y). \end{aligned}$$

We let $X = \frac{d\mathbf{q}}{dt}$ and recall that

$$\mathrm{II}(X,X) = \frac{d^2\mathbf{q}}{dt^2} \cdot \mathbf{N}.$$

As $\frac{d\mathbf{q}}{dt}$ and $\mathbf{N}$ are perpendicular it follows that

$$\begin{aligned} 0 &= \frac{d}{dt}\left(\frac{d\mathbf{q}}{dt} \cdot \mathbf{N}\right) \\ &= \frac{d^2\mathbf{q}}{dt^2} \cdot \mathbf{N} + \frac{d\mathbf{q}}{dt} \cdot \frac{d\mathbf{N} \circ \mathbf{q}}{dt} \\ &= \mathrm{II}(X,X) - X \cdot L(X) \\ &= \mathrm{II}(X,X) - \mathrm{I}(L(X), X). \end{aligned}$$

This proves the claim. $\square$

All in all this is still a bit abstract, but the Weingarten equations and the relationship between the Weingarten map and the first and second fundamental forms allow us to give explicit formulas for a parametrized surface.

Given a parametrized surface $\mathbf{q}(u,v)$ the entries in the matrix representation of the Weingarten map are defined as

$$\begin{aligned} \begin{bmatrix} L\left(\frac{\partial \mathbf{q}}{\partial u}\right) & L\left(\frac{\partial \mathbf{q}}{\partial v}\right) \end{bmatrix} &= \begin{bmatrix} \frac{\partial \mathbf{q}}{\partial u} & \frac{\partial \mathbf{q}}{\partial v} \end{bmatrix} [L] \\ &= \begin{bmatrix} \frac{\partial \mathbf{q}}{\partial u} & \frac{\partial \mathbf{q}}{\partial v} \end{bmatrix} \begin{bmatrix} L_u^u & L_v^u \\ L_u^v & L_v^v \end{bmatrix}. \end{aligned}$$

The matrix representation can be calculated as follows.

PROPOSITION 5.2.5. *The matrix representations of the Weingarten maps and the second fundamental form satisfy:*

$$[L] = [\mathrm{I}]^{-1}[\mathrm{II}]$$

and

$$
\begin{aligned}
[\mathrm{II}] &= -\left[\begin{array}{cc} \frac{\partial \mathbf{N}}{\partial u} & \frac{\partial \mathbf{N}}{\partial v} \end{array}\right]^t \left[\begin{array}{cc} \frac{\partial \mathbf{q}}{\partial u} & \frac{\partial \mathbf{q}}{\partial v} \end{array}\right] \\
&= -\left[\begin{array}{cc} \frac{\partial \mathbf{N}}{\partial u}\cdot\frac{\partial \mathbf{q}}{\partial u} & \frac{\partial \mathbf{N}}{\partial u}\cdot\frac{\partial \mathbf{q}}{\partial v} \\ \frac{\partial \mathbf{N}}{\partial v}\cdot\frac{\partial \mathbf{q}}{\partial u} & \frac{\partial \mathbf{N}}{\partial v}\cdot\frac{\partial \mathbf{q}}{\partial v} \end{array}\right] \\
&= -\left[\begin{array}{cc} \frac{\partial \mathbf{q}}{\partial u} & \frac{\partial \mathbf{q}}{\partial v} \end{array}\right]^t \left[\begin{array}{cc} \frac{\partial \mathbf{N}}{\partial u} & \frac{\partial \mathbf{N}}{\partial v} \end{array}\right].
\end{aligned}
$$

PROOF. To show the formula for [II] use that $\mathbf{N}$ is perpendicular to $\frac{\partial \mathbf{q}}{\partial w_2}$ and note that

$$
\begin{aligned}
L_{w_1 w_2} &= \frac{\partial^2 \mathbf{q}}{\partial w_1 \partial w_2}\cdot \mathbf{N} \\
&= \left(\frac{\partial}{\partial w_1}\left(\frac{\partial \mathbf{q}}{\partial w_2}\right)\right)\cdot \mathbf{N} \\
&= \frac{\partial}{\partial w_1}\left(\frac{\partial \mathbf{q}}{\partial w_2}\cdot \mathbf{N}\right) - \frac{\partial \mathbf{q}}{\partial w_2}\cdot\frac{\partial \mathbf{N}}{\partial w_1} \\
&= -\frac{\partial \mathbf{q}}{\partial w_2}\cdot\frac{\partial \mathbf{N}}{\partial w_1}.
\end{aligned}
$$

We then have that

$$
\begin{aligned}
[\mathrm{I}]^{-1}[\mathrm{II}] &= \left(\left[\begin{array}{cc} \frac{\partial \mathbf{q}}{\partial u} & \frac{\partial \mathbf{q}}{\partial v} \end{array}\right]^t \left[\begin{array}{cc} \frac{\partial \mathbf{q}}{\partial u} & \frac{\partial \mathbf{q}}{\partial v} \end{array}\right]\right)^{-1} \left[\begin{array}{cc} \frac{\partial \mathbf{q}}{\partial u} & \frac{\partial \mathbf{q}}{\partial v} \end{array}\right]^t \left[\begin{array}{cc} -\frac{\partial \mathbf{N}}{\partial u} & -\frac{\partial \mathbf{N}}{\partial v} \end{array}\right] \\
&= \left(\left[\begin{array}{cc} \frac{\partial \mathbf{q}}{\partial u} & \frac{\partial \mathbf{q}}{\partial v} \end{array}\right]^t \left[\begin{array}{cc} \frac{\partial \mathbf{q}}{\partial u} & \frac{\partial \mathbf{q}}{\partial v} \end{array}\right]\right)^{-1} \left[\begin{array}{cc} \frac{\partial \mathbf{q}}{\partial u} & \frac{\partial \mathbf{q}}{\partial v} \end{array}\right]^t \left[\begin{array}{cc} L\left(\frac{\partial \mathbf{q}}{\partial u}\right) & L\left(\frac{\partial \mathbf{q}}{\partial v}\right) \end{array}\right] \\
&= \left(\left[\begin{array}{cc} \frac{\partial \mathbf{q}}{\partial u} & \frac{\partial \mathbf{q}}{\partial v} \end{array}\right]^t \left[\begin{array}{cc} \frac{\partial \mathbf{q}}{\partial u} & \frac{\partial \mathbf{q}}{\partial v} \end{array}\right]\right)^{-1} \left[\begin{array}{cc} \frac{\partial \mathbf{q}}{\partial u} & \frac{\partial \mathbf{q}}{\partial v} \end{array}\right]^t \left[\begin{array}{cc} \frac{\partial \mathbf{q}}{\partial u} & \frac{\partial \mathbf{q}}{\partial v} \end{array}\right][L] \\
&= [L].
\end{aligned}
$$

□

REMARK 5.2.6. It is important to realize that while L is self-adjoint its matrix representation

$$[L] = [\mathrm{I}]^{-1}[\mathrm{II}]$$

need not be symmetric. In fact as [I] and [II] are symmetric it follows that

$$[L]^t = [\mathrm{II}][\mathrm{I}]^{-1}$$

so $[L]$ is only symmetric if [I] and [II] commute.

The Weingarten equations can now be written as

$$\frac{\partial \mathbf{N}}{\partial w} = -L_w^u \frac{\partial \mathbf{q}}{\partial u} - L_w^v \frac{\partial \mathbf{q}}{\partial v} = -L\left(\frac{\partial \mathbf{q}}{\partial w}\right).$$

Together the Gauss formulas and Weingarten equations tell us how the derivatives of our basis $\frac{\partial \mathbf{q}}{\partial u}, \frac{\partial \mathbf{q}}{\partial v}, \mathbf{N}$ relate back to the basis. They can be collected as follows:

COROLLARY 5.2.7. (The Gauss and Weingarten Formulas)

$$
\begin{aligned}
\frac{\partial}{\partial w}\left[\begin{array}{ccc} \frac{\partial \mathbf{q}}{\partial u} & \frac{\partial \mathbf{q}}{\partial v} & \mathbf{N} \end{array}\right] &= \left[\begin{array}{ccc} \frac{\partial \mathbf{q}}{\partial u} & \frac{\partial \mathbf{q}}{\partial v} & \mathbf{N} \end{array}\right][D_w] \\
&= \left[\begin{array}{ccc} \frac{\partial \mathbf{q}}{\partial u} & \frac{\partial \mathbf{q}}{\partial v} & \mathbf{N} \end{array}\right] \left[\begin{array}{ccc} \Gamma_{wu}^u & \Gamma_{wv}^u & -L_w^u \\ \Gamma_{wu}^v & \Gamma_{wv}^v & -L_w^v \\ L_{wu} & L_{wv} & 0 \end{array}\right].
\end{aligned}
$$

Exercises.

(1) For a surface of revolution
$$\mathbf{q}(t,\theta) = (r(t)\cos(\theta), r(t)\sin(\theta), z(t))$$
compute the first and second fundamental forms and the Weingarten map.

(2) Compute the matrix representation of the Weingarten map for a Monge patch $\mathbf{q}(x,y) = (x, y, f(x,y))$ with respect to the basis $\frac{\partial \mathbf{q}}{\partial x}, \frac{\partial \mathbf{q}}{\partial y}$.

(3) Show that if a surface satisfies $\mathrm{II} = \pm \frac{1}{R}\mathrm{I}$, then it is part of a sphere of radius R. Hint: Show that $\mathbf{N} \pm \frac{1}{R}\mathbf{q}$ is constant and use that to find the center of the sphere.

(4) Let M be a surface with normal $\mathbf{N}$ and $X, Y \in T_pM$. Show that if $\mathbf{q}(t)$ and a curve with velocity X at $t = 0$, then
$$\mathrm{II}(X,Y) = -Y \cdot \frac{d\mathbf{N} \circ \mathbf{q}}{dt}(0).$$

(5) Show that for a curve on a surface the geodesic torsion satisfies
$$\tau_g = \mathrm{II}(\mathbf{T}, \mathbf{S}).$$

5.3. The Gauss and Mean Curvatures

Definition 5.3.1. The *Gauss curvature* is defined as the determinant of the Weingarten map
$$K = \det L$$
and the *mean curvature* is related to the trace as follows
$$H = \frac{1}{2}\mathrm{tr}L$$

To calculate these quantities we have:

Proposition 5.3.2. *The Gauss and mean curvatures of a parametrized surface* $\mathbf{q}(u,v)$ *can be computed as*
$$K = \frac{\det[\mathrm{II}]}{\det[\mathrm{I}]} = \frac{L_{uu}L_{vv} - (L_{uv})^2}{g_{uu}g_{vv} - (g_{uv})^2}$$
and
$$H = \frac{1}{2}\frac{g_{vv}L_{uu} + g_{uu}L_{vv} - 2g_{uv}L_{uv}}{g_{uu}g_{vv} - (g_{uv})^2}.$$

Proof. To calculate the Gauss and mean curvatures we use the formulas for determinant and trace for a matrix representation:
$$K = \det[L] = L_u^u L_v^v - L_u^v L_v^u$$
and
$$H = \frac{1}{2}\mathrm{tr}[L] = \frac{1}{2}(L_u^u + L_v^v)$$
together with $[L] = [\mathrm{I}]^{-1}[\mathrm{II}]$ (see proposition 5.2.5). The formula for K now follows from standard determinant rules.

For H we just need to recall how the inverse of a matrix is computed
$$\begin{bmatrix} g_{uu} & g_{uv} \\ g_{vu} & g_{vv} \end{bmatrix}^{-1} = \frac{1}{g_{uu}g_{vv} - (g_{uv})^2}\begin{bmatrix} g_{vv} & -g_{uv} \\ -g_{vu} & g_{uu} \end{bmatrix}$$
to get the desired formula. □

The Gauss curvature can also be expressed more directly from the unit normal.

PROPOSITION 5.3.3. *(Gauss) The Gauss curvature satisfies*

$$K = \frac{\left(\frac{\partial \mathbf{N}}{\partial u} \times \frac{\partial \mathbf{N}}{\partial v}\right) \cdot \mathbf{N}}{\left|\frac{\partial \mathbf{q}}{\partial u} \times \frac{\partial \mathbf{q}}{\partial v}\right|}.$$

PROOF. Simply use the Weingarten equations to calculate

$$\begin{aligned}
\frac{\partial \mathbf{N}}{\partial u} \times \frac{\partial \mathbf{N}}{\partial v} &= \left(-L_u^u \frac{\partial \mathbf{q}}{\partial u} - L_u^v \frac{\partial \mathbf{q}}{\partial v}\right) \times \left(-L_v^u \frac{\partial \mathbf{q}}{\partial u} - L_v^v \frac{\partial \mathbf{q}}{\partial v}\right) \\
&= L_u^u L_v^v \frac{\partial \mathbf{q}}{\partial u} \times \frac{\partial \mathbf{q}}{\partial v} + L_u^v L_v^u \frac{\partial \mathbf{q}}{\partial v} \times \frac{\partial \mathbf{q}}{\partial u} \\
&= \left(L_u^u L_v^v - L_u^v L_v^u\right) \frac{\partial \mathbf{q}}{\partial u} \times \frac{\partial \mathbf{q}}{\partial v} \\
&= K \left|\frac{\partial \mathbf{q}}{\partial u} \times \frac{\partial \mathbf{q}}{\partial v}\right| \mathbf{N}.
\end{aligned}$$

□

Note that the denominator in

$$K = \frac{\left(\frac{\partial \mathbf{N}}{\partial u} \times \frac{\partial \mathbf{N}}{\partial v}\right) \cdot \mathbf{N}}{\left|\frac{\partial \mathbf{q}}{\partial u} \times \frac{\partial \mathbf{q}}{\partial v}\right|}$$

is already computed in terms of the first fundamental form

$$\left|\frac{\partial \mathbf{q}}{\partial u} \times \frac{\partial \mathbf{q}}{\partial v}\right|^2 = g_{uu} g_{vv} - \left(g_{uv}\right)^2$$

and can also be expressed as the volume of a parallelepiped

$$\left|\frac{\partial \mathbf{q}}{\partial u} \times \frac{\partial \mathbf{q}}{\partial v}\right| \cdot \mathbf{N}.$$

The numerator is the signed volume of the parallelepiped $\frac{\partial \mathbf{N}}{\partial u}, \frac{\partial \mathbf{N}}{\partial v}, \mathbf{N}$ corresponding to the Gauss map $\mathbf{N}(u, v) : U \to S^2(1) \subset \mathbb{R}^3$ of the surface. Thus it can be computed from the first fundamental form of $\mathbf{N}(u, v)$. However, there is a sign that depends on whether $\mathbf{N}$ points in the same direction as $\frac{\partial \mathbf{N}}{\partial u} \times \frac{\partial \mathbf{N}}{\partial v}$. Recall from curve theory that the tangent spherical image was also related to curvature in a similar way. Here the formulas are a bit more complicated as we use arbitrary parameters.

DEFINITION 5.3.4. The *third fundamental form* III is defined as the first fundamental form for $\mathbf{N}$

$$\begin{aligned}
[\mathrm{III}] &= \begin{bmatrix} \frac{\partial \mathbf{N}}{\partial u} & \frac{\partial \mathbf{N}}{\partial v} \end{bmatrix}^t \begin{bmatrix} \frac{\partial \mathbf{N}}{\partial u} & \frac{\partial \mathbf{N}}{\partial v} \end{bmatrix} \\
&= \begin{bmatrix} \frac{\partial \mathbf{N}}{\partial u} \cdot \frac{\partial \mathbf{N}}{\partial u} & \frac{\partial \mathbf{N}}{\partial u} \cdot \frac{\partial \mathbf{N}}{\partial v} \\ \frac{\partial \mathbf{N}}{\partial v} \cdot \frac{\partial \mathbf{N}}{\partial u} & \frac{\partial \mathbf{N}}{\partial v} \cdot \frac{\partial \mathbf{N}}{\partial v} \end{bmatrix}.
\end{aligned}$$

This certainly makes sense, but $\mathbf{N}$ might not be a genuine parametrization if the Gauss curvature vanishes. Nevertheless we alway have the relationship

$$\frac{\partial \mathbf{N}}{\partial u} \times \frac{\partial \mathbf{N}}{\partial v} = \left|\frac{\partial \mathbf{N}}{\partial u} \times \frac{\partial \mathbf{N}}{\partial v}\right| \mathbf{N}.$$

The three fundamental forms and two curvatures are related by a very interesting formula which also shows that the third fundamental form is almost redundant.

THEOREM 5.3.5. *All three fundamental forms are related by*

$$\mathrm{III} - 2H\mathrm{II} + K\mathrm{I} = 0.$$

PROOF. We first reduce this statement to the Cayley-Hamilton theorem for the linear operator L. This relies on showing

$$\begin{aligned} \mathrm{I}\left(L\left(X\right),Y\right) &= \mathrm{II}\left(X,Y\right), \\ \mathrm{I}\left(L^{2}\left(X\right),Y\right) &= \mathrm{III}\left(X,Y\right), \end{aligned}$$

and then proving that any 2×2 matrix satisfies:

$$L^{2}-\left(\mathrm{tr}\left(L\right)\right)L+\det\left(L\right)I=0,$$

where I is the identity matrix. This last step can be done by a straightforward calculation for a 2×2-matrix. (Only in higher dimensions is a more advanced proof necessary.)

We already proved that $\mathrm{I}\left(L\left(X\right),Y\right)=\mathrm{II}\left(X,Y\right)$, which in matrix form is equivalent to saying $[\mathrm{I}]\,[L]=[\mathrm{II}]$. We similarly have from the Weingarten equations that

$$\begin{aligned} [\mathrm{III}] &= \begin{bmatrix} L\left(\frac{\partial\mathbf{q}}{\partial u}\right) & L\left(\frac{\partial\mathbf{q}}{\partial v}\right)\end{bmatrix}^{t}\begin{bmatrix} L\left(\frac{\partial\mathbf{q}}{\partial u}\right) & L\left(\frac{\partial\mathbf{q}}{\partial v}\right)\end{bmatrix} \\ &= [L]^{t}\begin{bmatrix}\frac{\partial\mathbf{q}}{\partial u} & \frac{\partial\mathbf{q}}{\partial v}\end{bmatrix}^{t}\begin{bmatrix}\frac{\partial\mathbf{q}}{\partial u} & \frac{\partial\mathbf{q}}{\partial v}\end{bmatrix}[L] \\ &= [L]^{t}\,[\mathrm{I}]\,[L] \\ &= [\mathrm{II}]^{t}\left([\mathrm{I}]^{-1}\right)^{t}[\mathrm{I}]\,[L] \\ &= [\mathrm{II}]\,[\mathrm{I}]^{-1}\,[\mathrm{I}]\,[L] \\ &= [\mathrm{II}]\,[L] \\ &= [\mathrm{I}]\,[L]\,[L] \\ &= [\mathrm{I}]\,[L]^{2} \end{aligned}$$

showing that $\mathrm{I}\left(L^{2}\left(X\right),Y\right)=\mathrm{III}\left(X,Y\right).$ □

DEFINITION 5.3.6. A surface is called *minimal* if its mean curvature vanishes.

PROPOSITION 5.3.7. *A minimal surface has conformal Gauss map.*

PROOF. Let $\mathbf{q}\left(u,v\right)$ be a parametrization of the surface, then $\mathbf{N}\left(u,v\right)$ is a potential parametrization of the unit sphere via the Gauss map. The first fundamental form with respect to this parametrization is the third fundamental form. Using $H=0$ we obtain

$$[\mathrm{III}]+K\,[\mathrm{I}]=0,$$

which implies that the Gauss map is conformal. □

EXAMPLE 5.3.8. Note that the Gauss map for the unit sphere centered at the origin is simply the identity map on the sphere. Thus its Gauss map is an isometry and in particular conformal. However, the sphere is not a minimal surface. More generally, the Gauss map

$$\mathbf{N}\left(\mathbf{q}\right)=\pm\frac{\mathbf{q}-\mathbf{c}}{R}$$

for a sphere of radius R centered at $\mathbf{c}$ is also conformal as its derivative is given by $D\mathbf{N}=\pm\frac{1}{R}I$, where I is the identity map/matrix.

The name for minimal surfaces is partly justified by the next result. Meusnier in 1785 was the first to consider such surfaces and he also indicated with a geometric argument that their areas should be minimal. In fact Lagrange had already in 1761 come up with an (Euler-Lagrange) equation for surfaces that minimize area, but it was not until the mid 19th century with Bonnet and Beltrami that this was definitively connected to the condition that the mean curvature should vanish.

THEOREM 5.3.9. *A surface whose area is minimal among nearby surfaces is a minimal surface.*

PROOF. We assume that the surface is given by a parametrization $\mathbf{q}(u,v)$ and only consider nearby surfaces that are graphs over the given surface, i.e.,

$$\mathbf{q}^* = \mathbf{q} + \phi\mathbf{N}$$

for some function $\phi(u,v)$. From such a surface we can then create a family of surfaces

$$\mathbf{q}_\epsilon = \mathbf{q} + \epsilon\phi\mathbf{N}$$

that interpolates between these two surfaces. To calculate the area density as a function of ϵ we first note that

$$\frac{\partial \mathbf{q}_\epsilon}{\partial w} = \frac{\partial \mathbf{q}}{\partial w} + \epsilon\left(\frac{\partial \phi}{\partial w}\mathbf{N} + \phi\frac{\partial \mathbf{N}}{\partial w}\right).$$

Then the first fundamental form becomes

$$\begin{aligned}
g^\epsilon_{ww} &= g_{ww} + 2\epsilon\phi\frac{\partial \mathbf{q}}{\partial w}\cdot\frac{\partial \mathbf{N}}{\partial w} + \epsilon^2\left(\left(\frac{\partial \phi}{\partial w}\right)^2 + \phi^2\left|\frac{\partial \mathbf{N}}{\partial w}\right|^2\right) \\
&= g_{ww} - 2\epsilon\phi L_{ww} + \epsilon^2\left(\left(\frac{\partial \phi}{\partial w}\right)^2 - K\phi^2 g_{ww} + 2H\phi^2 L_{ww}\right), \\
g^\epsilon_{uv} &= g_{uv} + \epsilon\phi\left(\frac{\partial \mathbf{q}}{\partial u}\cdot\frac{\partial \mathbf{N}}{\partial v} + \frac{\partial \mathbf{q}}{\partial v}\cdot\frac{\partial \mathbf{N}}{\partial u}\right) + \epsilon^2\left(\frac{\partial \phi}{\partial u}\cdot\frac{\partial \phi}{\partial v} + \phi^2\frac{\partial \mathbf{N}}{\partial u}\cdot\frac{\partial \mathbf{N}}{\partial v}\right) \\
&= g_{uv} - 2\epsilon\phi L_{uv} + \epsilon^2\left(\frac{\partial \phi}{\partial u}\cdot\frac{\partial \phi}{\partial v} - K\phi^2 g_{uv} + 2H\phi^2 L_{uv}\right),
\end{aligned}$$

and the square of the area density

$$\begin{aligned}
g^\epsilon_{uu}g^\epsilon_{vv} - \left(g^\epsilon_{uv}\right)^2 &= g_{uu}g_{vv} - \left(g_{uv}\right)^2 - 2\epsilon\left(g_{uu}L_{vv} + g_{vv}L_{uu} - 2g_{uv}L_{uv}\right) + O\left(\epsilon^2\right) \\
&= \left(g_{uu}g_{vv} - \left(g_{uv}\right)^2\right)(1 - \epsilon\phi H) + O\left(\epsilon^2\right).
\end{aligned}$$

This shows that if $H \neq 0$ somewhere then we can select ϕ such that the area density will decrease for nearby surfaces. □

REMARK 5.3.10. Conversely note that when $H = 0$ everywhere, then the area density is critical. The term that involves ϵ^2 has a coefficient that looks like

$$\begin{aligned}
& g_{uu}\left(\left(\frac{\partial\phi}{\partial v}\right)^2 - \phi^2 K g_{vv}\right) + g_{vv}\left(\left(\frac{\partial\phi}{\partial u}\right)^2 - \phi^2 K g_{uu}\right) \\
& -2g_{uv}\left(\frac{\partial\phi}{\partial u}\cdot\frac{\partial\phi}{\partial v} - \phi^2 K g_{uv}\right) + 4\phi^2\left(L_{uu}L_{vv} - L_{uv}^2\right) \\
= & \left|-\frac{\partial\phi}{\partial v}\frac{\partial\mathbf{q}}{\partial u} + \frac{\partial\mathbf{q}}{\partial v}\frac{\partial\phi}{\partial u}\right|^2 - 2\phi^2 K\left(g_{uu}g_{vv} - g_{uv}^2\right) + 4\phi^2\left(L_{uu}L_{vv} - L_{uv}^2\right) \\
= & \left|-\frac{\partial\phi}{\partial v}\frac{\partial\mathbf{q}}{\partial u} + \frac{\partial\mathbf{q}}{\partial v}\frac{\partial\phi}{\partial u}\right|^2 + 2\phi^2 K\left(g_{uu}g_{vv} - g_{uv}^2\right)
\end{aligned}$$

and it is not clear that this is positive. In fact minimal surfaces have $K \le 0$ so the two terms compete.

Exercises.

(1) For a surface of revolution

$$\mathbf{q}(t,\mu) = (r(t)\cos\mu, r(t)\sin\mu, z(t))$$

compute the first and second fundamental forms as well as the Gauss and mean curvatures.

(2) Show that a surface is planar if and only if its Gauss and mean curvatures vanish. Hint: In a parametrization the Gauss and mean curvatures vanish if

$$\begin{aligned}
L_{uu}L_{vv} - (L_{uv})^2 &= 0, \\
g_{vv}L_{uu} + g_{uu}L_{vv} - 2g_{uv}L_{uv} &= 0.
\end{aligned}$$

Combine this with

$$g_{uu}g_{vv} - (g_{uv})^2 > 0$$

to reach a contradiction if the second fundamental form doesn't vanish.

(3) Compute the second fundamental form of a tangent developable $\mathbf{q}(s,t) = \alpha(t) + s\frac{d\alpha}{dt}$ of a unit speed curve $\alpha(t)$. Show that the mean curvature vanishes if and only if α is planar, and in that case the second fundamental form vanishes.

(4) Let $\mathbf{q}(t)$ be a curve on a surface with normal $\mathbf{N}$. Denote the Gauss image of the curve by $\mathbf{N}(t) = \mathbf{N}\circ\mathbf{q}(t)$. Show that the velocities of these curves are related by

$$\left|\frac{d\mathbf{N}}{dt}\right|^2 - 2H\frac{d\mathbf{N}}{dt}\cdot\frac{d\mathbf{q}}{dt} + K\left|\frac{d\mathbf{q}}{dt}\right|^2 = 0.$$

(5) Let $\mathbf{q}(t) = \mathbf{q}(u(t), v(t))$ be an asymptotic curve on a surface, i.e., $\kappa_n = 0$.
 (a) Show that $K \le 0$ along the curve.
 (b) (Beltrami-Enneper) If τ is the torsion of the curve as a space curve, then

$$\tau^2 = -K.$$

 Hint: Use the previous exercise.

(6) Let $\mathbf{q}(u,v)$ be a parametrized surface with negative Gauss curvature.

(a) Show that it can locally be reparametrized $\mathbf{q}(s,t)$ so that the parameter curves are asymptotic curves, i.e., the second fundamental form looks like

$$[\mathrm{II}] = \begin{bmatrix} 0 & L_{st} \\ L_{st} & 0 \end{bmatrix}.$$

(b) Show that in this case

$$[\mathrm{III}] = -K \begin{bmatrix} g_{ss} & -g_{st} \\ -g_{st} & g_{tt} \end{bmatrix}.$$

(7) (Meusnier, 1785) Show that the catenoid $\mathbf{q}(t,\mu) = (\cosh t \cos\mu, \cosh t \sin\mu, t)$ is minimal. Conversely, show that if a surface of revolution parametrized as

$$\mathbf{q}(t,\mu) = (r(t)\cos\mu, r(t)\sin\mu, t)$$

is minimal, then $r = a\cosh t + b\sinh t$ for constants a, b. Conclude that $b = 0$ if $r > 0$ for all $t \in \mathbb{R}$.

(8) (Meusnier, 1785) Show that the helicoid

$$\mathbf{q}(\phi,t) = (\sinh\phi\cos t, \sinh\phi\sin t, t)$$

is minimal.

(9) Show that a minimal surface satisfies $K \leq 0$.

(10) Compute the Gauss curvatures of the generalized cones (section 4.1 exercise 2), cylinders (section 4.1 exercise 1), and tangent developables (section 4.1 exercise 4). We shall show below that these are essentially the only surfaces with vanishing Gauss curvature.

(11) Show that

$$\begin{aligned} \frac{\partial \mathbf{N}}{\partial u} \times \frac{\partial \mathbf{N}}{\partial v} &= K \frac{\partial \mathbf{q}}{\partial u} \times \frac{\partial \mathbf{q}}{\partial v}, \\ \frac{\partial \mathbf{q}}{\partial u} \times \frac{\partial \mathbf{N}}{\partial v} + \frac{\partial \mathbf{N}}{\partial u} \times \frac{\partial \mathbf{q}}{\partial v} &= -2H \frac{\partial \mathbf{q}}{\partial u} \times \frac{\partial \mathbf{q}}{\partial v}, \end{aligned}$$

and more generally that

$$\begin{aligned} \frac{\partial \mathbf{q}}{\partial u} \times \frac{\partial \mathbf{N}}{\partial w} &= -L_w^v \frac{\partial \mathbf{q}}{\partial u} \times \frac{\partial \mathbf{q}}{\partial v}, \\ \frac{\partial \mathbf{N}}{\partial w} \times \frac{\partial \mathbf{q}}{\partial v} &= -L_w^u \frac{\partial \mathbf{q}}{\partial u} \times \frac{\partial \mathbf{q}}{\partial v}. \end{aligned}$$

(12) Compute the first and second fundamental forms as well as the Gauss and mean curvatures for the conoid

$$\begin{aligned} \mathbf{q}(s,t) &= (sx(t), sy(t), z(t)) \\ &= (0,0,z(t)) + s(x(t), y(t), 0) \end{aligned}$$

when $X = (x(t), y(t), 0)$ is a unit field.

(13) Let $X, Y \in T_pM$ be an orthonormal basis for the tangent space at p to the surface M. Prove that the mean and Gauss curvatures can be computed as follows:

$$\begin{aligned} H &= \frac{1}{2}(\mathrm{II}(X,X) + \mathrm{II}(Y,Y)), \\ K &= \mathrm{II}(X,X)\,\mathrm{II}(Y,Y) - (\mathrm{II}(X,Y))^2. \end{aligned}$$

(14) Show that Enneper's surface

$$\mathbf{q}(u,v) = \left(u - \frac{1}{3}u^3 + uv^2, v - \frac{1}{3}v^3 + vu^2, u^2 - v^2\right)$$

is minimal.

(15) Show that Scherk's surface $e^z \cos x = \cos y$ is minimal.

(16) Consider a unit speed curve $\alpha(s) : [0, L] \to \mathbb{R}^3$ with non-vanishing curvature and the tube of radius R around it

$$\mathbf{q}(s,\phi) = \alpha(s) + R(\mathbf{N}_\alpha \cos\phi + \mathbf{B}_\alpha \sin\phi)$$

(see section 4.1 exercise 8 and section 4.4 exercise 15).

(a) Use Gauss' formula for K to show that

$$K = \frac{-\kappa\cos\phi}{R(1-\kappa R)}.$$

(b) Show that

$$\int_0^{2\pi}\int_0^L K\sqrt{\det[\mathrm{I}]}\,ds\,d\phi = 0$$

and

$$\int_0^{2\pi}\int_0^L |K|\sqrt{\det[\mathrm{I}]}\,ds\,d\phi = 4\int_a^b \kappa\,ds.$$

(17) (Monge 1775) Consider a Monge patch $z = F(x,y)$. Define the two functions $p = \frac{\partial F}{\partial x}$ and $q = \frac{\partial F}{\partial y}$.

(a) Show that the Gauss curvature vanishes if and only if

$$\frac{\partial^2 F}{\partial x^2}\frac{\partial^2 F}{\partial y^2} - \left(\frac{\partial^2 F}{\partial x\partial y}\right)^2 = 0.$$

(b) Assume that $\frac{\partial^2 F}{\partial x\partial y} = 0$ on an open set.

(i) Show that $F = f(x) + h(y)$.

(ii) Show that the Gauss curvature vanishes if and only if $f'' = 0$ or $h'' = 0$.

(iii) Show that if the Gauss curvature vanishes, then it is a ruled surface.

(c) Assume that $\frac{\partial^2 F}{\partial x\partial y} \neq 0$ and that the Gauss curvature vanishes.

(i) Show that we can locally reparametrize the surface using the reparametrization $(u,q) = (x, q(x,y))$.

(ii) Show that $p = f(q)$ for some function f. Hint: In the (u,q)-coordinates $\frac{\partial p}{\partial u} = 0$. When doing this calculation keep in mind that y depends on u in(u,q)-coordinates as q depends on both x and y.

(iii) Show in the same way that $F(x,y) - (xp + qy) = h(q)$.

(iv) Show that in the new parametrization:

$$y = -h'(q) - uf'(q)$$

and

$$\begin{aligned} z &= xp + qy + h(q) \\ &= u(f(q) - qf'(q)) + h(q) - qh'(q). \end{aligned}$$

(v) Show that this is a ruled surface.
(vi) Show that this ruled surface is a generalized cylinder when f'' vanishes.
(vii) Show that it is a generalized cone when $h'' = af''$ for some constant a.
(viii) Show that otherwise it is a tangent developable by showing that the lines in the ruling are all tangent to the curve that corresponds to

$$u = -\frac{h''}{f''}.$$

5.4. Principal Curvatures

Definition 5.4.1. The *principal curvatures* at a point q on a surface are the eigenvalues of the Weingarten map $L : T_qM \to T_qM$ associated to that point. The *principal directions* are the corresponding eigenvectors. We say that q is umbilic if the principal curvatures coincide.

Definition 5.4.2. A curve on a surface with the property that its velocity is always an eigenvector for the Weingarten map, i.e., a principal direction, is called a *line of curvature.*

The fact that L is self-adjoint with respect to the first fundamental form guarantees that we can always find an orthonormal basis of principal directions and that the principal curvatures are real. This is a nice and general theorem from linear algebra, variously called diagonalization of symmetric matrices or the spectral theorem. Since the Weingarten map is a linear map on a two-dimensional vector space we can give a direct proof that does not use to more general constructions needed in higher dimensions.

Theorem 5.4.3. *For a fixed point $q \in M$, there exist orthonormal principal directions E_1, $E_2 \in T_qM$*

$$\begin{aligned} L(E_1) &= \kappa_1 E_1, \\ L(E_2) &= \kappa_2 E_2. \end{aligned}$$

Moreover, κ_1, κ_2 are both real.

Proof. The characteristic polynomial for L looks like

$$\lambda^2 - 2H\lambda + K = 0.$$

The roots of this polynomial are real if and only if the discriminant is non-negative:

$$\begin{aligned} 4H^2 - 4K &\geq 0, \text{ or} \\ H^2 &\geq K. \end{aligned}$$

If we select an orthonormal basis for T_qM (it doesn't have to be related to a parametrization), then the matrix representation for L is symmetric

$$[L] = \begin{bmatrix} a & b \\ b & d \end{bmatrix}.$$

Thus

$$\begin{aligned} H &= \frac{a+d}{2}, \\ K &= ad - b^2. \end{aligned}$$

This means we need to show that

$$ad - b^2 \leq \left(\frac{a+d}{2}\right)^2.$$

This follows directly from the trivial inequality:

$$-b^2 \leq \frac{a^2+d^2}{4}.$$

If the principal curvatures are equal, then all vectors are eigenvectors and so we can certainly find an orthonormal basis that diagonalizes L. If the principal curvatures are not equal, then the corresponding principal directions are forced to be orthogonal:

$$\kappa_1 \mathrm{I}(E_1, E_2) = \mathrm{I}(L(E_1), E_2) = \mathrm{I}(E_1, L(E_2)) = \kappa_2 \mathrm{I}(E_1, E_2)$$

or

$$(\kappa_1 - \kappa_2)\, \mathrm{I}(E_1, E_2) = 0.$$

□

Remark 5.4.4. The height function that measures the distance from a point on the surface to the tangent space T_qM is given by

$$f(\mathbf{q}) = (\mathbf{q} - q) \cdot \mathbf{N}(q).$$

Its partial derivatives with respect to a parametrization of the surface are

$$\begin{aligned} \frac{\partial f}{\partial w} &= \frac{\partial \mathbf{q}}{\partial w} \cdot \mathbf{N}(q), \\ \frac{\partial^2 f}{\partial w_1 \partial w_2} &= \frac{\partial^2 \mathbf{q}}{\partial w_1 \partial w_2} \cdot \mathbf{N}(q). \end{aligned}$$

So f has a critical point at q, and the second derivative matrix at q is simply $[\mathrm{II}]$. The second derivative test then tells us something about how the surface is placed in relation to T_qM. Specifically we see that if both principal curvatures have the same sign, or $K > 0$, then the surface must locally be on one side of the tangent plane, while if the principal curvatures have opposite signs, or $K < 0$, then the surface lies on both sides. In that case it'll look like a saddle.

We can also relate the second fundamental form in general directions to the principal curvatures.

Theorem 5.4.5. *(Euler, 1760) Let $X \in T_qM$ be a unit vector and κ_1, κ_2 the principal curvatures, then*

$$\mathrm{II}(X, X) = \left(\kappa_1 \cos^2\phi + \kappa_2 \sin^2\phi\right),$$

where ϕ is the angle between X and the principal direction corresponding to κ_1.

Proof. Simply select an orthonormal basis E_1, E_2 of principal directions and use that

$$\begin{aligned} X &= \cos\phi E_1 + \sin\phi E_2, \\ \mathrm{II}(E_1, E_1) &= \kappa_1, \\ \mathrm{II}(E_2, E_2) &= \kappa_2, \\ \mathrm{II}(E_1, E_2) &= 0 = \mathrm{II}(E_2, E_1). \end{aligned}$$

□

As an important corollary we get a nice characterization of the principal curvatures.

Corollary 5.4.6. *Assume that the principal curvatures are ordered $\kappa_1 \geq \kappa_2$, then*

$$\begin{aligned} \max_{|X|=1} \mathrm{II}(X, X) &= \kappa_1, \\ \min_{|X|=1} \mathrm{II}(X, X) &= \kappa_2. \end{aligned}$$

We can now give a rather surprising characterization of planes and spheres.

Theorem 5.4.7. *(Meusnier, 1776) If a surface $\mathbf{q}$ has the property that $\kappa_1 = \kappa_2$ at all points, then $\kappa = \kappa_1$ is constant and the surface is part of a plane or sphere.*

Proof. Since the principal curvatures agree at all points it follows that

$$-\frac{\partial \mathbf{N}}{\partial w} = L\left(\frac{\partial \mathbf{q}}{\partial w}\right) = \kappa \frac{\partial \mathbf{q}}{\partial w}.$$

By letting $w = u, v$ and taking partial derivatives of this equation we obtain

$$\begin{aligned} -\frac{\partial^2 \mathbf{N}}{\partial u \partial v} &= \frac{\partial \kappa}{\partial u}\frac{\partial \mathbf{q}}{\partial v} + \kappa \frac{\partial^2 \mathbf{q}}{\partial u \partial v}, \\ -\frac{\partial^2 \mathbf{N}}{\partial v \partial u} &= \frac{\partial \kappa}{\partial v}\frac{\partial \mathbf{q}}{\partial u} + \kappa \frac{\partial^2 \mathbf{q}}{\partial v \partial u}. \end{aligned}$$

As partial derivatives commute it follows that

$$\frac{\partial \kappa}{\partial u}\frac{\partial \mathbf{q}}{\partial v} = \frac{\partial \kappa}{\partial v}\frac{\partial \mathbf{q}}{\partial u}.$$

Since $\frac{\partial \mathbf{q}}{\partial u}, \frac{\partial \mathbf{q}}{\partial v}$ are linearly independent this forces $\frac{\partial \kappa}{\partial u} = \frac{\partial \kappa}{\partial v} = 0$. Thus κ is constant.

Returning to the equation

$$-\frac{\partial \mathbf{N}}{\partial w} = \kappa \frac{\partial \mathbf{q}}{\partial w}$$

we see that

$$\frac{\partial (\mathbf{N} + \kappa \mathbf{q})}{\partial w} = 0.$$

This implies that $\mathbf{N} + \kappa \mathbf{q}$ is constant. When $\kappa = 0$ this shows that $\mathbf{N}$ is constant and consequently the surface lies in the plane orthogonal to $\mathbf{N}$. When κ does not vanish we can assume that $\kappa = \pm \frac{1}{R}$, $R > 0$. We then have that

$$\pm R\mathbf{N} + \mathbf{q} = \mathbf{c}$$

for some $\mathbf{c} \in \mathbb{R}^3$. This shows that

$$|\mathbf{q} - \mathbf{c}|^2 = R^2.$$

Hence $\mathbf{q}$ lies on the sphere of radius R centered at $\mathbf{c}$. $\square$

Exercises.

(1) Show that the principal curvatures for a parametrized surface are the roots to the equation
$$\det\left([\mathrm{II}] - \kappa\,[\mathrm{I}]\right) = 0$$

(2) Consider
$$z\left(x^2+y^2\right) = \kappa_1 x^2 + \kappa_2 y^2$$
 (a) Show that this defines a surface when $x^2+y^2>0$.
 (b) Show that it is a ruled surface where the lines go through the z-axis and are perpendicular to the z-axis.
 (c) Show that if a general surface has principal curvatures $\kappa_1,\ \kappa_2$ at a point, then z corresponds to the possible values of the normal curvature at that point.

(3) (Rodrigues) Show that a curve $\mathbf{q}(t)$ on a surface with normal $\mathbf{N}$ is a line of curvature if and only if
$$-\kappa(t)\frac{d\mathbf{q}}{dt} = \frac{d(\mathbf{N}\circ\mathbf{q})}{dt}.$$

(4) Show that the principal curvatures are constant if and only if the Gauss and mean curvatures are constant.

(5) Consider the *pseudo-sphere*
$$\mathbf{q} = \left(\frac{\cos\mu}{\cosh t}, \frac{\sin\mu}{\cosh t}, t - \tanh t\right).$$
This is a model for a surface with constant negative Gauss curvature. Note that the surface
$$\mathbf{q} = \left(\frac{\cos\mu}{\cosh t}, \frac{\sin\mu}{\cosh t}, \tanh t\right)$$
is the sphere with a conformal (Mercator) parametrization.
 (a) Compute the first and second fundamental forms
 (b) Compute the principal curvatures, Gauss curvature, and mean curvature.

(6) A ruled surface $\mathbf{q}(u,v) = \alpha(v) + uX(v)$ is called *developable* if all of the u-curves $\mathbf{q}(u) = \mathbf{q}(u,v)$ for fixed v are lines of curvature with $\kappa = 0$. Show that ruled surfaces that are developable have vanishing Gauss curvature.

(7) (Monge) Show that a curve $\mathbf{q}(t)$ on a surface with normal $\mathbf{N}$ is a line of curvature if and only if the ruled surface generated by $\mathbf{q}$ and $\mathbf{N}$ is developable.

(8) Show that if a surface has conformal Gauss map, then it is either minimal or part of a sphere.

(9) Show that if $\mathrm{III} = \lambda\mathrm{II}$ for some function λ on the surface, then either $K=0$ or the surface is part of a sphere.

(10) Show that all curves on a sphere or plane are lines of curvature. Use this to show that if two spheres, a plane and a sphere, or two planes intersect in a curve, then they intersect at a constant angles along this curve.

(11) Consider a unit speed curve $\alpha(s) : [0, L] \to \mathbb{R}^3$ with non-vanishing curvature and the tube of radius R around it

$$\mathbf{q}(s, \phi) = \alpha(s) + R(\mathbf{N}_\alpha \cos\phi + \mathbf{B}_\alpha \sin\phi)$$

(see section 4.1 exercise 8 and section 4.4 exercise 15). Show that the principal directions are $-\mathbf{N}_\alpha \sin\phi + \mathbf{B}_\alpha \cos\phi$ and $\mathbf{T}_\alpha$ with principal curvatures $1/R$ and $-\frac{\kappa \cos\phi}{1-\kappa R}$.

(12) Consider a parametrized surface $\mathbf{q}(t, \phi)$ where the s- and ϕ-curves correspond to the principal directions. Assume that the principal curvatures are $\kappa_2 < \kappa_1$ and $\kappa_1 = 1/R$ is constant.
  (a) Consider $\mathbf{c}(t, \phi) = \mathbf{q}(t, \phi) + R\mathbf{N}(t, \phi)$ and show that

$$\frac{\partial \mathbf{c}}{\partial \phi} = 0, \ \frac{\partial \mathbf{c}}{\partial t} \neq 0.$$

  (b) Conclude that $\mathbf{q}$ is a tube of radius R (see section 4.1 exercise 8).
  (c) Show that a surface without umbilics where one of the principal curvatures is constant is a tube.
  (d) Is it necessary to assume that the surface has no umbilics?

(13) (Joachimsthal) Let $\mathbf{q}(t)$ be a curve that lies on two surfaces M_1 and M_2 that have normals $\mathbf{N}_1$ and $\mathbf{N}_2$ respectively. Define

$$\theta(t) = \angle(\mathbf{N}_1 \circ \mathbf{q}(t), \mathbf{N}_2 \circ \mathbf{q}(t))$$

and assume that $0 < \theta(t) < \pi$, in other words the surfaces are not tangent to each other along the curve.
  (a) Show that if $\mathbf{q}(t)$ is a line of curvature on both surfaces, then $\theta(t)$ is constant.
  (b) Show that if $\mathbf{q}(t)$ is a line of curvature on one of the surfaces and $\theta(t)$ is constant, then $\mathbf{q}(t)$ is also a line of curvature on the other surface.

(14) Show that the geodesic torsion of a curve on a surface satisfies

$$\tau_g = (\kappa_2 - \kappa_1)\sin\phi\cos\phi,$$

where ϕ is the angle between the tangent to the curve and the principal direction corresponding to κ_1.

(15) (Rodrigues) Show that a unit speed curve on a surface is a line of curvature if and only if its geodesic torsion vanishes.

(16) Show that the principal curvatures at a point are equal if and only if the mean and Gauss curvatures at the point are related by $H^2 = K$.

(17) Let $\mathbf{q}(u, v)$ be a parametrized surface where the principal curvatures never coincide. Show that $\frac{\partial \mathbf{q}}{\partial u}$ and $\frac{\partial \mathbf{q}}{\partial v}$ are the principal directions if and only if $g_{uv} = 0 = L_{uv}$.

(18) Let $\mathbf{q}(u, v)$ be a parametrized surface and $\mathbf{q}^\epsilon = \mathbf{q} + \epsilon\mathbf{N}$ the parallel surface at distance ϵ from $\mathbf{q}$.
  (a) Show that

$$\frac{\partial \mathbf{q}^\epsilon}{\partial w} = \frac{\partial \mathbf{q}}{\partial w} + \epsilon \frac{\partial \mathbf{N}}{\partial w} = (I + \epsilon L)\left(\frac{\partial \mathbf{q}}{\partial w}\right),$$

where I is the identity map $I(v) = v$.

(b) Show that $\mathbf{q}^\epsilon$ is a parametrized surface if $|\epsilon| < \min\left\{\frac{1}{|\kappa_1|}, \frac{1}{|\kappa_2|}\right\}$ and that $\mathbf{N}$ is also the natural normal to $\mathbf{q}^\epsilon$.

(c) Show that
$$L^\epsilon = L \circ (I + \epsilon L)^{-1}$$
by using that
$$L\left(\frac{\partial \mathbf{q}}{\partial w}\right) = \frac{\partial \mathbf{N}}{\partial w} = L^\epsilon\left(\frac{\partial \mathbf{q}^\epsilon}{\partial w}\right).$$

(d) Show that these surfaces all have the same principal directions and that the principal curvatures satisfy
$$\frac{1}{\kappa_i^\epsilon} = \frac{\kappa_i}{1 + \epsilon\kappa_i}.$$

(19) Let $\mathbf{q}(u, v)$ be a parametrized surface and $\mathbf{q}^\epsilon = \mathbf{q} + \epsilon\mathbf{N}$ the parallel surface at distance ϵ from $\mathbf{q}$.

(a) Show that
$$\mathrm{I}^\epsilon = \mathrm{I} - 2\epsilon\mathrm{II} + \epsilon^2\mathrm{III}.$$

(b) Show that
$$\mathrm{II}^\epsilon = \mathrm{II} - \epsilon\mathrm{III}.$$

(c) Show that
$$\mathrm{III}^\epsilon = \mathrm{III}.$$

(d) How do you reconcile this with the formula
$$L^\epsilon = L \circ (I + \epsilon L)^{-1}$$
from the previous exercise?

(e) Show that
$$K^\epsilon = \frac{K}{1 + 2\epsilon H + \epsilon^2 K}$$
and
$$H^\epsilon = \frac{H - \epsilon K}{1 + 2\epsilon H + \epsilon^2 K}.$$

5.5. Ruled Surfaces

A ruled surface is parameterized by selecting a curve $\alpha(v)$ and then considering the surface one obtains by adding a line through each of the points on the curve. If the directions of those lines are given by $X(v)$, then the surface can be parametrized by $\mathbf{q}(u, v) = \alpha(v) + uX(v)$. We can without loss of generality assume that X is a unit field. The condition for obtaining a parametrized surface is that $\frac{\partial \mathbf{q}}{\partial u} = X$ and $\frac{\partial \mathbf{q}}{\partial v} = \frac{d\alpha}{dv} + u\frac{dX}{dv}$ are linearly independent. Even though we don't always obtain a surface for all parameter values it is important to consider the extended lines in the rulings for all values of v.

Example 5.5.1. A *generalized cylinder* is a ruled surface where X is constant, i.e., $\frac{dX}{dv} = 0$. This will be a parametrized surface everywhere if X is never tangent to α.

Example 5.5.2. A *generalized cone* is a ruled surface where α can be chosen to be constant, i.e., $\frac{d\alpha}{dv} = 0$. This will clearly not be a parametrized surface when $u = 0$.

EXAMPLE 5.5.3. A *tangent developable*, is a ruled surface where X is always tangent to α, i.e., X and $\frac{d\alpha}{dv}$ are always proportional. This is also not a surface when $u = 0$. Note that generalized cones can be considered a special case of tangent developables. It is not unusual to also assume that that a tangent developable has the property that α is regular so as to avoid this overlap in definitions.

One of our goals is to understand when we obtain a tangent developable as that might not be obvious from a general parametrization.

An example of a cone that is not rotationally symmetric is the elliptic cone

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = z^2.$$

The elliptic hyperboloid

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = z^2 + 1$$

is an example of a surface that is ruled in two different ways, but which does not have zero Gauss curvature. We can let

$$\alpha(t) = (a\cos(t), b\sin(t), 0)$$

be the ellipse where $z = 0$. The fields generating the lines are given by

$$X = \frac{d\alpha}{dt} + (0, 0, \pm 1)$$

and it is not difficult to check that

$$\mathbf{q}(s,t) = \alpha(t) + s\left(\frac{d\alpha}{dt} + (0, 0, \pm 1)\right)$$

are both rulings of the elliptic hyperboloid.

PROPOSITION 5.5.4. *Ruled surfaces have non-positive Gauss curvature and the Gauss curvature vanishes if and only if*

$$\left(X \times \frac{d\alpha}{dv}\right) \cdot \frac{dX}{dv} = 0.$$

In particular, generalized cylinders, generalized cones, and tangent developables have vanishing Gauss curvature.

PROOF. Since $\frac{\partial^2 \mathbf{q}}{\partial u^2} = 0$ it follows that $L_{uu} = 0$. Thus

$$K = \frac{-L_{uv}^2}{g_{uu}g_{vv} - g_{uv}^2} \leq 0.$$

Moreover, K vanishes precisely when

$$\frac{\partial^2 \mathbf{q}}{\partial u \partial v} = \frac{dX}{dv}$$

is perpendicular to the normal. Since the normal is given by

$$\mathbf{N} = \frac{X \times \left(\frac{d\alpha}{dv} + u\frac{dX}{dv}\right)}{\left|X \times \left(\frac{d\alpha}{dv} + u\frac{dX}{dv}\right)\right|}$$

this translates to

$$\begin{aligned} 0 &= \left(X \times \left(\frac{d\alpha}{dv} + u\frac{dX}{dv}\right)\right) \cdot \frac{dX}{dv} \\ &= \left(X \times \frac{d\alpha}{dv}\right) \cdot \frac{dX}{dv}. \end{aligned}$$

which is what we wanted to prove. □

Definition 5.5.5. A ruled surface with the property that the normal is constant in the direction of the ruling, i.e.,

$$\frac{\partial \mathbf{N}}{\partial u} = 0,$$

is called a *developable* or *developable surface.*

The main focus in this section will be to understand what exactly characterizes developables and to narrow the types of surfaces with this property.

We start with a characterization in terms of Gauss curvature.

Lemma 5.5.6. *(Monge, 1775) A surface with vanishing Gauss curvature and no umbilics is a developable surface. Conversely any developable has vanishing Gauss curvature.*

Proof. First note that a developable has the property that the lines in the ruling are lines of curvature and that the principal value vanishes in the direction of the lines. The establishes the second claim and also guides us as to how to find the lines in a ruling.

Assume now that the surface has zero Gauss curvature. We shall show that the principal directions that correspond to the principal value 0 generate lines of curvature that are straight lines. This will create a ruling and the normal is by definition constant along these lines as they a lines of curvature for for the principal value that vanishes.

Since the surface has no umbilics the two principal unit directions are well-defined up to a choice of sign. In particular, we can find orthonormal vector fields E_1, E_2 of principal directions. We can then use theorem 4.2.7 select a parametrization where $\frac{\partial \mathbf{q}}{\partial u}, \frac{\partial \mathbf{q}}{\partial v}$ are also principal directions. This implies that $g_{uv} = L_{uv} = 0$. Using that the Gauss curvature vanishes further allows us to assume that

$$\begin{aligned} -\frac{\partial \mathbf{N}}{\partial u} &= L\left(\frac{\partial \mathbf{q}}{\partial u}\right) = 0, \\ -\frac{\partial \mathbf{N}}{\partial v} &= L\left(\frac{\partial \mathbf{q}}{\partial v}\right) = \kappa\frac{\partial \mathbf{q}}{\partial v},\ \kappa \neq 0. \end{aligned}$$

In particular,

$$\frac{\partial^2 \mathbf{N}}{\partial u \partial v} = 0.$$

This shows that

$$\begin{aligned}
\frac{\partial^2\mathbf{q}}{\partial u^2}\cdot\frac{\partial\mathbf{q}}{\partial v} &= \kappa^{-1}\frac{\partial^2\mathbf{q}}{\partial u^2}\cdot L\left(\frac{\partial\mathbf{q}}{\partial v}\right)\\
&= -\kappa^{-1}\frac{\partial^2\mathbf{q}}{\partial u^2}\cdot\frac{\partial\mathbf{N}}{\partial v}\\
&= -\kappa^{-1}\frac{\partial}{\partial u}\left(\frac{\partial\mathbf{q}}{\partial u}\cdot\frac{\partial\mathbf{N}}{\partial v}\right)+\kappa^{-1}\frac{\partial\mathbf{q}}{\partial u}\cdot\frac{\partial^2\mathbf{N}}{\partial v\partial u}\\
&= 0.
\end{aligned}$$

By assumption we also have

$$\frac{\partial^2\mathbf{q}}{\partial u^2}\cdot\mathbf{N}=0.$$

Thus $\frac{\partial^2\mathbf{q}}{\partial u^2}$ must be parallel to $\frac{\partial\mathbf{q}}{\partial u}$. This shows that the u-curves on the surface have zero curvature as curves in $\mathbb{R}^3$. This shows that they are straight lines. □

The next result shows that ruled surfaces admit a standard set of parameters. The goal is show that developables are locally forced to be either a generalized cylinder, generalized cone, or a tangent developable. It is always easy to recognize generalized cylinders as that occurs precisely when X is constant. However, as we already discussed, it is less obvious when the other two cases occur.

PROPOSITION 5.5.7. *A ruled surface* $\mathbf{q}(u,v)=\alpha(v)+uX(v)$ *can be reparametrized as*

$$\mathbf{q}(s,v)=c(v)+sX(v),$$

where $\frac{dc}{dv}\perp\frac{dX}{dv}$.

The ruled surface is a generalized cone if and only if c is constant. The ruled surface is a tangent developable if and only if $\frac{dc}{dv}$ and X are proportional at all points v.

PROOF. Note that no change in the parametrization is necessary if X is constant. When $\frac{dX}{dv}\neq 0$ define

$$c=\alpha-\frac{\frac{d\alpha}{dv}\cdot\frac{dX}{dv}}{\left|\frac{dX}{dv}\right|^2}X$$

and

$$s=u+\frac{\frac{d\alpha}{dv}\cdot\frac{dX}{dv}}{\left|\frac{dX}{dv}\right|^2}.$$

Then it is clear that $\mathbf{q}(u,v)=c(v)+sX(v)=\mathbf{q}(s,v)$. Moreover as X is a unit field it is perpendicular to its derivative so we have

$$\begin{aligned}
\frac{dc}{dv}\cdot\frac{dX}{dv} &= \left(\frac{d\alpha}{dv}-\frac{d}{dv}\left(\frac{\frac{d\alpha}{dv}\cdot\frac{dX}{dv}}{\left|\frac{dX}{dv}\right|^2}\right)X-\left(\frac{\frac{d\alpha}{dv}\cdot\frac{dX}{dv}}{\left|\frac{dX}{dv}\right|^2}\right)\frac{dX}{dv}\right)\cdot\frac{dX}{dv}\\
&= \frac{d\alpha}{dv}\cdot\frac{dX}{dv}-\left(\frac{\frac{d\alpha}{dv}\cdot\frac{dX}{dv}}{\left|\frac{dX}{dv}\right|^2}\right)\frac{dX}{dv}\cdot\frac{dX}{dv}\\
&= \frac{d\alpha}{dv}\cdot\frac{dX}{dv}-\left(\frac{d\alpha}{dv}\cdot\frac{dX}{dv}\right)\\
&= 0.
\end{aligned}$$

It is clear that we obtain a generalized cone when c is constant and a tangent developable if $\frac{dc}{dv}$ and X are parallel to each other.

Conversely if the ruled surface $\mathbf{q}(u,v)$ is a generalized cone, then there is a unique function $u = u(v)$ such that $\mathbf{q}(u(v), v)$ is constant. Thus

$$0 = \frac{d\alpha}{dv} + u(v)\frac{dX}{dv} + \frac{du(v)}{dv}X.$$

If we multiply by $\frac{dX}{dv}$, then we obtain

$$u(v) = -\frac{\frac{d\alpha}{dv} \cdot \frac{dX}{dv}}{\left|\frac{dX}{dv}\right|^2}.$$

This corresponds exactly to $s = 0$ in the parametrization $\mathbf{q}(s,v) = c(v) + sX(v)$. So it follows that $c(v)$ is constant.

When the ruled surface is a tangent developable it is possible to find $u = u(v)$ such that the curve $\beta(v) = \mathbf{q}(u(v), v)$ is tangent to the extended lines in the ruling, i.e., $\frac{d\beta}{dv}$ and X are proportional. In particular,

$$\begin{aligned} 0 &= \frac{d\beta}{dv} \cdot \frac{dX}{dv} \\ &= \left(\frac{d\alpha}{dv} + u(v)\frac{dX}{dv} + \frac{du(v)}{dv}X\right) \cdot \frac{dX}{dv} \\ &= \frac{d\alpha}{dv} \cdot \frac{dX}{dv} + u(v)\left|\frac{dX}{dv}\right|^2. \end{aligned}$$

So again we obtain that $u(v)$ corresponds exactly to $s = 0$, which forces β to be c. $\square$

THEOREM 5.5.8. *(Monge, 1775) A developable surface is a generalized cylinder, generalized cone, or a tangent developable at almost all points of the surface.*

PROOF. We can assume that the surface is given by

$$\mathbf{q}(s,v) = c(v) + sX(v),$$

where $\frac{dc}{dv} \perp \frac{dX}{dv}$. The Gauss curvature vanishes precisely when

$$\left(X \times \frac{dc}{dv}\right) \cdot \frac{dX}{dv} = 0.$$

If $\frac{dX}{dv} = 0$ on an interval, then the surface is a generalized cylinder. So we can assume that $\frac{dX}{dv} \neq 0$. This implies that X and $\frac{dX}{dv}$ are linearly independent as they are orthogonal. The condition

$$\left(X \times \frac{dc}{dv}\right) \cdot \frac{dX}{dv} = 0$$

on the other hand implies that the three vectors are linearly dependent. We already know that $\frac{dc}{dv} \perp \frac{dX}{dv}$, so this forces

$$\frac{dc}{dv} = \left(\frac{dc}{dv} \cdot X\right) X.$$

When $\frac{dc}{dv} \neq 0$ then X is tangent to c and so we have a tangent developable. On the other hand, if $\frac{dc}{ds} = 0$ on an interval, then the surface must be a generalized cone on that interval.

Thus the surface is divided into regions each of which can be identified with our three basic types of ruled surfaces and then glued together along lines that go through parameter values where either $\frac{dX}{dv}=0$ or $\frac{dc}{dv}=0$. $\square$

There is also a similar and very interesting result for ruled minimal surfaces.

THEOREM 5.5.9. *(Catalan) Any ruled surface that is minimal is planar or a helicoid at almost all points of the surface.*

PROOF. Assume that we have a parametrization $\mathbf{q}(s,v)=c(v)+sX(v)$ where $\frac{dc}{dv}\cdot\frac{dX}{dv}=0$. In case the surface also has vanishing Gauss curvature it follows that it is planar as the second fundamental form vanishes. Therefore, we can assume that both c and X are regular curves and additionally that $\frac{dc}{dv}$ is not parallel to X.

The mean curvature is given by the general formula

$$H=\frac{L_{ss}g_{vv}-2L_{sv}g_{sv}+L_{vv}g_{ss}}{2\left(g_{ss}g_{vv}-g_{sv}^{2}\right)},$$

where

$$\begin{aligned} g_{ss} &= 1,\\ g_{sv} &= \frac{dc}{dv}\cdot X,\\ g_{vv} &= \left|\frac{dc}{dv}\right|^{2}+s^{2}, \end{aligned}$$

$$\mathbf{N}=\frac{X\times\left(\frac{dc}{dv}+s\frac{dX}{dv}\right)}{\left|X\times\left(\frac{dc}{dv}+s\frac{dX}{dv}\right)\right|},$$

$$\begin{aligned} L_{ss} &= 0,\\ L_{sv} &= -\frac{dX}{dv}\cdot\mathbf{N},\\ L_{vv} &= -\left(\frac{d^{2}c}{dv^{2}}+s\frac{d^{2}X}{dv^{2}}\right)\cdot\mathbf{N}. \end{aligned}$$

Thus $H=0$ precisely when

$$-2\left(\frac{dc}{dv}\cdot X\right)\left(\frac{dX}{dv}\cdot\mathbf{N}\right)=-\left(\frac{d^{2}c}{dv^{2}}+s\frac{d^{2}X}{dv^{2}}\right)\cdot\mathbf{N},$$

which implies

$$2\left(\frac{dc}{dv}\cdot X\right)\left(\frac{dX}{dv}\cdot\left(X\times\left(\frac{dc}{dv}+s\frac{dX}{dv}\right)\right)\right)=\left(\frac{d^{2}c}{dv^{2}}+s\frac{d^{2}X}{dv^{2}}\right)\cdot\left(X\times\left(\frac{dc}{dv}+s\frac{dX}{dv}\right)\right).$$

The left hand side can be simplified

$$2\left(\frac{dc}{dv}\cdot X\right)\left(\frac{dX}{dv}\cdot\left(X\times\left(\frac{dc}{dv}+s\frac{dX}{dv}\right)\right)\right)=2\left(\frac{dc}{dv}\cdot X\right)\left(\frac{dX}{dv}\cdot\left(X\times\frac{dc}{dv}\right)\right).$$

This shows that it is independent of s. The right hand side can be expanded in terms of s as follows

$$\begin{aligned}\left(\frac{d^2c}{dv^2}+s\frac{d^2X}{dv^2}\right)\cdot\left(X\times\left(\frac{dc}{dv}+s\frac{dX}{dv}\right)\right) &= \frac{d^2c}{dv^2}\cdot\left(\frac{dc}{dv}\times X\right)\\ &\quad +s\left(\frac{d^2c}{dv^2}\cdot\left(X\times\frac{dX}{dv}\right)+\frac{d^2X}{dv^2}\cdot\left(X\times\frac{dc}{dv}\right)\right)\\ &\quad +s^2\frac{d^2X}{dv^2}\cdot\left(X\times\frac{dX}{dv}\right).\end{aligned}$$

This leads us to 3 identities depending on the powers of s. From the s^2-term we conclude

$$\frac{d^2X}{dv^2}\in\operatorname{span}\left\{X,\frac{dX}{dv}\right\}.$$

At this point it is convenient to assume that v is the arclength parameter for X. With that in mind we have

$$\begin{aligned}\frac{d^2X}{dv^2} &= \left(\frac{d^2X}{dv^2}\cdot X\right)X+\left(\frac{d^2X}{dv^2}\cdot\frac{dX}{dv}\right)\frac{dX}{dv}\\ &= -\left(\frac{dX}{dv}\cdot\frac{dX}{dv}\right)X\\ &= -X.\end{aligned}$$

This implies that X is in fact a planar circle or radius 1. For simplicity let us further assume that it is the unit circle in the (x,y)-plane, i.e.,

$$X(v)=(\cos v,\sin v,0).$$

From the s-term we obtain

$$\begin{aligned}0 &= \frac{d^2c}{dv^2}\cdot\left(X\times\frac{dX}{dv}\right)+\frac{d^2X}{dv^2}\cdot\left(X\times\frac{dc}{dv}\right)\\ &= \frac{d^2c}{dv^2}\cdot\left(X\times\frac{dX}{dv}\right)-X\cdot\left(X\times\frac{dc}{dv}\right)\\ &= \frac{d^2c}{dv^2}\cdot\left(X\times\frac{dX}{dv}\right)\\ &= \frac{d^2c}{dv^2}\cdot\begin{bmatrix}0\\0\\1\end{bmatrix}\end{aligned}$$

showing that $\frac{d^2c}{dv^2}$ also lies in the (x,y)-plane. In particular,

$$\frac{dc}{dv}\cdot\begin{bmatrix}0\\0\\1\end{bmatrix}=h$$

is constant. Since $\frac{dc}{dv}\perp\frac{dX}{dv}$ we obtain

$$\frac{dc}{dv}=\left(\frac{dc}{dv}\cdot X\right)X+\begin{bmatrix}0\\0\\h\end{bmatrix}$$

and

$$\frac{dc}{dv} \times X = \begin{bmatrix} 0 \\ 0 \\ h \end{bmatrix} \times X = h\frac{dX}{dv}.$$

This considerably simplifies the terms that are independent of s in the mean curvature equation

$$2\left(\frac{dc}{dv} \cdot X\right)\left(\frac{dX}{dv} \cdot \left(X \times \frac{dc}{dv}\right)\right) = \frac{d^2c}{dv^2} \cdot \left(X \times \frac{dc}{dv}\right)$$

as we now obtain

$$\begin{aligned} 2h\frac{dc}{dv} \cdot X &= h\frac{d^2c}{dv^2} \cdot \frac{dX}{dv} \\ &= -h\frac{dc}{dv} \cdot \frac{d^2X}{dv^2} \\ &= h\frac{dc}{dv} \cdot X. \end{aligned}$$

When $h = 0$ the curve c also lies in the (x, y)-plane and the surface is planar. Otherwise $\frac{dc}{dv} \cdot X = 0$ which implies that

$$\frac{dc}{dv} = \left(\frac{dc}{dv} \cdot X\right) X + \begin{bmatrix} 0 \\ 0 \\ h \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ h \end{bmatrix}$$

and

$$c = \begin{bmatrix} 0 \\ 0 \\ hv + v_0 \end{bmatrix}$$

for a constant v_0.

The surface is then given by

$$\mathbf{q}(s, v) = \begin{bmatrix} s\cos v \\ s\sin v \\ hv + v_0 \end{bmatrix},$$

which shows explicitly that it is a helicoid. □

Exercises.

(1) Let $\mathbf{q}(s)$ be a unit speed asymptotic line (see section 5.1 exercise 1) of a ruled surface $\mathbf{q}(u, v) = \alpha(v) + uX(v)$. Note that u-curves are asymptotic lines.

(a) Show that

$$\det\begin{bmatrix} \ddot{\mathbf{q}}, & X, & \frac{d\alpha}{dv} + u\frac{dX}{dv} \end{bmatrix} = 0.$$

(b) Assume for the remainder of the exercise that $K < 0$. Show that there is a unique asymptotic line through through every point that is not tangent to X.

(c) Show that this asymptotic line can locally be reparametrized as

$$\alpha(v) + u(v)X(v),$$

where

$$\frac{du}{dv} = \frac{\det \left[\begin{array}{ccc} X, & \frac{d\alpha}{dv} + u(v)\frac{dX}{dv}, & \frac{d^2\alpha}{dv^2} + u(v)\frac{d^2X}{dv^2} \end{array} \right]}{2\det \left[\begin{array}{ccc} \frac{d\alpha}{dv}, & X, & \frac{dX}{dv} \end{array} \right]}.$$

(2) Show that a generalized cylinder $\mathbf{q}(u,v) = \alpha(v) + uX$ where X is a fixed unit vector admits a parametrization $\mathbf{q}(s,t) = c(t) + sX$, where c is parametrized by arclength and lies a plane orthogonal to X.

(3) Let $\mathbf{q}(u,v)$ be a parametrized surface surface where at each point q there is a regular curve $\mathbf{q}(t) = \mathbf{q}(u(t), v(t))$ with $\mathbf{q}(0) = q$ and $\ddot{\mathbf{q}}(0) = \dddot{\mathbf{q}}(0) = 0$ as a curve in $\mathbb{R}^3$. Show that $\dot{\mathbf{q}}(0)$ is a principal direction with principal curvature 0.

(4) Consider a parameterized surface $\mathbf{q}(u,v)$. Show that the Gauss curvature vanishes if and only if $\frac{\partial \mathbf{N}}{\partial u}, \frac{\partial \mathbf{N}}{\partial v}$ are linearly dependent everywhere.

(5) Consider

$$\mathbf{q}(u,v) = \left(u+v, u^2+2uv, u^3+3u^2v\right).$$

(a) Determine when it defines a surface.
(b) Show that the Gauss curvature vanishes.
(c) What type of ruled surface is it?

(6) Consider the Monge patch

$$z = \sum_{k=2}^{n} (ax+by)^k + cx + dy + f.$$

(a) Show that the Gauss curvature vanishes.
(b) Depending on the values of a and b determine the type of ruled surface.

(7) Consider the equation

$$xy = (z-c)^2.$$

(a) Determine when it defines a surface.
(b) Show that the Gauss curvature vanishes.
(c) What type of ruled surface is it?

(8) Consider the equation

$$4\left(y-x^2\right)\left(xz-y^2\right) = (xy-z)^2.$$

(a) Determine when it defines a surface.
(b) Show that the Gauss curvature vanishes.
(c) What type of ruled surface is it?

(9) Show that a surface given by an equation

$$F(x,y,z) = R$$

has has vanishing Gauss curvature if and only if

$$\det \begin{bmatrix} \frac{\partial^2 F}{\partial x^2} & \frac{\partial^2 F}{\partial y \partial x} & \frac{\partial^2 F}{\partial z \partial x} & \frac{\partial F}{\partial x} \\ \frac{\partial^2 F}{\partial x \partial y} & \frac{\partial^2 F}{\partial y^2} & \frac{\partial^2 F}{\partial z \partial y} & \frac{\partial F}{\partial y} \\ \frac{\partial^2 F}{\partial x \partial z} & \frac{\partial^2 F}{\partial y \partial z} & \frac{\partial^2 F}{\partial z^2} & \frac{\partial F}{\partial z} \\ \frac{\partial F}{\partial x} & \frac{\partial F}{\partial y} & \frac{\partial F}{\partial z} & 0 \end{bmatrix} = 0.$$

(10) (Euler, 1775) Let $\alpha(t)$ be a unit speed space curve with curvature $\kappa(t) > 0$. Show that the tangent developable

$$\mathbf{q}(s,t) = \alpha(t) + s\frac{d\alpha}{dt}$$

admits Cartesian coordinates. Hint: There is a unit speed planar curve $\beta(t)$ whose curvature is $\kappa(t)$. Show that there is a natural isometry between the part of the plane parametrized by

$$\mathbf{q}^*(s,t) = \beta(t) + s\frac{d\beta}{dt}$$

and the tangent developable $\mathbf{q}(s,t)$.

(11) Use the previous exercise to show that a surface with $K = 0$ and no umbilics locally admits Cartesian coordinates at almost all points.

(12) Show that a ruled surface with constant and non-zero mean curvature is a generalized cylinder.

(13) Show directly that if a minimal surface has vanishing Gauss curvature, then it is part of a plane.

(14) Assume that we have a ruled surface

$$\mathbf{q}(u,v) = \alpha(v) + uX(v)$$

where $|X| = 1$.

(a) Show that if we use

$$c = \alpha + \left(a - \int \frac{d\alpha}{dv}\cdot X dv\right) X$$

for some constant a, then

$$\mathbf{q}(u,v) = c(v) + uX(v)$$

parametrizes the same surface and has the property that all v-curves are orthogonal to X and thus to the lines in the ruling.

(b) Show that if $\frac{dX}{dv} \neq 0$, then we can reparametrize X by arclength and thus obtain a parametrization

$$\mathbf{q}(u,t) = c(t) + uX(t),$$

where the t-curves are orthogonal to the ruling and X is a unit field parametrized by arclength.

(c) Show that if c is regular and has positive curvature and s denotes the arclength parameter for c we obtain $X(s) = \cos(\phi(s))\mathbf{N} + \sin(\phi(s))\mathbf{B}$.

(15) Assume that we have a minimal ruled surface

$$\mathbf{q}(u,t) = c(t) + uX(t)$$

as in the previous exercise with t-curves perpendicular to X and X a unit field parametrized by arclength. Reprove Catalan's theorem using this parametrization. Hint: One strategy is to first show that X is a unit circle, then show that $\ddot{c}$ is proportional to X, and finally conclude that the t-curves are all Bertrand mates to each other.

(16) Consider the quartic equation with variable t:

$$x + yt + zt^2 + t^4 = 0$$

and discriminant:

$$D = \left(x + \frac{z^2}{12}\right)^3 - 27\left(\frac{xz}{6} - \frac{y^2}{16} - \frac{z^3}{216}\right)^2.$$

Show that $D = 0$ corresponds to the tangent developable of the curve $\left(-3t^4, 8t^3, -6t^2\right)$.

(17) Consider a family of planes in the (x, y, z)-space parametrized by t:

$$F(x, y, z, t) = a(t)\,x + b(t)\,y + c(t)\,z + d(t) = 0.$$

An *envelope* to this family is a surface such that its tangents are precisely the planes of this family.

(a) Show that an envelope exists and can be determined by the equations:

$$\begin{aligned} F = a(t)\,x + b(t)\,y + c(t)\,z + d(t) &= 0 \\ \frac{\partial F}{\partial t} = \dot{a}(t)\,x + \dot{b}(t)\,y + \dot{c}(t)\,z + \dot{d}(t) &= 0 \end{aligned}$$

when

$$\begin{bmatrix} a & b & c \\ \dot{a} & \dot{b} & \dot{c} \end{bmatrix}$$

has rank 2. Hint: use t and one of the coordinates x, y, z as parameters. The parametrization might be singular for some parameter values.

(b) Show that the envelope is a ruled surface.

(c) Show that the envelope is a generalized cylinder when the three functions $a, b,$ and c are linearly dependent.

(d) Show that the envelope is a generalized cone when the function d is a linear combination of $a, b,$ and c.

(e) Show that the envelope is a tangent developable when the Wronskian

$$\det \begin{bmatrix} a & b & c & d \\ \dot{a} & \dot{b} & \dot{c} & \dot{d} \\ \ddot{a} & \ddot{b} & \ddot{c} & \ddot{d} \\ \dddot{a} & \dddot{b} & \dddot{c} & \dddot{d} \end{bmatrix} \neq 0$$

Hint: Show that the equations:

$$\begin{aligned} F = a(t)\,x + b(t)\,y + c(t)\,z + d(t) &= 0 \\ \frac{\partial F}{\partial t} = \dot{a}(t)\,x + \dot{b}(t)\,y + \dot{c}(t)\,z + \dot{d}(t) &= 0 \\ \frac{\partial^2 F}{\partial t^2} = \ddot{a}(t)\,x + \ddot{b}(t)\,y + \ddot{c}(t)\,z + \ddot{d}(t) &= 0 \end{aligned}$$

determine the curve that generates the tangent developable.

(f) Show that for fixed (x_0, y_0, z_0) the number of solutions or roots to the equation $F(x_0, y_0, z_0, t) = 0$ corresponds to the number of tangent planes to the envelope that pass through (x_0, y_0, z_0).

(18) Let $\mathbf{q}(u,v) = \alpha(v) + uX(v)$ be a developable surface, i.e., the u-curves are lines of curvature with $\kappa = 0$. Show that there exist functions $a(v), b(v), c(v),$ and $d(v)$ such that the surface is an envelope of the planes

$$a(v)x + b(v)y + c(v)z + d(v) = 0.$$

CHAPTER 6

Intrinsic Geometry

The goal in this chapter is to show that many of the the calculations we do on a surface can be done intrinsically. This means that they can be done without reference to the normal vector and thus are only allowed to depend on the first fundamental form. The highlights are the Gauss equations, Codazzi-Mainardi equations, and the Gauss-Bonnet theorem.

6.1. Calculating Christoffel Symbols and The Gauss Curvature

We start by showing that the tangential components of the derivatives

$$\frac{\partial^2 \mathbf{q}}{\partial w_1 \partial w_2}$$

can be calculated intrinsically. For the Gauss formulas this amounts to showing that the Christoffel symbols can be calculated from the first fundamental form (see section 4.6). In particular, this shows that they can be computed knowing only the first derivatives of $\mathbf{q}(u,v)$ despite the fact that they are defined using the second derivatives!

Proposition 6.1.1. *The Christoffel symbols satisfy*

$$\begin{aligned}
\Gamma_{uuu} &= \frac{1}{2}\frac{\partial g_{uu}}{\partial u}, \\
\Gamma_{uvu} &= \frac{1}{2}\frac{\partial g_{uu}}{\partial v} = \Gamma_{vuu}, \\
\Gamma_{vvv} &= \frac{1}{2}\frac{\partial g_{vv}}{\partial v}, \\
\Gamma_{uvv} &= \frac{1}{2}\frac{\partial g_{vv}}{\partial u} = \Gamma_{vuv}, \\
\Gamma_{uuv} &= \frac{\partial g_{uv}}{\partial u} - \frac{1}{2}\frac{\partial g_{uu}}{\partial v}, \\
\Gamma_{vvu} &= \frac{\partial g_{uv}}{\partial v} - \frac{1}{2}\frac{\partial g_{vv}}{\partial u}.
\end{aligned}$$

Proof. We prove only two of these as the proofs are all similar. First use the product rule to see

$$\Gamma_{uvu} = \frac{\partial^2 \mathbf{q}}{\partial u \partial v} \cdot \frac{\partial \mathbf{q}}{\partial u} = \left(\frac{\partial}{\partial v}\left(\frac{\partial \mathbf{q}}{\partial u}\right)\right) \cdot \frac{\partial \mathbf{q}}{\partial u} = \frac{1}{2}\frac{\partial}{\partial v}\left(\frac{\partial \mathbf{q}}{\partial u} \cdot \frac{\partial \mathbf{q}}{\partial u}\right) = \frac{1}{2}\frac{\partial g_{uu}}{\partial v}.$$

Now use this together with the product rule to find

$$\begin{aligned}
\Gamma_{uuv} &= \frac{\partial^2 \mathbf{q}}{\partial u \partial u} \cdot \frac{\partial \mathbf{q}}{\partial v} = \left(\frac{\partial}{\partial u} \left(\frac{\partial \mathbf{q}}{\partial u} \right) \right) \cdot \frac{\partial \mathbf{q}}{\partial v} \\
&= \frac{\partial}{\partial u} \left(\frac{\partial \mathbf{q}}{\partial u} \cdot \frac{\partial \mathbf{q}}{\partial v} \right) - \left(\frac{\partial \mathbf{q}}{\partial u} \cdot \frac{\partial}{\partial u} \frac{\partial \mathbf{q}}{\partial v} \right) \\
&= \frac{\partial g_{uv}}{\partial u} - \frac{\partial \mathbf{q}}{\partial u} \cdot \frac{\partial^2 \mathbf{q}}{\partial u \partial v} \\
&= \frac{\partial g_{uv}}{\partial u} - \frac{1}{2} \frac{\partial g_{uu}}{\partial v}.
\end{aligned}$$

A similar proof starts with the observation that we have 8 equations of the form

$$\frac{\partial g_{w_1 w_2}}{\partial w} = \frac{\partial}{\partial w} \left(\frac{\partial \mathbf{q}}{\partial w_1} \cdot \frac{\partial \mathbf{q}}{\partial w_2} \right) = \Gamma_{w w_2 w_1} + \Gamma_{w w_1 w_2}$$

The symmetry on the left hand side $g_{uv} = g_{vu}$ reduces this to only 6 equations. However, on the right hand side there are also only 6 Christoffel symbols due to the symmetry

$$\Gamma_{uvw} = \frac{\partial^2 \mathbf{q}}{\partial u \partial v} \cdot \frac{\partial \mathbf{q}}{\partial w} = \Gamma_{vuw}.$$

More explicitly we have 4 equations that come in pairs:

$$\frac{\partial g_{uu}}{\partial w} = \Gamma_{wuu} + \Gamma_{wuu} = 2\Gamma_{wuu},$$

$$\frac{\partial g_{vv}}{\partial w} = \Gamma_{wvv} + \Gamma_{wvv} = 2\Gamma_{wvv}.$$

These give us 2 of the Christoffel symbols. After using the symmetries $\Gamma_{wuu} = \Gamma_{uwu}$ and $\Gamma_{wvv} = \Gamma_{vwv}$ we then obtain 2 more Christoffel symbols. The last pair of equations:

$$\frac{\partial g_{uv}}{\partial w} = \Gamma_{wuv} + \Gamma_{wvu}$$

give us the last 2 Christoffel symbols by noting that 2 of the 4 symbols involved were computed from the first 4 equations. □

Remark 6.1.2. There is a unified formula for all of these equations. It is convenient to simplify notation by using i, j, k, l instead of u, v and also $\partial_w = \frac{\partial}{\partial w}$. The expression is a bit more complicated and is less useful for actual calculations:

$$\Gamma_{ijk} = \frac{1}{2} \left(\partial_i g_{jk} + \partial_j g_{ik} - \partial_k g_{ij} \right).$$

The proposition can also be used to find the Christoffel symbols of the second kind. For example

$$\begin{aligned}
\Gamma_{uv}^{u} &= g^{uu} \Gamma_{uvu} + g^{uv} \Gamma_{uvv} \\
&= \frac{1}{2} \left(g^{uu} \frac{\partial g_{uu}}{\partial v} + g^{uv} \frac{\partial g_{vv}}{\partial u} \right)
\end{aligned}$$

and for the general formula

$$\Gamma_{ij}^{k} = g^{kl} \Gamma_{ijl} = g^{ku} \Gamma_{iju} + g^{kv} \Gamma_{ijv}.$$

While these formulas for the Christoffel symbols can't be made simpler as such, it is possible to be a bit more efficient when calculations are done. Specifically we often do calculations in orthogonal coordinates, i.e., $g_{uv} \equiv 0$. In such coordinates

$$\begin{aligned} g^{uv} &= 0, \\ g^{uu} &= (g_{uu})^{-1}, \\ g^{vv} &= (g_{vv})^{-1}, \end{aligned}$$

$$\begin{aligned} \Gamma_{uuu} &= \frac{1}{2}\frac{\partial g_{uu}}{\partial u}, \\ \Gamma_{uvu} &= \frac{1}{2}\frac{\partial g_{uu}}{\partial v} = \Gamma_{vuu}, \\ \Gamma_{vvv} &= \frac{1}{2}\frac{\partial g_{vv}}{\partial v}, \\ \Gamma_{uvv} &= \frac{1}{2}\frac{\partial g_{vv}}{\partial u} = \Gamma_{vuv}, \\ \Gamma_{uuv} &= -\frac{1}{2}\frac{\partial g_{uu}}{\partial v}, \\ \Gamma_{vvu} &= -\frac{1}{2}\frac{\partial g_{vv}}{\partial u}, \end{aligned}$$

and

$$\begin{aligned} \Gamma_{uu}^{u} &= \frac{1}{2}g^{uu}\frac{\partial g_{uu}}{\partial u} = \frac{1}{2}\frac{\partial \ln g_{uu}}{\partial u}, \\ \Gamma_{uu}^{v} &= -\frac{1}{2}g^{vv}\frac{\partial g_{uu}}{\partial v}, \\ \Gamma_{vv}^{v} &= \frac{1}{2}g^{vv}\frac{\partial g_{vv}}{\partial v} = \frac{1}{2}\frac{\partial \ln g_{vv}}{\partial v}, \\ \Gamma_{vv}^{u} &= -\frac{1}{2}g^{uu}\frac{\partial g_{vv}}{\partial u}, \\ \Gamma_{uv}^{u} &= \frac{1}{2}g^{uu}\frac{\partial g_{uu}}{\partial v} = \frac{1}{2}\frac{\partial \ln g_{uu}}{\partial v}, \\ \Gamma_{uv}^{v} &= \frac{1}{2}g^{vv}\frac{\partial g_{vv}}{\partial u} = \frac{1}{2}\frac{\partial \ln g_{vv}}{\partial u}. \end{aligned}$$

Often there might be even more specific information. This could be that the metric coefficients only depend on one of the parameters, or that $g_{uu} = 1$. In such circumstances it is quite manageable to calculate the Christoffel symbols. What is more, it is always possible to find parametrizations where $g_{uu} \equiv 1$ and $g_{uv} \equiv 0$ as we shall see.

In a related vein we use our knowledge of Christoffel symbols to prove Gauss' amazing discovery that the Gauss curvature can be computed knowing only the first fundamental form. Given the definition of K this is certainly a big surprise. A different proof that uses our abstract framework will be given in section 6.4. Here we use a more direct approach.

THEOREM 6.1.3. *(Theorema Egregium, Gauss 1827) The Gauss curvature can be computed knowing only the first fundamental form.*

PROOF. We start with the observation that

$$
\begin{aligned}
K &= \det L = \det[\mathrm{I}]^{-1}\det[\mathrm{II}],\\
\det[\mathrm{I}] &= g_{uu}g_{vv} - (g_{uv})^2.
\end{aligned}
$$

So we concentrate on

$$
\begin{aligned}
\det[\mathrm{II}] &= \det\begin{bmatrix} L_{uu} & L_{uv} \\ L_{vu} & L_{vv}\end{bmatrix}\\
&= \det\begin{bmatrix} \frac{\partial^2\mathbf{q}}{\partial u^2}\cdot\mathbf{N} & \frac{\partial^2\mathbf{q}}{\partial u\partial v}\cdot\mathbf{N} \\ \frac{\partial^2\mathbf{q}}{\partial v\partial u}\cdot\mathbf{N} & \frac{\partial^2\mathbf{q}}{\partial v^2}\cdot\mathbf{N}\end{bmatrix}\\
&= \frac{1}{g_{uu}g_{vv}-(g_{uv})^2}\det\begin{bmatrix} \frac{\partial^2\mathbf{q}}{\partial u^2}\cdot\left(\frac{\partial\mathbf{q}}{\partial u}\times\frac{\partial\mathbf{q}}{\partial v}\right) & \frac{\partial^2\mathbf{q}}{\partial u\partial v}\cdot\left(\frac{\partial\mathbf{q}}{\partial u}\times\frac{\partial\mathbf{q}}{\partial v}\right) \\ \frac{\partial^2\mathbf{q}}{\partial v\partial u}\cdot\left(\frac{\partial\mathbf{q}}{\partial u}\times\frac{\partial\mathbf{q}}{\partial v}\right) & \frac{\partial^2\mathbf{q}}{\partial v^2}\cdot\left(\frac{\partial\mathbf{q}}{\partial u}\times\frac{\partial\mathbf{q}}{\partial v}\right)\end{bmatrix}.
\end{aligned}
$$

This allows us to consider

$$
\det\begin{bmatrix} \frac{\partial^2\mathbf{q}}{\partial u^2}\cdot\left(\frac{\partial\mathbf{q}}{\partial u}\times\frac{\partial\mathbf{q}}{\partial v}\right) & \frac{\partial^2\mathbf{q}}{\partial u\partial v}\cdot\left(\frac{\partial\mathbf{q}}{\partial u}\times\frac{\partial\mathbf{q}}{\partial v}\right) \\ \frac{\partial^2\mathbf{q}}{\partial v\partial u}\cdot\left(\frac{\partial\mathbf{q}}{\partial u}\times\frac{\partial\mathbf{q}}{\partial v}\right) & \frac{\partial^2\mathbf{q}}{\partial v^2}\cdot\left(\frac{\partial\mathbf{q}}{\partial u}\times\frac{\partial\mathbf{q}}{\partial v}\right)\end{bmatrix}.
$$

Here each entry in the matrix is a triple product and hence a determinant of a 3×3 matrix:

$$
\det\begin{bmatrix} \frac{\partial^2\mathbf{q}}{\partial w_1\partial w_2} & \frac{\partial\mathbf{q}}{\partial u} & \frac{\partial\mathbf{q}}{\partial v}\end{bmatrix} = \frac{\partial^2\mathbf{q}}{\partial w_1\partial w_2}\cdot\left(\frac{\partial\mathbf{q}}{\partial u}\times\frac{\partial\mathbf{q}}{\partial v}\right).
$$

With this observation and the fact that a matrix and its transpose have the same determinant we can calculate the products that appear in our 2×2 determinant

$$
\begin{aligned}
&\left(\frac{\partial^2\mathbf{q}}{\partial u^2}\cdot\left(\frac{\partial\mathbf{q}}{\partial u}\times\frac{\partial\mathbf{q}}{\partial v}\right)\right)\left(\frac{\partial^2\mathbf{q}}{\partial v^2}\cdot\left(\frac{\partial\mathbf{q}}{\partial u}\times\frac{\partial\mathbf{q}}{\partial v}\right)\right)\\
&= \det\begin{bmatrix} \frac{\partial^2\mathbf{q}}{\partial u^2} & \frac{\partial\mathbf{q}}{\partial u} & \frac{\partial\mathbf{q}}{\partial v}\end{bmatrix}\det\begin{bmatrix} \frac{\partial^2\mathbf{q}}{\partial v^2} & \frac{\partial\mathbf{q}}{\partial u} & \frac{\partial\mathbf{q}}{\partial v}\end{bmatrix}\\
&= \det\begin{bmatrix} \frac{\partial^2\mathbf{q}}{\partial u^2} & \frac{\partial\mathbf{q}}{\partial u} & \frac{\partial\mathbf{q}}{\partial v}\end{bmatrix}^t\det\begin{bmatrix} \frac{\partial^2\mathbf{q}}{\partial v^2} & \frac{\partial\mathbf{q}}{\partial u} & \frac{\partial\mathbf{q}}{\partial v}\end{bmatrix}\\
&= \det\left(\begin{bmatrix} \frac{\partial^2\mathbf{q}}{\partial u^2} & \frac{\partial\mathbf{q}}{\partial u} & \frac{\partial\mathbf{q}}{\partial v}\end{bmatrix}^t\begin{bmatrix} \frac{\partial^2\mathbf{q}}{\partial v^2} & \frac{\partial\mathbf{q}}{\partial u} & \frac{\partial\mathbf{q}}{\partial v}\end{bmatrix}\right)\\
&= \det\begin{bmatrix} \frac{\partial^2\mathbf{q}}{\partial u^2}\cdot\frac{\partial^2\mathbf{q}}{\partial v^2} & \frac{\partial\mathbf{q}}{\partial u}\cdot\frac{\partial^2\mathbf{q}}{\partial v^2} & \frac{\partial\mathbf{q}}{\partial v}\cdot\frac{\partial^2\mathbf{q}}{\partial v^2} \\ \frac{\partial^2\mathbf{q}}{\partial u^2}\cdot\frac{\partial\mathbf{q}}{\partial u} & \frac{\partial\mathbf{q}}{\partial u}\cdot\frac{\partial\mathbf{q}}{\partial u} & \frac{\partial\mathbf{q}}{\partial v}\cdot\frac{\partial\mathbf{q}}{\partial v} \\ \frac{\partial^2\mathbf{q}}{\partial u^2}\cdot\frac{\partial\mathbf{q}}{\partial v} & \frac{\partial\mathbf{q}}{\partial u}\cdot\frac{\partial\mathbf{q}}{\partial v} & \frac{\partial\mathbf{q}}{\partial v}\cdot\frac{\partial\mathbf{q}}{\partial v}\end{bmatrix}\\
&= \det\begin{bmatrix} \frac{\partial^2\mathbf{q}}{\partial u^2}\cdot\frac{\partial^2\mathbf{q}}{\partial v^2} & \Gamma_{vvu} & \Gamma_{vvv} \\ \Gamma_{uuu} & g_{uu} & g_{uv} \\ \Gamma_{uuv} & g_{vu} & g_{vv}\end{bmatrix}\\
&= \frac{\partial^2\mathbf{q}}{\partial u^2}\cdot\frac{\partial^2\mathbf{q}}{\partial v^2}\det[\mathrm{I}] + \det\begin{bmatrix} 0 & \Gamma_{vvu} & \Gamma_{vvv} \\ \Gamma_{uuu} & g_{uu} & g_{uv} \\ \Gamma_{uuv} & g_{vu} & g_{vv}\end{bmatrix}
\end{aligned}
$$

and similarly

$$
\begin{aligned}
&\left(\frac{\partial^2\mathbf{q}}{\partial u\partial v}\cdot\left(\frac{\partial\mathbf{q}}{\partial u}\times\frac{\partial\mathbf{q}}{\partial v}\right)\right)\left(\frac{\partial^2\mathbf{q}}{\partial v\partial u}\cdot\left(\frac{\partial\mathbf{q}}{\partial u}\times\frac{\partial\mathbf{q}}{\partial v}\right)\right)\\
&= \det\begin{bmatrix} \frac{\partial^2\mathbf{q}}{\partial u\partial v}\cdot\frac{\partial^2\mathbf{q}}{\partial u\partial v} & \frac{\partial\mathbf{q}}{\partial u}\cdot\frac{\partial^2\mathbf{q}}{\partial u\partial v} & \frac{\partial\mathbf{q}}{\partial v}\cdot\frac{\partial^2\mathbf{q}}{\partial u\partial v}\\ \frac{\partial^2\mathbf{q}}{\partial u\partial v}\cdot\frac{\partial\mathbf{q}}{\partial u} & \frac{\partial\mathbf{q}}{\partial u}\cdot\frac{\partial\mathbf{q}}{\partial u} & \frac{\partial\mathbf{q}}{\partial v}\cdot\frac{\partial\mathbf{q}}{\partial v}\\ \frac{\partial^2\mathbf{q}}{\partial u\partial v}\cdot\frac{\partial\mathbf{q}}{\partial v} & \frac{\partial\mathbf{q}}{\partial u}\cdot\frac{\partial\mathbf{q}}{\partial v} & \frac{\partial\mathbf{q}}{\partial v}\cdot\frac{\partial\mathbf{q}}{\partial v}\end{bmatrix}\\
&= \det\begin{bmatrix} \frac{\partial^2\mathbf{q}}{\partial u\partial v}\cdot\frac{\partial^2\mathbf{q}}{\partial u\partial v} & \Gamma_{uvu} & \Gamma_{uvv}\\ \Gamma_{uvu} & g_{uu} & g_{uv}\\ \Gamma_{uvv} & g_{vu} & g_{vv}\end{bmatrix}\\
&= \frac{\partial^2\mathbf{q}}{\partial u\partial v}\cdot\frac{\partial^2\mathbf{q}}{\partial u\partial v}\det[\mathrm{I}]+\det\begin{bmatrix} 0 & \Gamma_{uvu} & \Gamma_{uvv}\\ \Gamma_{uvu} & g_{uu} & g_{uv}\\ \Gamma_{uvv} & g_{vu} & g_{vv}\end{bmatrix}.
\end{aligned}
$$

We need to subtract these quantities but now only need to check the difference

$$
\begin{aligned}
&\frac{\partial^2\mathbf{q}}{\partial u^2}\cdot\frac{\partial^2\mathbf{q}}{\partial v^2}-\frac{\partial^2\mathbf{q}}{\partial u\partial v}\cdot\frac{\partial^2\mathbf{q}}{\partial u\partial v}\\
&= \frac{\partial}{\partial v}\left(\frac{\partial^2\mathbf{q}}{\partial u^2}\cdot\frac{\partial\mathbf{q}}{\partial v}\right)-\frac{\partial^3\mathbf{q}}{\partial v\partial u^2}\cdot\frac{\partial\mathbf{q}}{\partial v}\\
&\quad -\frac{\partial}{\partial u}\left(\frac{\partial^2\mathbf{q}}{\partial u\partial v}\cdot\frac{\partial\mathbf{q}}{\partial v}\right)+\frac{\partial^3\mathbf{q}}{\partial^2 u\partial v}\cdot\frac{\partial\mathbf{q}}{\partial v}\\
&= \frac{\partial}{\partial v}\Gamma_{uuv}-\frac{\partial}{\partial u}\Gamma_{uvv}
\end{aligned}
$$

or equivalently

$$
\begin{aligned}
&\frac{\partial^2\mathbf{q}}{\partial u^2}\cdot\frac{\partial^2\mathbf{q}}{\partial v^2}-\frac{\partial^2\mathbf{q}}{\partial u\partial v}\cdot\frac{\partial^2\mathbf{q}}{\partial u\partial v}\\
&= \frac{\partial}{\partial u}\left(\frac{\partial^2\mathbf{q}}{\partial v^2}\cdot\frac{\partial\mathbf{q}}{\partial u}\right)-\frac{\partial^3\mathbf{q}}{\partial u\partial v^2}\cdot\frac{\partial\mathbf{q}}{\partial u}\\
&\quad -\frac{\partial}{\partial v}\left(\frac{\partial^2\mathbf{q}}{\partial u\partial v}\cdot\frac{\partial\mathbf{q}}{\partial u}\right)+\frac{\partial^3\mathbf{q}}{\partial^2 v\partial u}\cdot\frac{\partial\mathbf{q}}{\partial u}\\
&= \frac{\partial}{\partial u}\Gamma_{vvu}-\frac{\partial}{\partial v}\Gamma_{uvu}.
\end{aligned}
$$

Thus

$$
\begin{aligned}
K &= \frac{\det\begin{bmatrix} \frac{\partial^2\mathbf{q}}{\partial u^2}\cdot\left(\frac{\partial\mathbf{q}}{\partial u}\times\frac{\partial\mathbf{q}}{\partial v}\right) & \frac{\partial^2\mathbf{q}}{\partial u\partial v}\cdot\left(\frac{\partial\mathbf{q}}{\partial u}\times\frac{\partial\mathbf{q}}{\partial v}\right)\\ \frac{\partial^2\mathbf{q}}{\partial v\partial u}\cdot\left(\frac{\partial\mathbf{q}}{\partial u}\times\frac{\partial\mathbf{q}}{\partial v}\right) & \frac{\partial^2\mathbf{q}}{\partial v^2}\cdot\left(\frac{\partial\mathbf{q}}{\partial u}\times\frac{\partial\mathbf{q}}{\partial v}\right)\end{bmatrix}}{(\det[\mathrm{I}])^2}\\
&= \frac{\left(\frac{\partial}{\partial v}\Gamma_{uuv}-\frac{\partial}{\partial u}\Gamma_{uvv}\right)}{\det[\mathrm{I}]}\\
&\quad +\frac{\det\begin{bmatrix} 0 & \Gamma_{vvu} & \Gamma_{vvv}\\ \Gamma_{uuu} & g_{uu} & g_{uv}\\ \Gamma_{uuv} & g_{vu} & g_{vv}\end{bmatrix}-\det\begin{bmatrix} 0 & \Gamma_{uvu} & \Gamma_{uvv}\\ \Gamma_{uvu} & g_{uu} & g_{uv}\\ \Gamma_{uvv} & g_{vu} & g_{vv}\end{bmatrix}}{(\det[\mathrm{I}])^2}.
\end{aligned}
$$

□

COROLLARY 6.1.4. *(Gauss, 1827) If a surface admits a Cartesian parametrization around every point, then the Gauss curvature vanishes.*

REMARK 6.1.5. The converse is also true and partly follows from Monge's classification of flat surfaces (see section 5.5). A more abstract proof will be given later. The proof will also have the advantage of working for the generalized and abstract surfaces that we discuss in the next section.

Exercises. In the following exercises calculate the Christoffel symbols of both kinds as well as the Gauss curvature. It is implicitly assumed that the functions and constants are chosen so as to represent the first fundamental form of a parametrized surface $\mathbf{q}(u, v)$. In the cases where you actually know the function $\mathbf{q}(u, v)$ try to compute these quantities first using that knowledge and then afterwards only using the first fundamental form given in the problem.

(1) Show directly from the formulas for the Christoffel symbols in terms of the first fundamental form that
$$\frac{\partial}{\partial v}\Gamma_{uuv} - \frac{\partial}{\partial u}\Gamma_{uvv} = \frac{\partial}{\partial u}\Gamma_{vvu} - \frac{\partial}{\partial v}\Gamma_{uvu}$$
and
$$\frac{\partial}{\partial v}\Gamma_{uuv} - \frac{\partial}{\partial u}\Gamma_{uvv} = -\frac{1}{2}\frac{\partial^2 g_{uu}}{\partial v^2} + \frac{\partial^2 g_{uv}}{\partial u \partial v} - \frac{1}{2}\frac{\partial^2 g_{vv}}{\partial u^2}.$$

(2) (Oblique Cartesian coordinates) The first fundamental form is given by
$$[\mathrm{I}] = \begin{bmatrix} a & b \\ b & d \end{bmatrix}$$
where a, b, d are constants with $a, d > 0$ and $ad > b^2$.

(3) (Surface of revolution) The first fundamental form is given by
$$[\mathrm{I}] = \begin{bmatrix} 1 & 0 \\ 0 & r^2 \end{bmatrix}$$
where $r = r(u) > 0$.

(4) (Polar and Fermi coordinates) The first fundamental form is given by
$$[\mathrm{I}] = \begin{bmatrix} 1 & 0 \\ 0 & r^2 \end{bmatrix}$$
where $r = r(u, v) > 0$. Gauss showed that such coordinates exist around any point in a surface with r denoting the "intrinsic" distance to the point. Fermi created such coordinates in a neighborhood of a geodesic with r denoting the "intrinsic" distance to the geodesic. The terminology will be explained later.

(5) The first fundamental form is given by
$$[\mathrm{I}] = \begin{bmatrix} r^2 & 0 \\ 0 & r^2 \end{bmatrix}$$
where $r = r(u) > 0$.

(6) (Isothermal coordinates) The first fundamental form is given by
$$[\mathrm{I}] = \begin{bmatrix} r^2 & 0 \\ 0 & r^2 \end{bmatrix}$$
where $r = r(u, v) > 0$.

(7) (Liouville surfaces) The first fundamental form is given by

$$[\mathrm{I}] = \begin{bmatrix} r^2 & 0 \\ 0 & r^2 \end{bmatrix}$$

where $r^2 = f(u) + g(v) > 0$.

(8) (Monge patch) The first fundamental form is given by

$$[\mathrm{I}] = \begin{bmatrix} 1+f^2 & fg \\ fg & 1+g^2 \end{bmatrix}$$

where $f = \frac{\partial F}{\partial u}, g = \frac{\partial F}{\partial v}$ and $F = F(u,v)$.

(9) (Gauss) Show that if we define $|g|^2 = \det[\mathrm{I}]$, then

$$\begin{aligned}
4|g|^4 K &= g_{uu}\left(\frac{\partial g_{uu}}{\partial v}\frac{\partial g_{vv}}{\partial v} - 2\frac{\partial g_{uv}}{\partial u}\frac{\partial g_{vv}}{\partial v} + \left(\frac{\partial g_{vv}}{\partial u}\right)^2\right) \\
&\quad + g_{uv}\left(\frac{\partial g_{uu}}{\partial u}\frac{\partial g_{vv}}{\partial v} - \frac{\partial g_{vv}}{\partial u}\frac{\partial g_{uu}}{\partial v} - 2\frac{\partial g_{uu}}{\partial v}\frac{\partial g_{uv}}{\partial v} - 2\frac{\partial g_{vv}}{\partial u}\frac{\partial g_{uv}}{\partial u} + 4\frac{\partial g_{uv}}{\partial u}\frac{\partial g_{uv}}{\partial v}\right) \\
&\quad + g_{vv}\left(\frac{\partial g_{uu}}{\partial u}\frac{\partial g_{vv}}{\partial u} - 2\frac{\partial g_{uu}}{\partial u}\frac{\partial g_{uv}}{\partial v} + \left(\frac{\partial g_{uu}}{\partial v}\right)^2\right) \\
&\quad - 2|g|^2\left(\frac{\partial^2 g_{uu}}{\partial v^2} - 2\frac{\partial^2 g_{uv}}{\partial u \partial v} + \frac{\partial^2 g_{vv}}{\partial u^2}\right).
\end{aligned}$$

(10) (Frobenius) Show that if we define $|g|^2 = \det[\mathrm{I}]$, then

$$\begin{aligned}
K &= -\frac{1}{4|g|^2}\det\begin{bmatrix} g_{uu} & g_{uv} & g_{vv} \\ \frac{\partial g_{uu}}{\partial u} & \frac{\partial g_{uv}}{\partial u} & \frac{\partial g_{vv}}{\partial u} \\ \frac{\partial g_{uu}}{\partial v} & \frac{\partial g_{uv}}{\partial v} & \frac{\partial g_{vv}}{\partial v} \end{bmatrix} \\
&\quad - \frac{1}{2|g|}\left(\frac{\partial}{\partial u}\left(\frac{\frac{\partial g_{vv}}{\partial u} - \frac{\partial g_{uv}}{\partial v}}{|g|}\right) + \frac{\partial}{\partial v}\left(\frac{\frac{\partial g_{uu}}{\partial v} - \frac{\partial g_{uv}}{\partial u}}{|g|}\right)\right).
\end{aligned}$$

(11) (Liouville) Show that if we define $|g|^2 = \det[\mathrm{I}]$, then

$$\begin{aligned}
K &= \frac{1}{|g|}\left(\frac{\partial}{\partial v}\left(\frac{|g|}{g_{uu}}\Gamma_{uu}^{v}\right) - \frac{\partial}{\partial u}\left(\frac{|g|}{g_{uu}}\Gamma_{uv}^{v}\right)\right) \\
&= \frac{1}{|g|}\left(\frac{\partial}{\partial v}\left(\frac{|g|}{g_{vv}}\Gamma_{vv}^{u}\right) + \frac{\partial}{\partial u}\left(\frac{|g|}{g_{vv}}\Gamma_{uv}^{u}\right)\right).
\end{aligned}$$

6.2. Generalized and Abstract Surfaces

It is possible to work with *generalized surfaces* in Euclidean spaces of arbitrary dimension: $\mathbf{q}(u,v) : U \to \mathbb{R}^k$ for any $k \geq 2$. What changes is that we no longer have a single normal vector $\mathbf{N}$. In fact for $k \geq 4$ there will be a whole family of normal vectors, not unlike what happened for space curves. What all of these surfaces do have in common is that we can define the first fundamental form. Thus we can also calculate the Christoffel symbols of the first and second kind using the formulas in terms of derivatives of g. This leads us to the possibility of an abstract definition of a surface that is independent of a particular map into some coordinate space $\mathbb{R}^k$.

One of the simplest examples of a generalized surface is the flat torus in $\mathbb{R}^4$. It is parametrized by

$$\mathbf{q}(u,v) = (\cos u, \sin u, \cos v, \sin v)$$

and its first fundamental form is

$$\mathrm{I} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

So this yields a Cartesian parametrization of the entire torus. This is why it is called the flat torus. It is in fact not possible to have a flat torus in $\mathbb{R}^3$.

An *abstract parametrized surface* consists of a domain $U \subset \mathbb{R}^2$ and a first fundamental form

$$[\mathrm{I}] = \begin{bmatrix} g_{uu} & g_{uv} \\ g_{uv} & g_{vv} \end{bmatrix},$$

where g_{uu}, g_{vv}, and g_{uv} are functions on U. The inner product of vectors $X = (X^u, X^v)$ and $Y = (Y^u, Y^v)$ thought of as having the same base point $p \in U$ is defined as

$$\mathrm{I}(X,Y) = \begin{bmatrix} X^u & X^v \end{bmatrix} \begin{bmatrix} g_{uu}(p) & g_{uv}(p) \\ g_{vu}(p) & g_{vv}(p) \end{bmatrix} \begin{bmatrix} Y^u \\ Y^v \end{bmatrix}.$$

For this to give us an inner product we also have to make sure that it is positive definite:

$$\begin{aligned} 0 &< \mathrm{I}(X,X) \\ &= \begin{bmatrix} X^u & X^v \end{bmatrix} \begin{bmatrix} g_{uu} & g_{uv} \\ g_{uv} & g_{vv} \end{bmatrix} \begin{bmatrix} X^u \\ X^v \end{bmatrix} \\ &= X^u X^u g_{uu} + 2X^u X^v g_{uv} + X^v X^v g_{vv}. \end{aligned}$$

PROPOSITION 6.2.1. I *is positive definite if and only if* $g_{uu} + g_{vv}$ *and* $g_{uu}g_{vv} - (g_{uv})^2$ *are positive.*

PROOF. If I is positive definite, then it follows that g_{uu} and g_{vv} are positive by letting $X = (1,0)$ and $(0,1)$. Next select $X = \left(\sqrt{g_{vv}}, \pm\sqrt{g_{uu}}\right)$ to get

$$0 < \mathrm{I}(X,X) = 2g_{uu}g_{vv} \pm 2\sqrt{g_{uu}}\sqrt{g_{vv}}g_{uv}.$$

Thus

$$\pm g_{uv} < \sqrt{g_{uu}}\sqrt{g_{vv}}$$

showing that

$$g_{uu}g_{vv} > (g_{uv})^2.$$

To check that I is positive definite when $g_{uu} + g_{vv}$, and $g_{uu}g_{vv} - (g_{uv})^2$ are positive we start by observing that

$$g_{uu}g_{vv} > g_{uv}^2 \geq 0.$$

Thus g_{uu} and g_{vv} have the same sign. As their sum is positive both terms are positive. It then follows that

$$\begin{aligned} \mathrm{I}(X,X) &= X^u X^u g_{uu} + 2X^u X^v g_{uv} + X^v X^v g_{vv} \\ &\geq X^u X^u g_{uu} - 2|X^u||X^v|\sqrt{g_{uu}g_{vv}} + X^v X^v g_{vv} \\ &= \left(|X^u|\sqrt{g_{uu}} - |X^v|\sqrt{g_{vv}}\right)^2 \\ &\geq 0. \end{aligned}$$

Here is first inequality is in fact $>$ unless $X^u = 0$ or $X^v = 0$. In case, say, $X^u = 0$ we obtain

$$\mathrm{I}(X,X) = (X^v)^2 g_{vv} > 0$$

unless also $X^v = 0$. □

EXAMPLE 6.2.2. The hyperbolic space $H \subset \mathbb{R}^{2,1}$ is defined as the imaginary unit sphere with $z > 0$, specifically it is the rotationally symmetric surface

$$x^2 + y^2 - z^2 = -1, \ z \geq 1$$

or equivalently the Monge patch

$$z = \sqrt{1 + x^2 + y^2}.$$

The metric on this surface, however, is inherited from a different inner product structure on $\mathbb{R}^3$ which is why we use the notation $\mathbb{R}^{2,1}$. Specifically:

$$X \cdot Y = X^x Y^x + X^y Y^y - X^z Y^z.$$

The x- and y-coordinates are the "space" part and the z-coordinate the "time" part. We say that a vector is space-like, null, or time-like if $|X|^2 = X \cdot X$ is positive, zero, or negative. Thus $(x, y, 0)$ is space-like while $(0, 0, z)$ is time-like. Null vectors satisfy the equation

$$X^x X^x + X^y X^y - X^z X^z = 0.$$

This describes a cone. The two insides of this cone consists of the time-like vectors, while the outside contains the space-like vectors.

Our surface H given by the equation

$$F(x, y, z) = x^2 + y^2 - z^2 = -1, \ z \geq 1$$

therefore consists of time-like points. However, all of the tangent spaces turn out to consist of time-like vectors. This means that we obtain a surface with a valid first fundamental form. In the Monge patch representation we have

$$\frac{\partial z}{\partial x} = \frac{x}{\sqrt{1 + x^2 + y^2}} = \frac{x}{z}, \ \frac{\partial z}{\partial y} = \frac{y}{\sqrt{1 + x^2 + y^2}} = \frac{y}{z}.$$

Thus the tangent space at $q = (x, y, z) = \left(x, y, \sqrt{1 + x^2 + y^2}\right)$ is given by

$$\begin{aligned}
T_q H &= \operatorname{span}\left\{\left(1, 0, \frac{x}{z}\right), \left(1, 0, \frac{y}{z}\right)\right\} \\
&= \left\{X^x \left(1, 0, \frac{x}{z}\right) + X^y \left(0, 1, \frac{y}{z}\right) \mid X^x, X^y \in \mathbb{R}\right\}
\end{aligned}$$

and consequently

$$\begin{aligned}
X \cdot X &= (X^x)^2 + (X^y)^2 - \left(X^x \frac{x}{z} + X^y \frac{y}{z}\right)^2 \\
&= (X^x)^2 \left(1 - \frac{x^2}{z^2}\right) + (X^y)^2 \left(1 - \frac{y^2}{z^2}\right) \\
&\quad -2X^x X^y \frac{xy}{z^2} \\
&= (X^x)^2 \frac{1 + y^2}{z^2} + (X^y)^2 \frac{1 + x^2}{z^2} - 2X^x X^y \frac{xy}{z^2} \\
&= \frac{1}{z^2} \left((X^x)^2 + (X^y)^2 + (yX^x - xX^y)^2\right).
\end{aligned}$$

This is clearly positive unless $X = 0$. The first fundamental form is

$$\begin{bmatrix} 1 - \frac{x^2}{z^2} & -\frac{xy}{z^2} \\ -\frac{xy}{z^2} & 1 - \frac{y^2}{z^2} \end{bmatrix}$$

which is also easily checked to be positive using proposition 6.2.1.

In order to find a nicer expression of the first fundamental form we switch to a surface of revolution parametrization

$$\mathbf{q}(\phi,\mu)=\begin{bmatrix}\cos\mu\sinh\phi\\ \sin\mu\sinh\phi\\ \cosh\phi\end{bmatrix},\ \mu\in\mathbb{R},\ \phi>0,$$

where $\phi=0$ corresponds to the point $(0,0,1)$ which we can think of as a pole. In this parametrization we obtain

$$\frac{\partial\mathbf{q}}{\partial\phi}=\begin{bmatrix}\cos\mu\cosh\phi\\ \sin\mu\cosh\phi\\ \sinh\phi\end{bmatrix},\ \frac{\partial\mathbf{q}}{\partial\mu}=\begin{bmatrix}\sin\mu\sinh\phi\\ -\cos\mu\sinh\phi\\ 0\end{bmatrix}$$

which gives us the first fundamental form

$$\begin{bmatrix}\frac{\partial\mathbf{q}}{\partial\phi}\cdot\frac{\partial\mathbf{q}}{\partial\phi} & \frac{\partial\mathbf{q}}{\partial\phi}\cdot\frac{\partial\mathbf{q}}{\partial\mu}\\ \frac{\partial\mathbf{q}}{\partial\mu}\cdot\frac{\partial\mathbf{q}}{\partial\phi} & \frac{\partial\mathbf{q}}{\partial\mu}\cdot\frac{\partial\mathbf{q}}{\partial\mu}\end{bmatrix}=\begin{bmatrix}1 & 0\\ 0 & \sinh^2\phi\end{bmatrix}.$$

Remark 6.2.3. It is not possible for a surface of revolution to have this first fundamental form in $\mathbb{R}^3$. But we shall see later that the pseudo-sphere from section 5.4 exercise 5 is a local Euclidean model that is locally isometric to H.

On the other hand a theorem of Hilbert (see theorem 6.4.6) shows that one cannot represent the entire surface H in $\mathbb{R}^3$, i.e., there is no parametrization $\mathbf{q}(x,y):H\to\mathbb{R}^3$ defined for all $(x,y)\in\mathbb{R}^2$ such that

$$\begin{bmatrix}\frac{\partial\mathbf{q}}{\partial x} & \frac{\partial\mathbf{q}}{\partial y}\end{bmatrix}^t\begin{bmatrix}\frac{\partial\mathbf{q}}{\partial x} & \frac{\partial\mathbf{q}}{\partial y}\end{bmatrix}=\begin{bmatrix}\frac{\partial\mathbf{q}}{\partial x}\cdot\frac{\partial\mathbf{q}}{\partial x} & \frac{\partial\mathbf{q}}{\partial x}\cdot\frac{\partial\mathbf{q}}{\partial y}\\ \frac{\partial\mathbf{q}}{\partial y}\cdot\frac{\partial\mathbf{q}}{\partial x} & \frac{\partial\mathbf{q}}{\partial y}\cdot\frac{\partial\mathbf{q}}{\partial y}\end{bmatrix}=\begin{bmatrix}1-\frac{x^2}{z^2} & -\frac{xy}{z^2}\\ -\frac{xy}{z^2} & 1-\frac{y^2}{z^2}\end{bmatrix}.$$

Janet-Burstin-Cartan showed that if the metric coefficients of an abstract surface are analytic, then one can always locally represent the abstract surface in $\mathbb{R}^3$. Nash showed that any abstract surface can be represented by a map $\mathbf{q}(u,v):U\to\mathbb{R}^k$ on the entire domain, but only at the expense of making k very large. Based in part on Nash's work Greene and Gromov both showed that one can always locally represent an abstract surface in $\mathbb{R}^5$. An even more recent development by M. Khuri in 2007 is that it is in fact possible to find an abstract surface that cannot be locally realized as a surface in $\mathbb{R}^3$.

Definition 6.2.4. We say that a surface $M\subset\mathbb{R}^{2,1}$ is space-like if all tangent vectors are space-like. This means that if we use the first fundamental form that comes from the inner product in $\mathbb{R}^{2,1}$, then we obtain an abstract surface.

Finally we also have to define what we mean by an abstract surface. There are many competing definitions. The more general and abstract ones unfortunately also have a very steep learning curve before a metric can be introduced so we stay with the more classical context. Essentially we define a surface, as was done classically, as a set of points where we can use the language of first fundamental form, convergence etc. This is generally too vague for modern mathematicians but at least allows us to move on to the issues that are relevant in differential geometry. There are several other standard concepts included in this definition so as to have them all in one place

Definition 6.2.5. A *surface* with a first fundamental form is a space M where we can locally work as if it is an abstract parametrized surface, i.e., every point is included in a parametrization $\mathbf{q} : U \subset \mathbb{R}^2 \to M$. When a point $q \in M$ is covered by more than one parametrization, then they are pairwise reparametrizations of each other near q and the first fundamental forms are the same via this reparametrization. Globally we are allowed to talk about convergence of sequences as we do in $\mathbb{R}^2$. A sequence converges to q if eventually it lies in a parametrization around q and converges to q in that parametrization. Moreover, if the sequence eventually lies in more than one parametrization then its limit will be q in all of these parametrizations. This allows us to talk about continuous maps $F : M \to \mathbb{R}^k$ or into $F : \mathbb{R}^l \to M$. Such a map is smooth if it smooth within the given parametrizations. Finally we want the surface to be path connected in the sense that any two points are joined by a piecewise smooth curve.

A surface is said to be *closed* if it is compact, i.e., any sequence has a convergent subsequence.

A surface M is said to be *orientable* if all the parametrizations can be chosen so that the differential of all the reparametrizations has positive determinant, e.g., if $\mathbf{q}(u,v) = \mathbf{q}(u(s,t), v(s,t)) = \mathbf{q}(s,t)$, then

$$\det \begin{bmatrix} \frac{\partial u}{\partial s} & \frac{\partial u}{\partial t} \\ \frac{\partial v}{\partial s} & \frac{\partial v}{\partial t} \end{bmatrix} > 0, \ \det \begin{bmatrix} \frac{\partial s}{\partial u} & \frac{\partial s}{\partial v} \\ \frac{\partial t}{\partial u} & \frac{\partial t}{\partial v} \end{bmatrix} > 0.$$

Such a choice of parametrizations that cover all of M will be called an *orientation* for M. Note that this tells us that if we have tangent vectors $v, w \in T_pM$ that are not proportional, then w either lies to the right or left of v.

The tangent space T_qM at a point $q \in M$ in a parametrization is defined as $T_qM = \operatorname{span}\{\partial_u \mathbf{q}, \partial_v \mathbf{q}\}$. In a different parametrization the two bases are related by

$$\begin{aligned} \begin{bmatrix} \partial_u \mathbf{q} & \partial_v \mathbf{q} \end{bmatrix} &= \begin{bmatrix} \partial_s \mathbf{q} & \partial_t \mathbf{q} \end{bmatrix} \begin{bmatrix} \frac{\partial s}{\partial u} & \frac{\partial s}{\partial v} \\ \frac{\partial t}{\partial u} & \frac{\partial t}{\partial v} \end{bmatrix}, \\ \begin{bmatrix} \partial_s \mathbf{q} & \partial_t \mathbf{q} \end{bmatrix} &= \begin{bmatrix} \partial_u \mathbf{q} & \partial_v \mathbf{q} \end{bmatrix} \begin{bmatrix} \frac{\partial u}{\partial s} & \frac{\partial u}{\partial t} \\ \frac{\partial v}{\partial s} & \frac{\partial v}{\partial t} \end{bmatrix}. \end{aligned}$$

A tangent vector $X \in T_qM$ can thus be written

$$\begin{aligned} X &= X^u \partial_u \mathbf{q} + X^v \partial_v \mathbf{q} \\ &= \begin{bmatrix} \partial_u \mathbf{q} & \partial_v \mathbf{q} \end{bmatrix} \begin{bmatrix} X^u \\ X^v \end{bmatrix} \\ &= \begin{bmatrix} \partial_s \mathbf{q} & \partial_t \mathbf{q} \end{bmatrix} \begin{bmatrix} \frac{\partial s}{\partial u} & \frac{\partial s}{\partial v} \\ \frac{\partial t}{\partial u} & \frac{\partial t}{\partial v} \end{bmatrix} \begin{bmatrix} X^u \\ X^v \end{bmatrix} \\ &= \begin{bmatrix} \partial_s \mathbf{q} & \partial_t \mathbf{q} \end{bmatrix} \begin{bmatrix} X^u \frac{\partial s}{\partial u} + X^v \frac{\partial s}{\partial v} \\ X^u \frac{\partial t}{\partial u} + X^v \frac{\partial t}{\partial v} \end{bmatrix} \\ &= \left(X^u \frac{\partial s}{\partial u} + X^v \frac{\partial s}{\partial v} \right) \partial_s \mathbf{q} + \left(X^u \frac{\partial t}{\partial u} + X^v \frac{\partial t}{\partial v} \right) \partial_t \mathbf{q} \\ &= X^s \partial_s \mathbf{q} + X^t \partial_t \mathbf{q}. \end{aligned}$$

A surface is said to be *isometrically embedded* in $\mathbb{R}^3$ if it can be represented as a surface $M \subset \mathbb{R}^3$ in such a way that that the induced first fundamental form agrees with the abstract one on M. Specifically, we seek a map $F : M \to F(M) \subset \mathbb{R}^3$ such that F is a diffeomorphism from M to $F(M)$ and $\mathrm{I}_M(X,Y) = \mathrm{I}_{F(M)}(DF(X), DF(Y))$.

A surface is said to be *isometrically immersed* in $\mathbb{R}^3$ if there is a map $F : M \to \mathbb{R}^3$ such that $\mathrm{I}_M(X,Y) = \mathrm{I}_{F(M)}(DF(X), DF(Y))$. In this case F will be a local diffeomorphism onto its image, but globally it might not be one-to-one (see also 4.1.5).

We define the Christoffel symbols on abstract surfaces using the formulas

$$\begin{aligned}\Gamma_{ijk} &= \frac{1}{2}\left(\partial_i g_{jk} + \partial_j g_{ik} - \partial_k g_{ij}\right).\\ \Gamma_{ij}^k &= g^{kl}\Gamma_{ijl}.\end{aligned}$$

When defined this way we also obtain the formula

$$\partial_k g_{ij} = \Gamma_{kij} + \Gamma_{kji}$$

since

$$\partial_k g_{ij} = \frac{1}{2}\left(\partial_k g_{ij} + \partial_i g_{kj} - \partial_j g_{ki}\right) + \frac{1}{2}\left(\partial_k g_{ij} + \partial_j g_{ki} - \partial_i g_{kj}\right).$$

So this no longer follows from the product rule. Instead we prove this product rule from our definition of the Christoffel symbols.

Remark 6.2.6. In modern usage a surface does not necessarily come with a first fundamental form. We could have called our surfaces Riemannian surfaces (Riemannian manifolds are their higher dimensional analogues), but that too can be confused with Riemann surfaces which are surfaces where the reparametrizations are holomorphic, i.e., satisfy the Cauchy-Riemann equations.

Exercises.

(1) Show that the surfaces in $\mathbb{R}^{2,1}$ given by the equation

$$x^2 + y^2 - z^2 = -R^2$$

have constant Gauss curvature $-R^{-2}$.

(2) Assume that $g_{uu} = g_{vv} = 1$ on a domain $U \subset \mathbb{R}^2$. Show that the corresponding first fundamental form represents an abstract surface if $|g_{uv}| < 1$ on U.

(3) Assume the first fundamental form is given by the conditions in the previous exercise and that $g_{uv} = \cos\theta$, where $\theta : U \to \mathbb{R}$. Show that

$$\begin{aligned}\Gamma_{uvw} &= \Gamma_{uuu} = \Gamma_{vvv} = 0,\\ \Gamma_{uuv} &= -\frac{\partial\theta}{\partial u}\sin\theta,\\ \Gamma_{vvu} &= -\frac{\partial\theta}{\partial v}\sin\theta,\\ \frac{\partial^2\theta}{\partial u\partial v} &= -K\sin\theta.\end{aligned}$$

(4) Assume that a parametrized surface $\mathbf{q} : U \to \mathbb{R}^n$ has a first fundamental form where $g_{uu} = g_{vv} = 1$ on U. Show that $\frac{\partial^2\mathbf{q}}{\partial u\partial v}$ is perpendicular to the surface. Hint: Use the previous exercise.

6.3. Acceleration

The goal here is to show that the tangential component of the acceleration of a curve on a parametrized surface $\mathbf{q}(u,v): U \to \mathbb{R}^3$ can be calculated intrinsically. The curve is parametrized in U as $(u(t), v(t))$ and becomes a space curve $\mathbf{q}(t) = \mathbf{q}(u(t), v(t))$ that lies on our parametrized surface.

The velocity is

$$\dot{\mathbf{q}} = \frac{d\mathbf{q}}{dt} = \frac{d\mathbf{q}}{dt} = \frac{\partial \mathbf{q}}{\partial u}\frac{du}{dt} + \frac{\partial \mathbf{q}}{\partial v}\frac{dv}{dt} = \begin{bmatrix} \frac{\partial \mathbf{q}}{\partial u} & \frac{\partial \mathbf{q}}{\partial v} \end{bmatrix} \begin{bmatrix} \frac{du}{dt} \\ \frac{dv}{dt} \end{bmatrix}.$$

The acceleration can be calculated as if it were a space curve and we explored that in chapter 5. Using the velocity representation we just gave and separating the tangential and normal components of the acceleration we obtain:

$$\begin{aligned} \ddot{\mathbf{q}} &= \ddot{\mathbf{q}}^{\mathrm{I}} + \ddot{\mathbf{q}}^{\mathrm{II}} \\ &= \begin{bmatrix} \frac{\partial \mathbf{q}}{\partial u} & \frac{\partial \mathbf{q}}{\partial v} \end{bmatrix} [\mathrm{I}]^{-1} \begin{bmatrix} \frac{\partial \mathbf{q}}{\partial u} & \frac{\partial \mathbf{q}}{\partial v} \end{bmatrix}^t \ddot{\mathbf{q}} + (\ddot{\mathbf{q}} \cdot \mathbf{N})\,\mathbf{N}. \end{aligned}$$

In chapter 5 we focused on the normal component. Here we shall mostly concentrate on the tangential part.

THEOREM 6.3.1. *The acceleration can be calculated as*

$$\begin{aligned} \ddot{\mathbf{q}} &= \begin{bmatrix} \frac{\partial \mathbf{q}}{\partial u} & \frac{\partial \mathbf{q}}{\partial v} & \mathbf{N} \end{bmatrix} \begin{bmatrix} \frac{d^2u}{dt^2} + \Gamma^u(\dot{\mathbf{q}}, \dot{\mathbf{q}}) \\ \frac{d^2v}{dt^2} + \Gamma^v(\dot{\mathbf{q}}, \dot{\mathbf{q}}) \\ \mathrm{II}(\dot{\mathbf{q}}, \dot{\mathbf{q}}) \end{bmatrix} \\ &= \left(\frac{d^2u}{dt^2} + \Gamma^u(\dot{\mathbf{q}}, \dot{\mathbf{q}})\right)\frac{\partial \mathbf{q}}{\partial u} + \left(\frac{d^2v}{dt^2} + \Gamma^v(\dot{\mathbf{q}}, \dot{\mathbf{q}})\right)\frac{\partial \mathbf{q}}{\partial v} + \mathbf{N}\mathrm{II}(\dot{\mathbf{q}}, \dot{\mathbf{q}}), \end{aligned}$$

where

$$\Gamma^w(\dot{\mathbf{q}}, \dot{\mathbf{q}}) = \sum_{w_1, w_2 = u, v} \Gamma^w_{w_1 w_2}\frac{dw_1}{dt}\frac{dw_2}{dt} = \begin{bmatrix} \frac{du}{dt} & \frac{dv}{dt} \end{bmatrix} \begin{bmatrix} \Gamma^w_{uu} & \Gamma^w_{uv} \\ \Gamma^w_{vu} & \Gamma^w_{vv} \end{bmatrix} \begin{bmatrix} \frac{du}{dt} \\ \frac{dv}{dt} \end{bmatrix}.$$

PROOF. We start from the formula for the velocity and take derivatives. This clearly requires us to be able to calculate derivatives of the tangent fields $\frac{\partial \mathbf{q}}{\partial u}, \frac{\partial \mathbf{q}}{\partial v}$. Fortunately the Gauss formulas tell us how that is done. This leads us to the acceleration as follows

$$\begin{aligned} \ddot{\mathbf{q}} &= \frac{d}{dt}\left(\begin{bmatrix} \frac{\partial \mathbf{q}}{\partial u} & \frac{\partial \mathbf{q}}{\partial v} \end{bmatrix} \begin{bmatrix} \frac{du}{dt} \\ \frac{dv}{dt} \end{bmatrix}\right) \\ &= \begin{bmatrix} \frac{\partial \mathbf{q}}{\partial u} & \frac{\partial \mathbf{q}}{\partial v} \end{bmatrix} \begin{bmatrix} \frac{d^2u}{dt^2} \\ \frac{d^2v}{dt^2} \end{bmatrix} + \left(\frac{d}{dt}\begin{bmatrix} \frac{\partial \mathbf{q}}{\partial u} & \frac{\partial \mathbf{q}}{\partial v} \end{bmatrix}\right) \begin{bmatrix} \frac{du}{dt} \\ \frac{dv}{dt} \end{bmatrix} \end{aligned}$$

which after using the chain rule

$$\frac{d}{dt} = \frac{du}{dt}\frac{\partial}{\partial u} + \frac{dv}{dt}\frac{\partial}{\partial v}$$

becomes

$$\begin{aligned}\ddot{\mathbf{q}} &= \begin{bmatrix} \frac{\partial \mathbf{q}}{\partial u} & \frac{\partial \mathbf{q}}{\partial v} \end{bmatrix} \begin{bmatrix} \frac{d^2u}{dt^2} \\ \frac{d^2v}{dt^2} \end{bmatrix} \\ &\quad + \frac{du}{dt}\left(\frac{\partial}{\partial u}\begin{bmatrix} \frac{\partial \mathbf{q}}{\partial u} & \frac{\partial \mathbf{q}}{\partial v} \end{bmatrix}\right)\begin{bmatrix} \frac{du}{dt} \\ \frac{dv}{dt} \end{bmatrix} \\ &\quad + \frac{dv}{dt}\left(\frac{\partial}{\partial v}\begin{bmatrix} \frac{\partial \mathbf{q}}{\partial u} & \frac{\partial \mathbf{q}}{\partial v} \end{bmatrix}\right)\begin{bmatrix} \frac{du}{dt} \\ \frac{dv}{dt} \end{bmatrix}.\end{aligned}$$

The Gauss formulas help us with the last two terms

$$\begin{aligned}\left(\frac{\partial}{\partial w}\begin{bmatrix} \frac{\partial \mathbf{q}}{\partial u} & \frac{\partial \mathbf{q}}{\partial v} \end{bmatrix}\right)\begin{bmatrix} \frac{du}{dt} \\ \frac{dv}{dt} \end{bmatrix} &= \begin{bmatrix} \frac{\partial \mathbf{q}}{\partial u} & \frac{\partial \mathbf{q}}{\partial v} & \mathbf{N} \end{bmatrix}\begin{bmatrix} \Gamma^u_{wu} & \Gamma^u_{wv} \\ \Gamma^v_{wu} & \Gamma^v_{wv} \\ L_{wu} & L_{wv} \end{bmatrix}\begin{bmatrix} \frac{du}{dt} \\ \frac{dv}{dt} \end{bmatrix} \\ &= \frac{\partial \mathbf{q}}{\partial u}\begin{bmatrix} \Gamma^u_{wu} & \Gamma^u_{wv} \end{bmatrix}\begin{bmatrix} \frac{du}{dt} \\ \frac{dv}{dt} \end{bmatrix} \\ &\quad + \frac{\partial \mathbf{q}}{\partial v}\begin{bmatrix} \Gamma^v_{wu} & \Gamma^v_{wv} \end{bmatrix}\begin{bmatrix} \frac{du}{dt} \\ \frac{dv}{dt} \end{bmatrix} \\ &\quad + \mathbf{N}\begin{bmatrix} L_{wu} & L_{wv} \end{bmatrix}\begin{bmatrix} \frac{du}{dt} \\ \frac{dv}{dt} \end{bmatrix}\end{aligned}$$

which after further rearranging allows us to conclude

$$\begin{aligned}\ddot{\mathbf{q}} &= \begin{bmatrix} \frac{\partial \mathbf{q}}{\partial u} & \frac{\partial \mathbf{q}}{\partial v} \end{bmatrix} \begin{bmatrix} \frac{d^2u}{dt^2} \\ \frac{d^2v}{dt^2} \end{bmatrix} \\ &\quad + \frac{\partial \mathbf{q}}{\partial u}\begin{bmatrix} \frac{du}{dt} & \frac{dv}{dt} \end{bmatrix}\begin{bmatrix} \Gamma^u_{uu} & \Gamma^u_{uv} \\ \Gamma^u_{vu} & \Gamma^u_{vv} \end{bmatrix}\begin{bmatrix} \frac{du}{dt} \\ \frac{dv}{dt} \end{bmatrix} \\ &\quad + \frac{\partial \mathbf{q}}{\partial v}\begin{bmatrix} \frac{du}{dt} & \frac{dv}{dt} \end{bmatrix}\begin{bmatrix} \Gamma^v_{uu} & \Gamma^v_{vu} \\ \Gamma^v_{vu} & \Gamma^v_{vv} \end{bmatrix}\begin{bmatrix} \frac{du}{dt} \\ \frac{dv}{dt} \end{bmatrix} \\ &\quad + \mathbf{N}\begin{bmatrix} \frac{du}{dt} & \frac{dv}{dt} \end{bmatrix}\begin{bmatrix} L_{uu} & L_{uv} \\ L_{vu} & L_{vv} \end{bmatrix}\begin{bmatrix} \frac{du}{dt} \\ \frac{dv}{dt} \end{bmatrix} \\ &= \begin{bmatrix} \frac{\partial \mathbf{q}}{\partial u} & \frac{\partial \mathbf{q}}{\partial v} & \mathbf{N} \end{bmatrix}\begin{bmatrix} \frac{d^2u}{dt^2} + \Gamma^u(\dot{\mathbf{q}}, \dot{\mathbf{q}}) \\ \frac{d^2v}{dt^2} + \Gamma^v(\dot{\mathbf{q}}, \dot{\mathbf{q}}) \\ \mathrm{II}(\dot{\mathbf{q}}, \dot{\mathbf{q}}) \end{bmatrix}.\end{aligned}$$

Alternately the whole calculation could have been done using summations

$$\begin{aligned}
\ddot{\mathbf{q}} &= \frac{d^2\mathbf{q}}{dt^2} \\
&= \frac{\partial \mathbf{q}}{\partial u}\frac{d^2u}{dt^2} + \frac{\partial \mathbf{q}}{\partial v}\frac{d^2v}{dt^2} \\
&\quad + \left(\frac{\partial^2 \mathbf{q}}{\partial u^2}\frac{du}{dt} + \frac{\partial^2 \mathbf{q}}{\partial u \partial v}\frac{dv}{dt}\right)\frac{du}{dt} + \left(\frac{\partial^2 \mathbf{q}}{\partial u \partial v}\frac{du}{dt} + \frac{\partial^2 \mathbf{q}}{\partial v^2}\frac{dv}{dt}\right)\frac{dv}{dt} \\
&= \frac{\partial \mathbf{q}}{\partial u}\frac{d^2u}{dt^2} + \frac{\partial \mathbf{q}}{\partial v}\frac{d^2v}{dt^2} + \sum_{w_1,w_2=u,v} \frac{\partial^2 \mathbf{q}}{\partial w_1 \partial w_2}\frac{dw_1}{dt}\frac{dw_2}{dt} \\
&= \frac{\partial \mathbf{q}}{\partial u}\left(\frac{d^2u}{dt^2} + \sum_{w_1,w_2=u,v} \Gamma^u_{w_1w_2}\frac{dw_1}{dt}\frac{dw_2}{dt}\right) \\
&\quad + \frac{\partial \mathbf{q}}{\partial v}\left(\frac{d^2v}{dt^2} + \sum_{w_1,w_2=u,v} \Gamma^v_{w_1w_2}\frac{dw_1}{dt}\frac{dw_2}{dt}\right) \\
&\quad + \mathbf{N}\left(\sum_{w_1,w_2=u,v} L_{w_1w_2}\frac{dw_1}{dt}\frac{dw_2}{dt}\right) \\
&= \frac{\partial \mathbf{q}}{\partial u}\left(\frac{d^2u}{dt^2} + \Gamma^u(\dot{\mathbf{q}},\dot{\mathbf{q}})\right) + \frac{\partial \mathbf{q}}{\partial v}\left(\frac{d^2v}{dt^2} + \Gamma^v(\dot{\mathbf{q}},\dot{\mathbf{q}})\right) + \mathbf{N}\mathrm{II}(\dot{\mathbf{q}},\dot{\mathbf{q}}).
\end{aligned}$$

□

Note that we have again shown

Corollary 6.3.2. (*Euler, 1760 and Meusnier, 1776) The normal component of the acceleration satisfies*

$$(\ddot{\mathbf{q}} \cdot \mathbf{N})\,\mathbf{N} = \ddot{\mathbf{q}}^{\mathrm{II}} = \mathbf{N}\mathrm{II}(\dot{\mathbf{q}},\dot{\mathbf{q}}).$$

In particular, two curves with the same velocity at a point have the same normal acceleration components.

The tangential component is more complicated

$$\left[\begin{array}{cc} \frac{\partial \mathbf{q}}{\partial u} & \frac{\partial \mathbf{q}}{\partial v} \end{array}\right][\mathrm{I}]\left[\begin{array}{cc} \frac{\partial \mathbf{q}}{\partial u} & \frac{\partial \mathbf{q}}{\partial v} \end{array}\right]^t \ddot{\mathbf{q}} = \ddot{\mathbf{q}}^{\mathrm{I}} = \frac{\partial \mathbf{q}}{\partial u}\left(\frac{d^2u}{dt^2} + \Gamma^u(\dot{\mathbf{q}},\dot{\mathbf{q}})\right) + \frac{\partial \mathbf{q}}{\partial v}\left(\frac{d^2v}{dt^2} + \Gamma^v(\dot{\mathbf{q}},\dot{\mathbf{q}})\right).$$

But it seems to be a more genuine acceleration as it includes second derivatives. It actually tells us what acceleration we feel on the surface. Note that we have now proved that the tangential acceleration only depends on the first fundamental form.

Remark 6.3.3. We can also consider space-like surfaces $\mathbf{q}(u,v) : U \to \mathbb{R}^{2,1}$. These also have a normal $\mathbf{N}$, but it has the property that $|\mathbf{N}|^2 = \mathbf{N} \cdot \mathbf{N} = -1$ as well as the usual conditions: $\mathbf{N} \cdot \frac{\partial \mathbf{q}}{\partial u} = 0 = \mathbf{N} \cdot \frac{\partial \mathbf{q}}{\partial v}$. However, $\mathbf{N}$ cannot be calculated as easily from the standard vector calculus cross product $\frac{\partial \mathbf{q}}{\partial u} \times \frac{\partial \mathbf{q}}{\partial v}$. The projection formulas will look a little different. If we focus on a curve $\mathbf{q}(t)$ in this surface we still have

$$\dot{\mathbf{q}} = \frac{d\mathbf{q}}{dt} = \frac{d\mathbf{q}}{dt} = \frac{\partial \mathbf{q}}{\partial u}\frac{du}{dt} + \frac{\partial \mathbf{q}}{\partial v}\frac{dv}{dt} = \left[\begin{array}{cc} \frac{\partial \mathbf{q}}{\partial u} & \frac{\partial \mathbf{q}}{\partial v} \end{array}\right]\left[\begin{array}{c} \frac{du}{dt} \\ \frac{dv}{dt} \end{array}\right]$$

since this doesn't depend on any geometric structure. The acceleration however, now decomposes as

$$\begin{aligned}\ddot{\mathbf{q}} &= \ddot{\mathbf{q}}^{\mathrm{I}}+\ddot{\mathbf{q}}^{\mathrm{II}}\\ &= \begin{bmatrix} \frac{\partial \mathbf{q}}{\partial u} & \frac{\partial \mathbf{q}}{\partial v}\end{bmatrix}[\mathrm{I}]^{-1}\begin{bmatrix} \frac{\partial \mathbf{q}}{\partial u} & \frac{\partial \mathbf{q}}{\partial v}\end{bmatrix}^{t}\ddot{\mathbf{q}}-(\ddot{\mathbf{q}}\cdot\mathbf{N})\,\mathbf{N},\end{aligned}$$

where $\ddot{\mathbf{q}}^{\mathrm{I}}$ is tangent to the surface and $\ddot{\mathbf{q}}^{\mathrm{II}}$ proportional to $\mathbf{N}$. Here all inner products

$$\frac{\partial \mathbf{q}}{\partial u}\cdot\ddot{\mathbf{q}},\ \frac{\partial \mathbf{q}}{\partial v}\cdot\ddot{\mathbf{q}},\ \mathbf{N}\cdot\ddot{\mathbf{q}}$$

are the $\mathbb{R}^{2,1}$ inner product. The tangential part of the acceleration can also be calculated intrinsically with the same formula as above:

$$\ddot{\mathbf{q}}^{\mathrm{I}}=\begin{bmatrix} \frac{\partial \mathbf{q}}{\partial u} & \frac{\partial \mathbf{q}}{\partial v}\end{bmatrix}[\mathrm{I}]^{-1}\begin{bmatrix} \frac{\partial \mathbf{q}}{\partial u} & \frac{\partial \mathbf{q}}{\partial v}\end{bmatrix}^{t}\ddot{\mathbf{q}}=\frac{\partial \mathbf{q}}{\partial u}\left(\frac{d^2u}{dt^2}+\Gamma^{u}(\dot{\mathbf{q}},\dot{\mathbf{q}})\right)+\frac{\partial \mathbf{q}}{\partial v}\left(\frac{d^2v}{dt^2}+\Gamma^{v}(\dot{\mathbf{q}},\dot{\mathbf{q}})\right).$$

Exercises.

(1) Assume that a unit speed curve satisfies an equation $F(u,v)=R$ on a parameterized surface $\mathbf{q}(u,v)$. If we use $\partial_w F_w=\frac{\partial F}{\partial w}$ show that

$$\partial_u F\dot{u}+\partial_v F\dot{v}=0.$$

(a) Show that

$$\begin{aligned}\dot{\mathbf{q}} &= \dot{u}\frac{\partial \mathbf{q}}{\partial u}+\dot{v}\frac{\partial \mathbf{q}}{\partial v}\\ &= \frac{\pm 1}{\sqrt{g_{uu}\partial_v F\partial_v F-2g_{uv}\partial_u F\partial_v F+g_{vv}\partial_u F\partial_u F}}\left(-\partial_v F\frac{\partial \mathbf{q}}{\partial u}+\partial_u F\frac{\partial \mathbf{q}}{\partial v}\right).\end{aligned}$$

This means that the unit tangent can be calculated without reference to the parametrization of the curve.

(b) Show that if we use this formula for the velocity, then the geodesic curvature can be computed as

$$\kappa_g=\frac{\frac{\partial}{\partial u}\left(\dot{\mathbf{q}}\cdot\frac{\partial \mathbf{q}}{\partial v}\right)-\frac{\partial}{\partial v}\left(\dot{\mathbf{q}}\cdot\frac{\partial \mathbf{q}}{\partial u}\right)}{\sqrt{\det[\mathrm{I}]}}.$$

(c) Generalize this to the situation where a curve satisfies a differential relation

$$P\dot{u}+Q\dot{v}=0,$$

where $P=P(u,v)$ and $Q=Q(u,v)$.

(2) Define the Hessian of a function on a surface by

$$\mathrm{Hess}f(X,Y)=\mathrm{I}(D_X\nabla f,Y).$$

Show that the entries in the matrix $[\mathrm{Hess}f]$ defined by

$$\mathrm{Hess}f(X,Y)=\begin{bmatrix} X^u & X^v\end{bmatrix}[\mathrm{Hess}f]\begin{bmatrix} Y^u\\ Y^v\end{bmatrix}$$

are given as

$$\partial^2_{ij}f+\begin{bmatrix}\partial_u f & \partial_v f\end{bmatrix}\begin{bmatrix}\Gamma^u_{ij}\\ \Gamma^v_{ij}\end{bmatrix}.$$

Further relate these entries to the dot products

$$\partial_i\nabla f\cdot\partial_j\mathbf{q}.$$

6.4. The Gauss and Codazzi Equations

The goal in this section is to establish the classical Gauss formula and the accompanying Codaazi equations from the Gauss formulas and Weingarten equations. The Codazzi equations were historically first discovered by K.M. Peterson in 1853, then rediscovered by G. Mainardi in 1856, and then finally by D. Codazzi in 1867.

Recall from sections 4.6 and 5.2 the Gauss formulas and Weingarten equations in combined form:

$$\frac{\partial}{\partial w}\begin{bmatrix} \frac{\partial \mathbf{q}}{\partial u} & \frac{\partial \mathbf{q}}{\partial v} & \mathbf{N} \end{bmatrix} = \begin{bmatrix} \frac{\partial \mathbf{q}}{\partial u} & \frac{\partial \mathbf{q}}{\partial v} & \mathbf{N} \end{bmatrix}[D_w].$$

Taking one more derivative on both sides yields

$$\begin{aligned}
\frac{\partial^2}{\partial u \partial v}\begin{bmatrix} \frac{\partial \mathbf{q}}{\partial u} & \frac{\partial \mathbf{q}}{\partial v} & \mathbf{N} \end{bmatrix} &= \left(\frac{\partial}{\partial u}\begin{bmatrix} \frac{\partial \mathbf{q}}{\partial u} & \frac{\partial \mathbf{q}}{\partial v} & \mathbf{N} \end{bmatrix}\right)[D_v] \\
&\quad + \begin{bmatrix} \frac{\partial \mathbf{q}}{\partial u} & \frac{\partial \mathbf{q}}{\partial v} & \mathbf{N} \end{bmatrix}\left(\frac{\partial}{\partial u}[D_v]\right) \\
&= \begin{bmatrix} \frac{\partial \mathbf{q}}{\partial u} & \frac{\partial \mathbf{q}}{\partial v} & \mathbf{N} \end{bmatrix}[D_u][D_v] \\
&\quad + \begin{bmatrix} \frac{\partial \mathbf{q}}{\partial u} & \frac{\partial \mathbf{q}}{\partial v} & \mathbf{N} \end{bmatrix}\left(\frac{\partial}{\partial u}[D_v]\right)
\end{aligned}$$

and similarly

$$\frac{\partial^2}{\partial v \partial u}\begin{bmatrix} \frac{\partial \mathbf{q}}{\partial u} & \frac{\partial \mathbf{q}}{\partial v} & \mathbf{N} \end{bmatrix} = \begin{bmatrix} \frac{\partial \mathbf{q}}{\partial u} & \frac{\partial \mathbf{q}}{\partial v} & \mathbf{N} \end{bmatrix}\left([D_v][D_u] + \frac{\partial}{\partial v}[D_u]\right).$$

Now using that

$$\frac{\partial^2}{\partial u \partial v}\begin{bmatrix} \frac{\partial \mathbf{q}}{\partial u} & \frac{\partial \mathbf{q}}{\partial v} & \mathbf{N} \end{bmatrix} = \frac{\partial^2}{\partial v \partial u}\begin{bmatrix} \frac{\partial \mathbf{q}}{\partial u} & \frac{\partial \mathbf{q}}{\partial v} & \mathbf{N} \end{bmatrix}$$

we obtain after writing out the entries in the matrices

$$\begin{aligned}
&\begin{bmatrix} \partial_u \Gamma^u_{vu} & \partial_u \Gamma^u_{vv} & -\partial_u L^u_v \\ \partial_u \Gamma^v_{vu} & \partial_u \Gamma^v_{vv} & -\partial_u L^v_v \\ \partial_u L_{vu} & \partial_u L_{vv} & 0 \end{bmatrix} + \begin{bmatrix} \Gamma^u_{uu} & \Gamma^u_{uv} & -L^u_u \\ \Gamma^v_{uu} & \Gamma^v_{uv} & -L^v_u \\ L_{uu} & L_{uv} & 0 \end{bmatrix}\begin{bmatrix} \Gamma^u_{vu} & \Gamma^u_{vv} & -L^u_v \\ \Gamma^v_{vu} & \Gamma^v_{vv} & -L^v_v \\ L_{vu} & L_{vv} & 0 \end{bmatrix} \\
= &\begin{bmatrix} \partial_v \Gamma^u_{uu} & \partial_v \Gamma^u_{uv} & -\partial_v L^u_u \\ \partial_v \Gamma^v_{uu} & \partial_v \Gamma^v_{uv} & -\partial_v L^v_u \\ \partial_v L_{uu} & \partial_v L_{uv} & 0 \end{bmatrix} + \begin{bmatrix} \Gamma^u_{vu} & \Gamma^u_{vv} & -L^u_v \\ \Gamma^v_{vu} & \Gamma^v_{vv} & -L^v_v \\ L_{vu} & L_{vv} & 0 \end{bmatrix}\begin{bmatrix} \Gamma^u_{uu} & \Gamma^u_{uv} & -L^u_u \\ \Gamma^v_{uu} & \Gamma^v_{uv} & -L^v_u \\ L_{uu} & L_{uv} & 0 \end{bmatrix}.
\end{aligned}$$

If we restrict attention to the the general terms of the entries in the first two columns and rows we obtain 4 equations for the partial derivatives of Γ^k_{ij} where each i, j, k can be u, v. If we also write $\partial_w = \frac{\partial}{\partial w}$ we obtain

$$\partial_i \Gamma^l_{jk} + \begin{bmatrix} \Gamma^l_{iu} & \Gamma^l_{iv} & -L^l_i \end{bmatrix}\begin{bmatrix} \Gamma^u_{jk} \\ \Gamma^v_{jk} \\ L_{jk} \end{bmatrix} = \partial_j \Gamma^l_{ik} + \begin{bmatrix} \Gamma^l_{ju} & \Gamma^l_{jv} & -L^l_j \end{bmatrix}\begin{bmatrix} \Gamma^u_{ik} \\ \Gamma^v_{ik} \\ L_{ik} \end{bmatrix}$$

which can further be rearranged by isolating Γs on one side:

$$\partial_i \Gamma^l_{jk} - \partial_j \Gamma^l_{ik} + \begin{bmatrix} \Gamma^l_{iu} & \Gamma^l_{iv} \end{bmatrix}\begin{bmatrix} \Gamma^u_{jk} \\ \Gamma^v_{jk} \end{bmatrix} - \begin{bmatrix} \Gamma^l_{ju} & \Gamma^l_{jv} \end{bmatrix}\begin{bmatrix} \Gamma^u_{ik} \\ \Gamma^v_{ik} \end{bmatrix} = L^l_i L_{jk} - L^l_j L_{ik}.$$

These are called the *Gauss Equations*. Note that we only established these equations when $i = u$ and $j = v$. Clearly they also hold when $u = j$ and $v = i$ as both sides just change sign. They also hold trivially when $i = j$ as both sides vanish in that case. This means that the 4 original equations can be expanded to 16 equations where the 4 indices i, j, k, l can be both u, v.

The *Riemann curvature tensor* is defined as the left hand side of the Gauss equations

$$R^l_{ijk} = \partial_i \Gamma^l_{jk} - \partial_j \Gamma^l_{ik} + \begin{bmatrix} \Gamma^l_{iu} & \Gamma^l_{iv} \end{bmatrix} \begin{bmatrix} \Gamma^u_{jk} \\ \Gamma^v_{jk} \end{bmatrix} - \begin{bmatrix} \Gamma^l_{ju} & \Gamma^l_{jv} \end{bmatrix} \begin{bmatrix} \Gamma^u_{ik} \\ \Gamma^v_{ik} \end{bmatrix}.$$

It is clearly an object that can be calculated directly from the first fundamental form, although it is certainly not always easy to do so. We just discussed that if $i \neq j$, then

$$R^l_{ijk} = -R^l_{jik},$$

and

$$R^l_{iik} = 0.$$

More explicitly the skew-symmetry can be spelled out

$$\begin{aligned} R^u_{uvu} &= -R^u_{vuu}, \\ R^v_{uvu} &= -R^v_{vuu}, \\ R^u_{uvv} &= -R^u_{vuv}, \\ R^v_{uvv} &= -R^v_{vuv}. \end{aligned}$$

A slightly less obvious formula is the *Bianchi identity*

$$R^l_{ijk} + R^l_{kij} + R^l_{jki} = 0.$$

It too follows from the above definition, but with more calculations. Unfortunately it doesn't reduce our job of computing curvatures. The final reduction comes about by constructing

$$R_{ijkl} = R^u_{ijk} g_{ul} + R^v_{ijk} g_{vl}$$

and showing that

$$R_{ijkl} = -R_{ijlk}.$$

This means that the 4 possibly nontrivial curvatures are related by

$$R_{uvvu} = R_{vuuv} = -R_{uvuv} = -R_{vuvu}.$$

All of the curvatures of both types turn out to be related to an old friend.

THEOREM 6.4.1. *(Theorema Egregium) The Gauss curvature can be computed knowing only the first fundamental form*

$$\begin{aligned} K &= \frac{R^u_{uvv}}{g_{vv}} = \frac{R^v_{vuu}}{g_{uu}} \\ &= -\frac{R^v_{uvv}}{g_{vu}} = -\frac{R^u_{vuu}}{g_{vu}} \\ &= \frac{R_{uvvu}}{\det[\mathrm{I}]}. \end{aligned}$$

PROOF. We know that

$$K = L^u_u L^v_v - L^u_v L^v_u$$

and

$$\begin{aligned} \begin{bmatrix} L^u_u & L^u_v \\ L^v_u & L^v_v \end{bmatrix} &= \begin{bmatrix} g^{uu} & g^{uv} \\ g^{vu} & g^{vv} \end{bmatrix} \begin{bmatrix} L_{uu} & L_{uv} \\ L_{vu} & L_{vv} \end{bmatrix}, \\ [L] &= [\mathrm{I}]^{-1} [\mathrm{II}]. \end{aligned}$$

Now let $u = i = l$ and $v = j = k$ in the Gauss equation. We take the strange route of calculating so that we end up with second fundamental form terms. This is because [II] is always symmetric, while $[L]$ might not be symmetric. Thus several steps are somewhat simplified.

$$\begin{aligned}
R^u_{uvv} &= L^u_u L_{vv} - L^u_v L_{uv} \\
&= \left(g^{uu} L_{uu} + g^{uv} L_{vu}\right) L_{vv} - \left(g^{uu} L_{uv} + g^{uv} L_{vv}\right) L_{uv} \\
&= g^{uu} \left(L_{uu} L_{vv} - L_{uv} L_{uv}\right) \\
&= g^{uu} \det [\mathrm{II}] \\
&= g^{uu} \det [\mathrm{I}] \det L \\
&= g_{vv} \det L \\
&= g_{vv} K.
\end{aligned}$$

The second equality follows by a similar calculation. For the third (and in a similar way fourth) the Gauss equations can again be used to calculate

$$\begin{aligned}
R^v_{uvv} &= L^v_u L_{vv} - L^v_v L_{uv} \\
&= \left(g^{vu} L_{uu} + g^{vv} L_{vu}\right) L_{vv} - \left(g^{vu} L_{uv} + g^{vv} L_{vv}\right) L_{uv} \\
&= g^{vu} \left(L_{uu} L_{vv} - L_{uv} L_{uv}\right) \\
&= -g_{vu} K.
\end{aligned}$$

Finally note that

$$\begin{aligned}
R_{uvvu} &= R^u_{uvv} g_{uu} + R^v_{uvv} g_{vu} \\
&= K g_{vv} g_{uu} + R^v_{uvv} g_{vu} \\
&= K \left(g_{vv} g_{uu} - g_{uv} g_{vu}\right) \\
&= K \det [\mathrm{I}] .
\end{aligned}$$

□

COROLLARY 6.4.2. *If an abstract surface has constant Gauss curvature K, then the Riemann curvature tensor is given by*

$$R^l_{ijk} = K \left(\delta^l_i g_{jk} - \delta^l_j g_{ik}\right).$$

PROOF. It is easy to check that both sides vanish when for the same indices. For the other possibilities of the indices this follows from theorem 6.4.1. □

The other entries in the matrices above reduce to the *Codazzi Equations*

$$\partial_i L_{jk} + \begin{bmatrix} L_{iu} & L_{iv} & 0 \end{bmatrix} \begin{bmatrix} \Gamma^u_{jk} \\ \Gamma^v_{jk} \\ L_{jk} \end{bmatrix} = \partial_j L_{ik} + \begin{bmatrix} L_{ju} & L_{jv} & 0 \end{bmatrix} \begin{bmatrix} \Gamma^u_{ik} \\ \Gamma^v_{ik} \\ L_{ik} \end{bmatrix}$$

or rearranged

$$\partial_i L_{jk} - \partial_j L_{ik} = \begin{bmatrix} L_{ju} & L_{jv} \end{bmatrix} \begin{bmatrix} \Gamma^u_{ik} \\ \Gamma^v_{ik} \end{bmatrix} - \begin{bmatrix} L_{iu} & L_{iv} \end{bmatrix} \begin{bmatrix} \Gamma^u_{jk} \\ \Gamma^v_{jk} \end{bmatrix}.$$

Note again that while we only established these for $u = i$ and $j = v$ that they also hold when u, v are switched and that both sides vanish when $i = j$.

We are now ready to present the fundamental theorem of surface theory. It is analogous to theorem 2.2.2 for planar curves.

THEOREM 6.4.3. (Fundamental Theorem of Surface Theory, Bonnet, 1848) *A surface is uniquely determined by its first and second fundamental forms if its position and tangent space space are known at just one point. Conversely any set of abstract first and second fundamental forms that are related by the Gauss and Codazzi equations are locally the first and second fundamental forms of a surface.*

PROOF. We start by observing that the matrices $[D_w]$ can be defined as long as we are given [I] and [II]. So we seek solutions to a rather big system

$$\begin{aligned}
\frac{\partial \mathbf{q}}{\partial u} &= U, \\
\frac{\partial \mathbf{q}}{\partial v} &= V, \\
\frac{\partial}{\partial u}\begin{bmatrix} U & V & \mathbf{N} \end{bmatrix} &= \begin{bmatrix} U & V & \mathbf{N} \end{bmatrix}[D_u], \\
\frac{\partial}{\partial v}\begin{bmatrix} U & V & \mathbf{N} \end{bmatrix} &= \begin{bmatrix} U & V & \mathbf{N} \end{bmatrix}[D_v],
\end{aligned}$$

with some specific initial conditions

$$\begin{aligned}
\mathbf{q}(0,0) &= \mathbf{q}_0 \in \mathbb{R}^3, \\
U(0,0) &= U_0 \in \mathbb{R}^3, \\
V(0,0) &= V_0 \in \mathbb{R}^3, \\
\mathbf{N}(0,0) &= \mathbf{N}_0 \in \mathbb{R}^3,
\end{aligned}$$

where we additionally require that

$$\begin{aligned}
U_0 \cdot U_0 &= g_{uu}(0,0), \\
U_0 \cdot V_0 &= g_{uv}(0,0), \\
V_0 \cdot V_0 &= g_{vv}(0,0), \\
\mathbf{N}_0 &= \frac{U_0 \times V_0}{|U_0 \times V_0|}.
\end{aligned}$$

It is clear that this big system has a unique solution given the initial values. Conversely to solve it we must check that the necessary integrability conditions are satisfied. We can separate the problem into first solving

$$\begin{aligned}
\frac{\partial}{\partial u}\begin{bmatrix} U & V & \mathbf{N} \end{bmatrix} &= \begin{bmatrix} U & V & \mathbf{N} \end{bmatrix}[D_u], \\
\frac{\partial}{\partial v}\begin{bmatrix} U & V & \mathbf{N} \end{bmatrix} &= \begin{bmatrix} U & V & \mathbf{N} \end{bmatrix}[D_v].
\end{aligned}$$

Here the integrability conditions are satisfied as we assumed that

$$[D_u][D_v] + \frac{\partial}{\partial u}[D_v] = [D_v][D_u] + \frac{\partial}{\partial v}[D_u].$$

Having solved this system with the given initial values it remains to find the surface by solving

$$\begin{aligned}
\frac{\partial \mathbf{q}}{\partial u} &= U, \\
\frac{\partial \mathbf{q}}{\partial v} &= V.
\end{aligned}$$

Here the right hand side does not depend on $\mathbf{q}$ so the integrability conditions are simply

$$\frac{\partial U}{\partial v}=\frac{\partial V}{\partial u}.$$

However, we know that

$$\begin{aligned}
\frac{\partial U}{\partial v} &= \begin{bmatrix} U & V & \mathbf{N} \end{bmatrix} \begin{bmatrix} \Gamma^u_{vu} \\ \Gamma^v_{vu} \\ L_{vu} \end{bmatrix}, \\
\frac{\partial V}{\partial u} &= \begin{bmatrix} U & V & \mathbf{N} \end{bmatrix} \begin{bmatrix} \Gamma^u_{uv} \\ \Gamma^v_{uv} \\ L_{uv} \end{bmatrix},
\end{aligned}$$

where the right-hand sides are equal as $L_{uv}=L_{vu}$ and $\Gamma^w_{uv}=\Gamma^w_{vu}$.

Having solved the equations it then remains to show that the surface we have constructed has the correct first and second fundamental forms. This will of course depend on the extra conditions that we imposed:

$$\begin{aligned}
U_0\cdot U_0 &= g_{uu}(0,0), \\
U_0\cdot V_0 &= g_{uv}(0,0), \\
V_0\cdot V_0 &= g_{vv}(0,0), \\
\mathbf{N}_0 &= \frac{U_0\times V_0}{|U_0\times V_0|}.
\end{aligned}$$

In fact they show that at $(0,0)$ the surface has the correct first fundamental form and normal vector. More generally consider the 3×3 matrix of inner products

$$\begin{bmatrix} U & V & \mathbf{N} \end{bmatrix}^t \begin{bmatrix} U & V & \mathbf{N} \end{bmatrix},$$

where the block consisting of

$$\begin{bmatrix} U & V \end{bmatrix}^t \begin{bmatrix} U & V \end{bmatrix}$$

corresponds to the first fundamental form of the surface we have constructed. The derivative of this 3×3 matrix satisfies

$$\begin{aligned}
&\frac{\partial}{\partial w}\left(\begin{bmatrix} U & V & \mathbf{N} \end{bmatrix}^t \begin{bmatrix} U & V & \mathbf{N} \end{bmatrix}\right) \\
&= \left(\frac{\partial}{\partial w}\begin{bmatrix} U & V & \mathbf{N} \end{bmatrix}\right)^t \begin{bmatrix} U & V & \mathbf{N} \end{bmatrix} + \begin{bmatrix} U & V & \mathbf{N} \end{bmatrix}^t \frac{\partial}{\partial w}\begin{bmatrix} U & V & \mathbf{N} \end{bmatrix} \\
&= \left(\begin{bmatrix} U & V & \mathbf{N} \end{bmatrix}[D_w]\right)^t \begin{bmatrix} U & V & \mathbf{N} \end{bmatrix} + \begin{bmatrix} U & V & \mathbf{N} \end{bmatrix}^t \begin{bmatrix} U & V & \mathbf{N} \end{bmatrix}[D_w] \\
&= [D_w]^t \begin{bmatrix} U & V & \mathbf{N} \end{bmatrix}^t \begin{bmatrix} U & V & \mathbf{N} \end{bmatrix} + \begin{bmatrix} U & V & \mathbf{N} \end{bmatrix}^t \begin{bmatrix} U & V & \mathbf{N} \end{bmatrix}[D_w].
\end{aligned}$$

This is a differential equation of the type

$$\frac{\partial X}{\partial w}=[D_w]^t X + X[D_w],$$

where X is a 3×3 matrix. Now

$$X=\begin{bmatrix} g_{uu} & g_{uv} & 0 \\ g_{vu} & g_{vv} & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

also satisfies this equation as we constructed $[D_w]$ directly from the given first and second fundamental forms. However, these two solutions have the same initial value at $(0,0)$ so they must be equal. This shows that our surface has the correct first fundamental form and also that $\mathbf{N}$ is a unit normal to the surface. This in turn implies that we also obtain the correct second fundamental form since we now also know that

$$\begin{bmatrix} U & V & \mathbf{N} \end{bmatrix}^t \frac{\partial}{\partial w} \begin{bmatrix} U & V & \mathbf{N} \end{bmatrix} = \begin{bmatrix} U & V & \mathbf{N} \end{bmatrix}^t \begin{bmatrix} U & V & \mathbf{N} \end{bmatrix} [D_w],$$

where the right hand side is now known and the left hand side contains all of the terms we need for calculating the second fundamental form of the constructed surface. □

This theorem allows us to give a complete local characterization of abstract surfaces with constant non-negative Gauss curvature. In fact such surfaces are forced to be locally isometric to the plane or a sphere.

THEOREM 6.4.4. *An abstract surface of constant Gauss curvature $K \geq 0$, can locally be represented as part of a plane when $K = 0$ and part of a sphere of radius $1/\sqrt{K}$ when $K > 0$.*

PROOF. We are given I and have to guess II. We use $\mathrm{II} = \sqrt{K}\mathrm{I}$, i.e., $L^i_j = \sqrt{K}\delta^i_j$ and $L_{ij} = \sqrt{K}g_{ij}$. This allows us to calculate $[D_i]$ in a specific parametrization. We are then left with the goal of checking the integrability conditions, i.e., the Gauss and Codazzi equations. The Codazzi equations are obviously satisfied when $\mathrm{II} = 0$, and follow from the formula for the Christoffel symbols when $K > 0$. More precisely we start with the right hand side of the Codazzi equations and use the intrinsic formulas for the Christoffel symbols from remark 6.1.2 to show that they hold:

$$\begin{aligned}
&\sqrt{K}\begin{bmatrix} g_{ju} & g_{jv} \end{bmatrix}\begin{bmatrix} \Gamma^u_{ik} \\ \Gamma^v_{ik} \end{bmatrix} - \sqrt{K}\begin{bmatrix} g_{iu} & g_{iv} \end{bmatrix}\begin{bmatrix} \Gamma^u_{jk} \\ \Gamma^v_{jk} \end{bmatrix} \\
&= \sqrt{K}\begin{bmatrix} g_{ju} & g_{jv} \end{bmatrix}\begin{bmatrix} \Gamma^u_{ik} \\ \Gamma^v_{ik} \end{bmatrix} - \sqrt{K}\begin{bmatrix} g_{iu} & g_{iv} \end{bmatrix}\begin{bmatrix} \Gamma^u_{jk} \\ \Gamma^v_{jk} \end{bmatrix} \\
&= \sqrt{K}\left(\Gamma_{ikj} - \Gamma_{jki}\right) \\
&= \frac{\sqrt{K}}{2}\left(\left(\partial_i g_{kj} + \partial_k g_{ij} - \partial_j g_{ik}\right) - \left(\partial_j g_{ik} + \partial_k g_{ji} - \partial_i g_{jk}\right)\right) \\
&= \frac{\sqrt{K}}{2}\left(\left(\partial_i g_{kj} - \partial_j g_{ik}\right) - \left(\partial_j g_{ik} - \partial_i g_{jk}\right)\right) \\
&= \sqrt{K}\left(\partial_i g_{kj} - \partial_j g_{ik}\right) \\
&= \sqrt{K}\partial_i g_{jk} - R\partial_j g_{ik}.
\end{aligned}$$

Our assumptions about the the second fundamental form imply

$$L^l_i L_{jk} - L^l_j L_{ik} = K\delta^l_i g_{jk} - K\delta^l_j g_{ik}$$

and corollary 6.4.2 shows that this gives us the Gauss equations:

$$R^l_{ijk} = L^l_i L_{jk} - L^l_j L_{ik}.$$

Now that we have a local representation of the abstract surface as a parametrized surface in $\mathbb{R}^3$ with $\mathrm{II} = \sqrt{K}\mathrm{I}$ we can use theorem 5.4.7 to finish the proof. □

REMARK 6.4.5. It is possible to develop a theory for space-like surfaces in $\mathbb{R}^{2,1}$ that mirrors the theory for surfaces in $\mathbb{R}^3$. This includes new versions of the Gauss and Codazzi equations that also lead to exact analogies of theorems 6.4.3 and 6.4.4. Thus abstract surfaces of constant negative curvature $-R^{-2}$ can locally be represented as part of the surface in $\mathbb{R}^{2,1}$ given by the equation

$$x^2 + y^2 - z^2 = -R^2.$$

We end this long section with a profound theorem that relates to the concepts discussed here. In essence it shows that while it is occasionally possible to choose a second fundamental form locally so that it satisfies the Gauss and Codazzi equations, it might not be possible to extend it to be defined on the entire abstract surface. The result also indicates that in order to characterize hyperbolic space in a way that is similar to theorem 6.4.4 it is most convenient to use $\mathbb{R}^{2,1}$ as the ambient space.

THEOREM 6.4.6. *(Hilbert, 1901) It is not possible to select a second fundamental form II on all of hyperbolic space H such that I and II satisfy the Gauss and Codazzi equations.*

PROOF. We argue by contradiction and assume that such a second fundamental form exists. The Gauss equations imply that at each point there is a positive and negative principal direction for II. Let the positive principal curvature be $\kappa : H \to (0, \infty)$ and the negative $-1/\kappa$. Since $\kappa > 0$ we can find a unique smooth function $\theta : H \to (0, \pi/2)$ such that

$$\cos^2\theta = \frac{1}{1+\kappa^2},\ \sin^2\theta = \frac{\kappa^2}{1+\kappa^2}.$$

Fix a parametrization (x, y) of H, e.g., the one that makes hyperbolic space a Monge patch. At $(0, 0)$ make a choice of orthonormal principal directions E_1, E_2. Extend this choice to be consistent along the x-axis, and then finally along vertical lines to obtain a consistent choice on all of H. Next consider the two vector fields $P = \cos\theta E_1$ and $Q = \sin\theta E_2$. We claim that there is a global parametrization where these are the coordinate vector fields. This would follow directly from the global version of theorem A.5.3 if we could check the integrability conditions

$$\frac{\partial P}{\partial x}Q^x + \frac{\partial P}{\partial y}Q^y = \frac{\partial Q}{\partial x}P^x + \frac{\partial Q}{\partial y}P^y$$

and find M, C such that

$$\sqrt{(P^x)^2 + (P^y)^2},\ \sqrt{(Q^x)^2 + (Q^y)^2} \le M + C\sqrt{x^2+y^2}.$$

Note that in this case P, Q are independent of (u, v). To prove the bounds for P, Q we show that a unit vector field X in the hyperbolic metric satisfies these bounds. From example 6.2.2 we have that

$$\begin{aligned} 1 &= X \cdot X \\ &= \frac{1}{z^2}\left((X^x)^2 + (X^y)^2 + (yX^x - xX^y)^2\right). \end{aligned}$$

Since $z^2 = 1 + x^2 + y^2$ this shows that

$$(X^x)^2 + (X^y)^2 \le 1 + x^2 + y^2$$

which implies our claim.

The integrability conditions are a consequence of the Codazzi equations. We do the calculation by an indirect method where we show that there are local parametrizations $\mathbf{q}(u,v)$ of H where $\partial_u \mathbf{q} = \cos\theta E_1$ and $\partial_v \mathbf{q} = \sin\theta E_2$.

By appealing to remark 4.2.8 we can for any $q \in H$ find a local parametrization $\mathbf{q}(u,v)$ with $\mathbf{q}(0,0) = q$, where the parameter curves are lines of curvature and

$$\begin{aligned}\partial_u \mathbf{q}(u,0) &= \cos\theta(u,0)\, E_1(u,0),\\ \partial_v \mathbf{q}(0,v) &= \sin\theta(0,v)\, E_2(0,v).\end{aligned}$$

The Codazzi equations for such a local parametrization is (see section 6.4 exercise 4 below)

$$\begin{aligned}\frac{\partial \kappa}{\partial v} &= \frac{1}{2}\left(-\frac{1}{\kappa} - \kappa\right)\frac{\partial \ln g_{uu}}{\partial v},\\ \frac{\partial \kappa^{-1}}{\partial u} &= \frac{1}{2}\left(\kappa + \frac{1}{\kappa}\right)\frac{\partial \ln g_{vv}}{\partial u}.\end{aligned}$$

This shows that

$$\begin{aligned}-\frac{2\kappa}{1+\kappa^2}\frac{\partial \kappa}{\partial v} &= \frac{\partial \ln g_{uu}}{\partial v},\\ \frac{2\kappa^{-1}}{1+\kappa^{-2}}\frac{\partial \kappa^{-1}}{\partial u} &= \frac{\partial \ln g_{vv}}{\partial u}.\end{aligned}$$

Consequently,

$$g_{uu} = \frac{f(u)}{1+\kappa^2},\ g_{vv} = \frac{\kappa^2 h(v)}{1+\kappa^2}.$$

Evaluating at $(u,0)$ gives us $f = 1$ and and evaluating at $(0,v)$ that $h = 1$. This tells us that

$$g_{uu} = \cos^2\theta,\ g_{vv} = \sin^2\theta,\ g_{uv} = 0,$$

and

$$L_{uu} = \sin\theta\cos\theta,\ L_{vv} = \sin\theta\cos\theta,\ L_{uv} = 0.$$

This gives us the desired local parametrization and we conclude that there is a global parametrization with the same properties.

If we switch coordinates to $s = u+v$ and $t = u - v$, then $g_{ss} = 1$, $g_{tt} = 1$, $g_{st} = \cos(2\theta)$, $L_{ss} = L_{tt} = 0$, and $L_{st} = \sin(2\theta)$. The formula for the Gauss curvature in such coordinates reduce to the formula

$$2\partial^2_{st}\theta = -K\sin(2\theta) = \sin(2\theta),$$

where $2\theta \in (0,\pi)$. In particular, $\partial^2_{st}\theta > 0$. This shows that $t \mapsto \partial_s\theta(s,t)$ is strictly increasing. Integrating $\partial_s\theta(s,t)$ over an interval $s \in [a,b]$ then shows that for $c < d$ we have

$$\theta(b,c) - \theta(a,c) < \theta(b,d) - \theta(a,d).$$

Now assume that a, c are chosen so that $\partial_s\theta(a,c) \neq 0$. Assume that $\partial_s\theta(a,c) > 0$ and select $b > a$ so that $\partial_s\theta(s,c) > 0$ for all $s \in [a,b]$. It follows that $\partial_s\theta(s,t) > 0$ for all $(s,t) \in [a,b]\times[c,\infty)$. Finally fix $a < a_1 < b_1 < b$ and $\epsilon > 0$ so that

$$\theta(b,c) - \theta(b_1,c) > \epsilon,\ \theta(a_1,c) - \theta(a,c) > \epsilon.$$

Then

$$\theta(b,t) - \theta(b_1,t) > \epsilon,\ \theta(a_1,t) - \theta(a,t) > \epsilon$$

for all $t > c$. In particular,

$$\theta(b_1,t) < \tfrac{\pi}{2} - \epsilon,\ \theta(a_1,t) > \epsilon.$$

As $s \mapsto \theta(s,t)$ is increasing for $s \in [a,b]$ and $t \geq c$, this implies that

$$\epsilon < \theta(s,t) < \tfrac{\pi}{2} - \epsilon$$

for all $(s,t) \in [a_1,b_1] \times [c,\infty)$. In particular, $\sin 2\theta > \sin\epsilon$ for all $(s,t) \in [a_1,b_1] \times [c,\infty)$. This shows that

$$\begin{aligned} \left(\theta(b_1,t) - \theta(a_1,t)\right)\Big|_{t=c}^{t=T} &= \int_c^T \int_{a_1}^{b_1} \partial_{st}\theta\, ds dt \\ &= \frac{1}{2}\int_c^T \int_{a_1}^{b_1} \sin 2\theta\, ds dt \\ &\geq \frac{1}{2}\int_c^T \int_{a_1}^{b_1} \sin\epsilon\, ds dt \\ &= \frac{1}{2}(T-c)(b_1-a_1)\sin\epsilon. \end{aligned}$$

However, the left hand side is bounded by π so we conclude that

$$T < c + \frac{2\pi}{(b_1-a_1)\sin\epsilon}.$$

This contradicts that the inequality holds for all $T > 0$.

When $\partial_s\theta(a,c) < 0$, can redefine $s = -u-v$ and $t = v-u$ so that we get $\partial_s\theta(-a,-c) > 0$ with this new choice of s. $\square$

COROLLARY 6.4.7. *There is no Riemannian immersion from hyperbolic space H to $\mathbb{R}^3$.*

Exercises.

(1) We saw above that all of the Gauss equations reduced to just one relevant equation. Reduce all of the 8 Codazzi equations

$$\partial_i L_{jk} - \partial_j L_{ik} = \begin{bmatrix} L_{ju} & L_{jv} \end{bmatrix} \begin{bmatrix} \Gamma^u_{ik} \\ \Gamma^v_{ik} \end{bmatrix} - \begin{bmatrix} L_{iu} & L_{iv} \end{bmatrix} \begin{bmatrix} \Gamma^u_{jk} \\ \Gamma^v_{jk} \end{bmatrix}$$

to the following two equations

$$\begin{aligned} \frac{\partial L_{vu}}{\partial u} - \frac{\partial L_{uu}}{\partial v} + \begin{bmatrix} L_{uu} & L_{uv} \end{bmatrix}\begin{bmatrix} \Gamma^u_{vu} \\ \Gamma^v_{vu} \end{bmatrix} - \begin{bmatrix} L_{vu} & L_{vv} \end{bmatrix}\begin{bmatrix} \Gamma^u_{uu} \\ \Gamma^v_{uu} \end{bmatrix} &= 0, \\ \frac{\partial L_{vv}}{\partial u} - \frac{\partial L_{uv}}{\partial v} + \begin{bmatrix} L_{uu} & L_{uv} \end{bmatrix}\begin{bmatrix} \Gamma^u_{vv} \\ \Gamma^v_{vv} \end{bmatrix} - \begin{bmatrix} L_{vu} & L_{vv} \end{bmatrix}\begin{bmatrix} \Gamma^u_{uv} \\ \Gamma^v_{uv} \end{bmatrix} &= 0 \end{aligned}$$

(2) Show that having zero Gauss curvature is the integrability condition for admitting Cartesian coordinates on an abstract surface. Hint: Think of U,V as 2-dimensional vectors and consider the system

$$\begin{aligned} \partial_w \begin{bmatrix} U & V \end{bmatrix} &= \begin{bmatrix} U & V \end{bmatrix}\begin{bmatrix} \Gamma^u_{wu} & \Gamma^u_{wv} \\ \Gamma^v_{wu} & \Gamma^v_{wv} \end{bmatrix}, \\ \partial_u \mathbf{q} &= U, \\ \partial_v \mathbf{q} &= V. \end{aligned}$$

These are simply the Gauss equations with the last columns and rows erased. Show that the integrability equations for U,V are $K=0$. Then use the last equations to find $\mathbf{q} : U \to \mathbb{R}^2$ after having checked the integrability conditions are satisfied. Finally, show that $(x,y) = \mathbf{q}(u,v)$ is a

Cartesian parametrization provided the correct initial conditions for U, V have be specified.

(3) Use the Codazzi equations to show that if the principal curvatures $\kappa_1 = \kappa_2$ are equal on a surface, then they are constant. Hint: In this case $L_{ij} = \kappa g_{ij}$ for some function κ.

(4) If the principal curvatures κ_1 and κ_2 are not equal on some part of the surface, then we can use theorem 4.2.7 to construct an orthogonal parametrization where the tangent fields are principal directions or said differently the coordinate curves are lines of curvature:

$$\begin{aligned} L\left(\frac{\partial \mathbf{q}}{\partial u}\right) &= \kappa_1 \frac{\partial \mathbf{q}}{\partial u}, \\ L\left(\frac{\partial \mathbf{q}}{\partial v}\right) &= \kappa_2 \frac{\partial \mathbf{q}}{\partial v}. \end{aligned}$$

Show that in this case the Codazzi equations can be written as

$$\begin{aligned} \frac{\partial \kappa_1}{\partial v} &= \frac{1}{2}(\kappa_2 - \kappa_1)\frac{\partial \ln g_{uu}}{\partial v}, \\ \frac{\partial \kappa_2}{\partial u} &= \frac{1}{2}(\kappa_1 - \kappa_2)\frac{\partial \ln g_{vv}}{\partial u}. \end{aligned}$$

(5) (Hilbert, 1901) The goal is to show: If there is a point p on a surface, where K is positive, κ_1 has a maximum, and κ_2 a minimum, then the principal curvatures are equal and constant. We assume otherwise:

$$\sup \kappa_1 = \kappa_1(p) > \kappa_2(p) = \inf \kappa_2,$$

and appeal to exercise 4 for a coordinate system around p where the coordinate curves are lines of curvature.

(a) Show that at p

$$\begin{aligned} \frac{\partial \kappa_1}{\partial u} &= \frac{\partial \kappa_1}{\partial v} = 0, \ \frac{\partial^2 \kappa_1}{\partial v^2} \le 0, \\ \frac{\partial \kappa_2}{\partial u} &= \frac{\partial \kappa_2}{\partial v} = 0, \ \frac{\partial^2 \kappa_2}{\partial u^2} \ge 0. \end{aligned}$$

(b) Using the Codazzi equations from the previous exercise to show that at p

$$\frac{\partial \ln g_{uu}}{\partial v} = 0 = \frac{\partial \ln g_{vv}}{\partial u}$$

and after differentiation also at p that

$$\frac{\partial^2 \ln g_{uu}}{\partial v^2} \ge 0, \ \frac{\partial^2 \ln g_{vv}}{\partial u^2} \ge 0.$$

(c) Next show that at p

$$K = -\frac{1}{2}\left(\frac{1}{g_{vv}}\frac{\partial^2 \ln g_{uu}}{\partial v^2} + \frac{1}{g_{uu}}\frac{\partial^2 \ln g_{vv}}{\partial u^2}\right) \le 0.$$

This contradicts our assumption about the Gauss curvature.

(d) Show that the principal curvatures are equal and constant.

(6) Show that a surface with constant principal curvatures must be part of a plane, sphere, or right circular cylinder. Note that the two former cases happen when the principal curvatures are equal.

(7) Let $\mathbf{q}(u,v)$ be a parametrized surface in $\mathbb{R}^3$. Assume E_1 and E_2 are tangent vector fields forming an orthonormal basis for the tangent space everywhere and

$$E_1 \times E_2 = \mathbf{N} = \frac{\frac{\partial \mathbf{q}}{\partial u} \times \frac{\partial \mathbf{q}}{\partial v}}{\left| \frac{\partial \mathbf{q}}{\partial u} \times \frac{\partial \mathbf{q}}{\partial v} \right|}.$$

(a) Show that

$$\begin{aligned} \frac{\partial}{\partial w} \begin{bmatrix} E_1 & E_2 & \mathbf{N} \end{bmatrix} &= \begin{bmatrix} E_1 & E_2 & \mathbf{N} \end{bmatrix} [D_w], \\ [D_w] &= \begin{bmatrix} 0 & -\phi_w & -\phi_{w1} \\ \phi_w & 0 & -\phi_{w2} \\ \phi_{w1} & \phi_{w2} & 0 \end{bmatrix}, \end{aligned}$$

where

$$\begin{aligned} \phi_w &= \frac{\partial E_1}{\partial w} \cdot E_2 = -\frac{\partial E_2}{\partial w} \cdot E_1, \\ \phi_{w1} &= \frac{\partial E_1}{\partial w} \cdot \mathbf{N} = -\mathrm{I}\left(\frac{\partial \mathbf{N}}{\partial w}, E_1 \right), \\ \phi_{w2} &= \frac{\partial E_2}{\partial w} \cdot \mathbf{N} = -\mathrm{I}\left(\frac{\partial \mathbf{N}}{\partial w}, E_2 \right). \end{aligned}$$

(b) Show that

$$\begin{aligned} \phi_{w1} &= \mathrm{II}(\partial_w \mathbf{q}, E_1) = \mathrm{I}(\partial_w \mathbf{q}, L(E_1)), \\ \phi_{w2} &= \mathrm{II}(\partial_w \mathbf{q}, E_2) = \mathrm{I}(\partial_w \mathbf{q}, L(E_2)). \end{aligned}$$

(c) Use the Weingarten equations and $[L]$ as the matrix of the Weingarten map with respect to E_1, E_2 to show that

$$[L] \begin{bmatrix} E_1 & E_2 \end{bmatrix}^t \begin{bmatrix} \frac{\partial \mathbf{q}}{\partial u} & \frac{\partial \mathbf{q}}{\partial v} \end{bmatrix} = \begin{bmatrix} \phi_{u1} & \phi_{v1} \\ \phi_{u2} & \phi_{v2} \end{bmatrix}$$

and

$$K\sqrt{\det[\mathrm{I}]} = \phi_{u1}\phi_{v2} - \phi_{u2}\phi_{v1}.$$

(d) Show that the integrability conditions

$$\frac{\partial}{\partial u}[D_v] - \frac{\partial}{\partial v}[D_u] + [D_u][D_v] - [D_v][D_u] = 0$$

can be reduced to the three equations:

$$\begin{aligned} \frac{\partial \phi_v}{\partial u} - \frac{\partial \phi_u}{\partial v} &= \phi_{u2}\phi_{v1} - \phi_{v2}\phi_{u1}, \\ \frac{\partial \phi_{v1}}{\partial u} - \frac{\partial \phi_{u1}}{\partial v} &= \phi_{v2}\phi_u - \phi_{u2}\phi_v, \\ \frac{\partial \phi_{v2}}{\partial u} - \frac{\partial \phi_{u2}}{\partial v} &= -\phi_{v1}\phi_u + \phi_{u1}\phi_v. \end{aligned}$$

(e) Show that

$$\frac{\partial \phi_v}{\partial u} - \frac{\partial \phi_u}{\partial v} = \phi_{u2}\phi_{v1} - \phi_{v2}\phi_{u1}$$

corresponds to the Gauss equation.

(f) Show that

$$\frac{\partial \phi_{v1}}{\partial u} - \frac{\partial \phi_{u1}}{\partial v} = \phi_{v2}\phi_u - \phi_{u2}\phi_v,$$
$$\frac{\partial \phi_{v2}}{\partial u} - \frac{\partial \phi_{u2}}{\partial v} = -\phi_{v1}\phi_u + \phi_{u1}\phi_v$$

correspond to the Codazzi equations.

(8) Consider potential surfaces $\mathbf{q}(u,v)$ where

$$[\mathrm{I}] = \begin{bmatrix} \lambda^2 & 0 \\ 0 & \lambda^2 \end{bmatrix}, \ [\mathrm{II}] = \begin{bmatrix} \lambda^2\kappa & 0 \\ 0 & -\frac{\lambda^2}{\kappa} \end{bmatrix}.$$

(a) Show that

$$K = -1$$

and

$$\Delta \ln \lambda = \lambda^2.$$

(b) Show that if we choose

$$\lambda = \frac{1}{a\left(u^2+v^2\right) + b_u u + b_v v + c},$$

where a, b_u, b_v, c are constants such that

$$4ac - b_u^2 - b_v^2 = -1,$$

then the first fundamental form

$$[\mathrm{I}] = \begin{bmatrix} \lambda^2 & 0 \\ 0 & \lambda^2 \end{bmatrix}$$

has $K = -1$. It can in fact be shown that there are no other possibilities for λ given that $K = -1$.

(c) Show that we only obtain a surface in space when $a = 0$ and either $b_u = 0$ or $b_v = 0$.

(d) Show that the pseudo-sphere (see section 5.4 exercise 5) is an example of such a surface with $\lambda = \frac{1}{v}$, $v > 0$.

(e) When $\lambda = \frac{1}{v}$ show that $\kappa^2 + 1 = ev^2$ for some constant $e > 0$ and conclude that κ is not defined for all v.

6.5. Gauss-Bonnet

Inspired by the idea that the integral of the curvature of a planar curve is related to how the tangent moves we can try to prove a similar result on surfaces. First we point out that we cannot expect the same theorem to hold. Consider the equator on a sphere. This curve has acceleration normal to it self and lies in the (x,y)-plane, in particular, the acceleration is also normal to the sphere and so has no geodesic curvature. On the other hand the tangent field clearly turns around 360 degrees.

Throughout this section we assume that a parametrized surface is given:

$$\mathbf{q}(u,v) : (a_u, b_u) \times (a_v, b_v) \to \mathbb{R}^3$$

where the domain is a rectangle. The key is that the domain should not have any holes in it. We further assume that we have a smaller domain

$$R \subset (a_u, b_u) \times (a_v, b_v)$$

that is bounded by a piecewise smooth curve

$$(u(s), v(s)) : [0, L] \to (a_u, b_u) \times (a_v, b_v)$$

running counter clockwise in the plane and such that $\mathbf{q}(s) == \mathbf{q}(u(s), v(s))$ is unit speed.

Integration of functions on the surface is done by defining a suitable integral using the parametrization. To make this invariant under parametrizations we define

$$\int_{\mathbf{q}(R)} f dA = \int_R f(u, v) \sqrt{\det [\mathrm{I}]} du dv = \int_R f(u, v) \left| \frac{\partial \mathbf{q}}{\partial u} \times \frac{\partial \mathbf{q}}{\partial v} \right| du dv.$$

This ensures that if we use a different parametrization (s, t) where $\mathbf{q}(Q) = \mathbf{q}(R)$, then

$$\int_R f(u, v) \sqrt{\det [\mathrm{I}]} du dv = \int_Q f(s, t) \sqrt{\det [\mathrm{I}]} ds dt.$$

We start by calculating the geodesic curvature of $\mathbf{q}$ assuming further that the parametrization gives a geodesic coordinate system

$$[\mathrm{I}] = \begin{bmatrix} 1 & 0 \\ 0 & r^2 \end{bmatrix}.$$

The existence of such coordinate systems will be established in proposition 7.4.1.

Lemma 6.5.1. *Let θ be the angle between $\mathbf{q}$ and the u-curves, then*

$$\kappa_g = \frac{d\theta}{ds} + \frac{\partial r}{\partial u} \frac{1}{r} \sin\theta.$$

Proof. We start by pointing out that the velocity is

$$\begin{aligned} \frac{d\mathbf{q}}{ds} &= \frac{du}{ds}\frac{\partial \mathbf{q}}{\partial u} + \frac{dv}{ds}\frac{\partial \mathbf{q}}{\partial v} \\ &= \cos\theta \frac{\partial \mathbf{q}}{\partial u} + \frac{1}{r}\sin\theta \frac{\partial \mathbf{q}}{\partial v}. \end{aligned}$$

The natural unit normal field to $\mathbf{q}$ in the surface is then given by

$$\mathbf{S} = -\sin\theta \frac{\partial \mathbf{q}}{\partial u} + \frac{1}{r}\cos\theta \frac{\partial \mathbf{q}}{\partial v}.$$

The geodesic curvature is then given by

$$\begin{aligned} \kappa_g &= \mathrm{I}\left(\mathbf{S}, \ddot{\mathbf{q}}^{\mathrm{I}}\right) \\ &= \mathbf{S} \cdot \left(\left(\frac{d^2 u}{ds^2} + \Gamma^u \left(\frac{d\mathbf{q}}{ds}, \frac{d\mathbf{q}}{ds} \right) \right) \frac{\partial \mathbf{q}}{\partial u} + \left(\frac{d^2 v}{ds^2} + \Gamma^v \left(\frac{d\mathbf{q}}{ds}, \frac{d\mathbf{q}}{ds} \right) \right) \frac{\partial \mathbf{q}}{\partial v} \right) \\ &= -\sin\theta \left(\frac{d^2 u}{ds^2} + \Gamma^u \left(\frac{d\mathbf{q}}{ds}, \frac{d\mathbf{q}}{ds} \right) \right) + r^2 \frac{1}{r}\cos\theta \left(\frac{d^2 v}{ds^2} + \Gamma^v \left(\frac{d\mathbf{q}}{ds}, \frac{d\mathbf{q}}{ds} \right) \right) \\ &= -\sin\theta \left(\frac{d^2 u}{ds^2} + \Gamma^u \left(\frac{d\mathbf{q}}{ds}, \frac{d\mathbf{q}}{ds} \right) \right) + r\cos\theta \left(\frac{d^2 v}{ds^2} + \Gamma^v \left(\frac{d\mathbf{q}}{ds}, \frac{d\mathbf{q}}{ds} \right) \right). \end{aligned}$$

We further have

$$\begin{aligned}
\frac{d^2u}{ds^2} &= \frac{d\cos\theta}{ds} = -\sin\theta\frac{d\theta}{ds},\\
\frac{d^2v}{ds^2} &= \frac{d\frac{1}{r}\sin\theta}{ds}\\
&= \frac{-1}{r^2}\frac{dr}{ds}\sin\theta + \frac{1}{r}\cos\theta\frac{d\theta}{ds}\\
&= \frac{-1}{r^2}\left(\frac{\partial r}{\partial u}\frac{du}{ds} + \frac{\partial r}{\partial v}\frac{dv}{ds}\right)\sin\theta + \frac{1}{r}\cos\theta\frac{d\theta}{ds}\\
&= \frac{-1}{r^2}\frac{\partial r}{\partial u}\cos\theta\sin\theta + \frac{-1}{r^3}\frac{\partial r}{\partial v}\sin^2\theta + \frac{1}{r}\cos\theta\frac{d\theta}{ds}.
\end{aligned}$$

And the Christoffel symbols are

$$\begin{aligned}
\Gamma^u\left(\frac{d\mathbf{q}}{ds},\frac{d\mathbf{q}}{ds}\right) &= \Gamma^u_{vv}\left(\frac{dv}{ds}\right)^2\\
&= -r\frac{\partial r}{\partial u}\frac{1}{r^2}\sin^2\theta\\
&= \frac{-1}{r}\frac{\partial r}{\partial u}\sin^2\theta,
\end{aligned}$$

$$\begin{aligned}
\Gamma^v\left(\frac{d\mathbf{q}}{ds},\frac{d\mathbf{q}}{ds}\right) &= 2\Gamma^v_{uv}\frac{du}{ds}\frac{dv}{ds} + \Gamma^v_{vv}\left(\frac{dv}{ds}\right)^2\\
&= \frac{2}{r}\frac{\partial r}{\partial u}\frac{du}{ds}\frac{dv}{ds} + \frac{1}{r}\frac{\partial r}{\partial v}\left(\frac{dv}{ds}\right)^2\\
&= \frac{2}{r^2}\frac{\partial r}{\partial u}\sin\theta\cos\theta + \frac{1}{r^3}\frac{\partial r}{\partial v}\sin^2\theta.
\end{aligned}$$

Thus

$$\begin{aligned}
\kappa_g &= -\sin\theta\left(-\sin\theta\frac{d\theta}{ds} - \frac{1}{r}\frac{\partial r}{\partial u}\sin^2\theta\right) + r\cos\theta\left(\frac{1}{r}\cos\theta\frac{d\theta}{ds} + \frac{1}{r^2}\frac{\partial r}{\partial u}\sin\theta\cos\theta\right)\\
&= \frac{d\theta}{ds} + \frac{1}{r}\frac{\partial r}{\partial u}\sin^3\theta + \frac{1}{r}\frac{\partial r}{\partial u}\sin\theta\cos^2\theta\\
&= \frac{d\theta}{ds} + \frac{\partial r}{\partial u}\frac{1}{r}\sin\theta.
\end{aligned}$$

□

We first prove the local Gauss-Bonnet theorem. It is stated in the way that Gauss and Bonnet proved it. Gauss considered regions bounded by geodesics thus eliminating the geodesic curvature, while Bonnet presented the version given below.

THEOREM 6.5.2. *(Gauss, 1825 and Bonnet, 1848) Assume as in the above lemma that the parametrization gives a geodesic coordinate system. Let θ_i be the exterior angles at the points where $\mathbf{q}$ has vertices, then*

$$\int_{\mathbf{q}(R)} KdA + \int_0^L \kappa_g ds = 2\pi - \sum\theta_i.$$

PROOF. We have that

$$\begin{aligned}
\int_{\mathbf{q}(R)} K dA &= \int_R K\sqrt{\det[\mathrm{I}]}dudv \\
&= -\int_R \frac{\frac{\partial^2 r}{\partial u^2}}{r} r dudv \\
&= -\int_R \frac{\partial^2 r}{\partial u^2} dudv.
\end{aligned}$$

The last integral can be turned into a line integral if we use Green's theorem

$$\int_R \frac{\partial^2 r}{\partial u^2} dudv = \int_{\partial R} \frac{\partial r}{\partial u} dv.$$

This line integral can now be recognized as one of the terms in the formula for the geodesic curvature

$$\begin{aligned}
\int_{\partial R} \frac{\partial r}{\partial u} dv &= \int_0^L \frac{\partial r}{\partial u}\frac{dv}{ds} ds \\
&= \int_0^L \frac{\partial r}{\partial u}\frac{1}{r}\sin\theta ds \\
&= \int_0^L \left(\kappa_g - \frac{d\theta}{ds}\right) ds \\
&= \int_0^L \kappa_g ds - \int_0^L \frac{d\theta}{ds} ds.
\end{aligned}$$

Thus we obtain

$$\begin{aligned}
\int_{\mathbf{q}(R)} K dA + \int_0^L \kappa_g ds &= -\int_R \frac{\partial^2 r}{\partial u^2} dudv + \int_{\partial R} \frac{\partial r}{\partial u} dv + \int_0^L \frac{d\theta}{ds} ds \\
&= \int_0^L \frac{d\theta}{ds} ds.
\end{aligned}$$

Finally we must show that

$$\int_0^L \frac{d\theta}{ds} ds + \sum \theta_i = 2\pi.$$

For a planar simple closed curve this is a consequence of knowing that the rotation index must be 1 for such curves if they are parametrized to run counterclockwise (see theorem 2.4.5 and section 2.4 exercise 3). In this case we still know that the right hand side must be a multiple of 2π. The trick now is to compute the right hand side for each of the abstract metrics

$$[\mathrm{I}_\epsilon] = \begin{bmatrix} 1 & 0 \\ 0 & 1-\epsilon+\epsilon r^2 \end{bmatrix}.$$

For each $\epsilon \in [0,1]$ this defines a metric on $(a_u, b_u) \times (a_v, b_v)$ and the rotation index for our curve has to be a multiple 2π. It is easy to see that the angles θ_ϵ between the curve and the u-curves is continuous in ϵ. Thus the right hand side also varies continuously. However, as it is always a multiple of 2π and is 2π in case $\epsilon = 0$ it follows that it is always 2π. $\square$

Clearly there are subtle things about the regions R we are allowed to use. Aside from the topological restriction on R there is also an orientation choice (counter clockwise) for ∂R in Green's theorem. If we reverse that orientation there is a sign change, and the geodesic curvature also changes sign when we run backwards.

We used rather special coordinates as well, but it is possible to extend the proof to work for all coordinate systems. The same strategy even works, but is complicated by the nasty formula we have for the Gauss curvature in general coordinates. However, if we used a conformal or isothermal parametrization then the argument about the winding is much simpler as angles would be the same in the plane and on the surface. Thus the winding number is clearly 1.

Using Cartan's approach with selecting orthonormal frames rather than special coordinates makes for a fairly simple proof that works within all coordinate systems. This is exploited in an exercise below, but to keep things in line with what we have already covered we still restrict attention to how this works in relation to a parametrization.

Let us now return to our examples from above. Without geodesic curvature and exterior angles we expect to end up with the formula

$$\int_{\mathbf{q}(R)} KdA = 2\pi.$$

But there has to be a region R bounding the closed geodesic. On the sphere we can clearly use the upper hemisphere. As $K = 1$ we end up with the well known fact that the upper hemisphere has area 2π. On the cylinder, however, there is no reasonable region bounding the closed geodesics despite the fact that we have a valid geodesic coordinate system. The issue is that the bounding curve cannot be set up to be a closed curve in a parametrization where there is a rectangle containing the curve.

It is possible to modify the Gauss-Bonnet formula so that more general regions can be used in the statement, but it requires topological information about the region R. This will be studied in detail later and also in some interesting cases in the exercises below.

Another very important observation about our proof is that it only referred to quantities related to the first fundamental form. In fact, the result holds without further ado for generalized surfaces and abstract surfaces as well, again with the proviso of working within coordinates and regions without holes.

It is, however, possible to also get the second fundamental form into the picture if we recall that

$$K\left|\frac{\partial \mathbf{q}}{\partial u}\times\frac{\partial \mathbf{q}}{\partial v}\right| = \left(\frac{\partial \mathbf{N}}{\partial u}\times\frac{\partial \mathbf{N}}{\partial v}\right)\cdot \mathbf{N} = \pm\left|\frac{\partial \mathbf{N}}{\partial u}\times\frac{\partial \mathbf{N}}{\partial v}\right|.$$

Thus $\int_R KdA$ also measures the signed area of the spherical image traced by the normal vector, or the image of the Gauss map.

Exercises.

(1) Consider a surface of revolution and two latitudes $\mathbf{q}_1$ and $\mathbf{q}_2$ on it. These curves bound a band or annular region $\mathbf{q}(R)$. By subdividing the region and using proper orientations and parametrizations on the curves show that

$$\int_{\mathbf{q}(R)} KdA = \int_{\mathbf{q}_1} \kappa_g ds_1 - \int_{\mathbf{q}_2} \kappa_g ds_2.$$

(2) Generalize the previous exercise to regions that are bounded both on the inside and outside by smooth (or even piecewise smooth) closed curves.
(3) Consider a geodesic triangle inside a parametrization with interior angles α, β, γ. Show that
$$\int_{\mathbf{q}(R)} K dA = \alpha + \beta + \gamma - \pi.$$
(4) Show that a closed surface with $\chi > 0$ must have a point with $K > 0$.
(5) Let $\mathbf{q}(u, v)$ be a parametrized surface without special assumptions about the parametrization. Create tangent vector fields E_1 and E_2 forming an orthonormal basis for the tangent space everywhere with the further property that E_1 is proportional to the first tangent field $\frac{\partial \mathbf{q}}{\partial u}$ and
$$E_1 \times E_2 = \mathbf{N} = \frac{\frac{\partial \mathbf{q}}{\partial u} \times \frac{\partial \mathbf{q}}{\partial v}}{\left| \frac{\partial \mathbf{q}}{\partial u} \times \frac{\partial \mathbf{q}}{\partial v} \right|}.$$
(a) Use section 6.4 exercise 7 to conclude that
$$\begin{aligned} \int_{\mathbf{q}(R)} K dA &= -\int_R \left(\frac{\partial \phi_v}{\partial u} - \frac{\partial \phi_u}{\partial v} \right) du dv \\ &= -\int_{\mathbf{q}} \phi_u du + \phi_v dv. \end{aligned}$$
(b) Finally prove the Gauss-Bonnet theorem by establishing
$$\int_{\mathbf{q}} \phi_u du + \phi_v dv = \int \left(k_g - \frac{d\theta}{ds} \right) ds,$$
where θ is the angle with E_1 or $\frac{\partial \mathbf{q}}{\partial u}$. To aid the last calculation show that
$$\begin{aligned} \frac{d\mathbf{q}}{ds} &= \cos\theta E_1 + \sin\theta E_2, \\ \mathbf{S} &= -\sin\theta E_1 + \cos\theta E_2, \\ \frac{d^2\mathbf{q}}{ds^2} &= \mathbf{S}\frac{d\theta}{ds} - \sin\theta\left(\cos\theta\phi_u + \sin\theta\phi_v\right) E_1 \\ &\quad + \cos\theta\left(\cos\theta\phi_u + \sin\theta\phi_v\right) E_2 + a\mathbf{N}, \end{aligned}$$
where the coefficient a in front of $\mathbf{N}$ is irrelevant for computing the inner product with $\mathbf{S}$ and hence the geodesic curvature.

6.6. Topology of Surfaces

So far we have only worked with the geometry of surfaces. When studying the global behavior of closed surfaces there are also some interesting numerical topological concepts that are important in our geometric understanding of these surfaces.

DEFINITION 6.6.1. A polygon, or n-gon is a piecewise smooth simple closed curve inside a rectangular parameterization as in the previous section. The number $n = 0, 1, 2, ...$ refers to the number of points where the curve is not differentiable. We call these points *vertices* of the closed curve and the connecting arcs the *edges*.

REMARK 6.6.2. The edges will not include the two boundary points, thus they are simple smooth curves defined on an open interval. A 0-gon is a smooth simple closed curve, it has 0 vertices and 0 edges. A 1-gon is a loop with one vertex and one edge etc. The inside is well defined by the Jordan curve theorem (see theorem 2.3.3) and is called the *face*. Thus the face of an n-gon, $n > 0$, has a boundary that consists of n vertices and n edges, each of the edges in turn has two vertices as boundary points. In the special case of a 0-gon the boundary is a smooth circle. Note that the inside of a circle or regular n-gon in the plane is homeomorphic to an open disc. Thus any face is homeomorphic to an open disc.

DEFINITION 6.6.3. A *polygonal subdivision* of an abstract surface is a disjoint decomposition of the surface into faces and their boundaries. More specifically, if a point lies inside a face, then it cannot be in any other face or on the boundary of any face. If a point is a vertex for one face, then it cannot lie on an edge of any other face, but it can be a vertex for several other faces.

REMARK 6.6.4. If a point lies on an edge of one face, then it can only lie on edges of other faces. In fact it can only lie on the edge for one other face since such a point will have a neighborhood that is homeomorphic to a disc; if 3 or more faces meet in a common edge, no point on that edge has a neighborhood that is homeomorphic to a disc.

DEFINITION 6.6.5. A *triangulation* is a polygonal subdivision into triangles (3-gons) with the added condition that two faces can have at most one edge in common. Note that one can subdivide the sphere into two triangles as in a triangular pillow, but this is not a triangulation. The tetrahedron is a triangulation of the sphere and in fact the triangulation with the smallest number of vertices, edges and faces.

In any given concrete situation it is not hard to find a triangulation, but for an abstract surface this is much less easy to see. We will take it for granted that our surfaces have polygonal subdivisions and triangulations. A polygonal subdivision in fact creates a triangulation if we add a vertex in each face and connect it to the vertices of the face with edges. A polygonal subdivision can be created using some of the geometric developments in the next chapter. One has to find a finite covering of small sets B_j that have boundaries with positive geodesic curvature. These form a big Venn-type subdivision of the surface. If the sets are chosen appropriately this will also be a polygonal subdivision.

DEFINITION 6.6.6. The *Euler characteristic* of a polygonal subdivision is defined as the alternating sum: $\chi = V - E + F$ where F is the number of faces and E, V the number of edges and vertices not counted with multiplicity.

EXAMPLE 6.6.7. If we take a smooth simple closed curve on a sphere then we obtain a polygonal subdivision where $F = 2$, $E = 0$, and $V = 0$. If we triangulate the sphere using the tetrahedron then $F = 4$, $E = 6$, and $V = 4$. In either case $\chi = 2$.

We will first use geometry to show that the Euler characteristic is a numerical topological invariant of the surface. Below we indicate how closed oriented surfaces are classified and how the Euler characteristic is constrained to be ≤ 2.

THEOREM 6.6.8. *Let M be an oriented closed surface, then*

$$\int_M K dA = 2\pi\chi,$$

for any polygonal subdivision of M. In particular, χ does not depend on the polygonal subdivision.

PROOF. The orientation is used so that integration has a consistent sign when we switch parametrizations.

We consider a polygonal subdivision with F polygons. Each n_j-gon is denoted by P_j. The local version of Gauss-Bonnet for each polygon can be written:

$$\begin{aligned}\int_{P^j} KdA &= -\int_0^{L_j} \kappa_g ds + 2\pi - \sum_{i_j=1}^{n_j} \theta_{i_j} \\ &= -\int_0^{L_j} \kappa_g ds + 2\pi - \pi n_j + \sum_{i_j=1}^{n_j} \alpha_{i_j},\end{aligned}$$

where α_{i_j} is the interior angle. The global formula is now gotten by adding up these contributions. When doing this it is important to orient each polygon so that the winding number is 1. Each edge occurs in exactly two adjacent polygons, but the edge will have the opposite orientation in each of the polygons when we insist that they both have winding number 1. Thus the geodesic curvature changes sign and those terms cancel each other in the sum.

$$\begin{aligned}\int_M KdA &= \sum_{j=1}^{F} \int_{P_j} KdA \\ &= 2\pi F - \sum_{j=1}^{F} \pi n_j + \sum_{j=1}^{F} \sum_{i_j=1}^{n_j} \alpha_{i_j} \\ &= 2\pi (F - E + V).\end{aligned}$$

Here the last equality follows from the fact that at each vertex the interior angles add up to 2π, while $n_1 + \cdots + n_F = 2E$ since each edge gets counted twice in that sum.

This shows that $F - E + V$ does not depend on what subdivision we picked. Given that information we observe that $\int_M KdA$ does not vary if we change the first fundamental form on a given abstract surface as we can always use the same subdivision regardless of what the first fundamental form is. $\square$

DEFINITION 6.6.9. The genus g of an orientable closed surface is the maximum number of disjoint simple closed curves whose complement is connected. Orientability is used to guarantee that any simple closed curve has a well-defined right and and left hand side, i.e., it locally divides the surface in two. Globally, the complement might still be connected. Note that the Jordan curve theorem implies that $g = 0$ for the sphere.

Using g surgeries (see proof below) one can obtain a closed surface with $g = 0$. We shall be concerned with the opposite question: What can we say about a closed oriented surface with $g = 0$ and more generally about a surface with genus g?

It is easy to construct surfaces of genus g by adding g handles to a sphere. The next theorem explains why there are no other orientable surfaces.

THEOREM 6.6.10. *An oriented surface with genus g has $\chi = 2 - 2g$ and is a sphere with g handles attached.*

Proof. We will fix a triangulation for a closed oriented surface. A *simple cycle* in a triangulation is a simple closed loop of edges, i.e. each vertex and edge only appears once as we run around in the loop. We can then redefine the genus as the maximum number of simple cycles whose complement is connected.

Surgery for a triangulation is defined by cutting along a simple cycle whose complement is connected and adding two pyramids to create a new surface with a triangulation. This reduces the genus and increases χ by 2. The latter is because the simple cycle has the same number of edges and vertices and thus does not contribute to χ. For each pyramid we add the same number of faces and edges and 1 vertex. Thus χ is increased by 1 for each of the two pyramids. Now g such surgeries will result in a triangulated surface with $g = 0$ where χ has been reduced by $2g$. In case we have a triangulated surface with $g = 0$ it is believable that it must be a sphere and we will give the proof below. If we reverse the surgeries, then we are adding handles to the sphere. Thus showing that the original surface was a sphere with g handles.

Recall that faces and edges do not include their boundary points. Consider a collection of faces, edges, and vertices whose union is homeomorphic to an open disc and has $\chi = V - E + F = 1$. The boundary consists of the edges and vertices that meet the faces in the collection, but are excluded from being part of the union. Since the collection is an open set it can't contain a vertex without also including all edges and faces that have the vertex on their boundaries. However, it is possible for it to contain two adjacent faces without the common edge. In particular, such a collection could contain all faces in the triangulation and still have nonempty boundary. Note that in defining χ for such a collection we only count the vertices and edges included, not the remaining vertices and edges that meet the faces, those are included in the boundary. The simplest example of such a collection is a single face.

The claim is that any surface contains a collection that forms an open disc with $\chi = 1$; includes all faces in the triangulation; and such that the boundary graph is connected and has no branches, i.e., there are no vertices that are met by just one edge.

Consider any collection whose union is an open disc with $\chi = 1$ and whose boundary is connected.

Faces outside this collection that meet the boundary either do so in one, two, or three edges. Regardless of which situation occurs we can add the face and exactly one of the edges that is also an edge for a face in the collection. This keeps the properties that the collection forms an open disc with $\chi = 1$. To be specific, note that the open half-disc

$$H = \left\{ (x, y) \in (-\infty, 0) \times \mathbb{R} \mid \sqrt{x^2 + y^2} < 1 \right\}$$

and when with a wedge added

$$H \cup \{(x, y) \in [0, 1] \times \mathbb{R} \mid |y| < a\,(1 - x)\,,\, 0 < a \leq 1\}$$

are both homeomorphic to open discs. Note that the boundary still has all of the original vertices, one edge is deleted, and the other two edges and vertex of the added face are added. Thus the boundary stays connected. Now continue this process until all faces in the triangulation of the surface have been included.

Next we eliminate branches from the boundary. If the boundary contains a vertex that is met by exactly one edge, then add the vertex and edge to the collection.

This keeps the properties that the collection forms an open disc with $\chi = 1$. To be specific, note that all of the open sets $\left\{(x,y) \in \mathbb{R}^2 \mid \sqrt{x^2+y^2} < 1,\ x \geq a > -1\right\}$ are homeomorphic to open discs. When we delete a vertex and edge, we are essentially just increasing a. Clearly the boundary stays connected. Continue this until there are no branches on the boundary.

We can now characterize the sphere as the only surface with $g = 0$. In this case the boundary of the open disc with $\chi = 1$ that includes all faces can't contain any simple cycles since the complement of the boundary is the open disc and hence connected. If there are no branches, then it can only be a single vertex. This implies that $\chi = 2$ and that the surface is a sphere. □

Exercises.

(1) Show that for a triangulation of a closed surface we have:
   (a) $E \leq \binom{V}{2}$,
   (b) $E \leq \binom{F}{2}$,
   (c) $2E = 3F$,
   (d) $E = 3(V - \chi)$,
   (e) $F \geq V$,
   (f) $V \geq \frac{1}{2}\left(7 + \sqrt{1+48g}\right)$,
   (g) When $g = 0$ show that at least one vertex has degree ≤ 5. The degree of a vertex is the number of edges that meet the vertex.
   (h) When $g \geq 1$ show that at least one vertex has degree $\leq \frac{1}{2}\left(7 + \sqrt{1+48g}\right) - 1$.

   The number $\frac{1}{2}\left(7 + \sqrt{1+48g}\right)$ is also known as the coloring number of the of the surface. The fact that any map on a surface can be colored with at most that many colors is the famous 4-coloring conjecture/theorem for the sphere. Heawood established the result for surfaces of genus $g \geq 1$ by showing that (h) holds. The same method shows that (g) implies that maps on the sphere can be 6 colored. Heawood also showed that maps on the sphere can be 5 colored. It was not until 1968 that Ringel and Youngs showed that this is the correct coloring number when $g \geq 1$. The 4 color problem ($g = 0$) was solved by Appel and Haken in 1977.

6.7. Closed and Convex Surfaces

PROPOSITION 6.7.1. *A closed surface $M \subset \mathbb{R}^3$ has the property that the Gauss map is onto. There are no closed space-like surfaces $M \subset \mathbb{R}^{2,1}$.*

PROOF. The proof in either case consider $f(p) = p \cdot n$ for a fixed $n \in \mathbb{R}^3$. At a maximum point this vector will be the normal vector.

In the case of $M \subset \mathbb{R}^{2,1}$ it follows that there will be both time-like and space-like normal vectors. That's impossible if all tangent spaces are space-like as that forces the normals to be time-like. □

PROPOSITION 6.7.2. *A closed surface $M \subset \mathbb{R}^3$ has points where both principal curvatures are positive.*

PROOF. Consider $f(p) = |p|^2$. Take a maximum and show that both principal curvatures are bigger than $1/|p|$ at such points (see also 2.5.3.) □

Abstract surface with $g = 0$ also has points with $K > 0$ by G-B.

THEOREM 6.7.3. *(Liebmann, 1900) If $M \subset \mathbb{R}^3$ is closed and has constant Gauss curvature, then it is a constant curvature sphere.*

PROOF. First note that the surface must has positive curvature. Next observe that since the surface is closed and $K = \kappa_1 \cdot \kappa_2$ is constant. It follows that when κ_1 has a minimum, then κ_2 has a maximum. Hilbert's lemma (see section 6.4 exercise 5) then tells us that the principal curvatures must be equal and constant. □

THEOREM 6.7.4. *If $M \subset \mathbb{R}^3$ is closed and has constant mean curvature, then it is a constant curvature sphere.*

PROOF. Same proof as above. □

THEOREM 6.7.5. *(Hadamard, 1897) Let $M \subset \mathbb{R}^3$ be a closed surface with $K > 0$, then the Gauss map is a diffeomorphism and M is convex.*

PROOF. First observe that $\mathbf{N} : M \to S^2(1)$ is has nonsingular differential everywhere as $\det DN = K > 0$. The global Gauss-Bonnet theorem tells us that

$$0 < \int_M K dA = 2\pi\chi(M).$$

This implies that $\chi(M) = 2$.

Show $\mathbf{N}$ is onto by above proposition

Show $\mathbf{N}$ is one-to-one by contradiction as otherwise we can find a small open set $O \subset M$ such that $\mathbf{N}$ is onto when restricted to $M - O$. This implies

$$4\pi = \int_{M-O} K dA + \int_O K dA = 4\pi + \int_O K dA > 4\pi.$$

Consider the signed height function to the tangent plane at a point $p \in M$:

$$f(x) = (x - p) \cdot \mathbf{N}_p$$

This has exactly two critical points where $\mathbf{N}_x = \pm \mathbf{N}_p$. These correspond to the maximum and minimum. Assume p is the minimum. Then $f(x) > 0$ for all $x \neq p$. □

THEOREM 6.7.6. *Any two simple closed geodesics on a closed surface with $K > 0$ intersect.*

PROOF. If they don't intersect then there is an annular region with $K > 0$ where the boundary curves have no geodesic curvature. This violates G-B. See also theorem 7.8.4 for a different proof. □

REMARK 6.7.7. One can reprove the results in this section for isometric immersions $F : M \to \mathbb{R}^3$ when M is oriented. In particular, it will follow that all such immersions are embeddings when $K > 0$.

CHAPTER 7

Geodesics and Metric Geometry

This chapter covers the basics of geodesics and their properties as shortest curves. We also give models for constant curvature spaces and calculate the geodesics in these models. We discuss isometries and the local/global classification of surfaces with constant Gauss curvature. The chapter ends with a treatment of a few classical comparison theorems. Virtually all results have analogues for higher dimensional Riemannian manifolds, but certain proofs are a bit easier for surfaces. It will be noted that there is no mention of parallel translation although we do introduce second partial derivatives for 2-parameter maps in to an abstract surface. This is more or less in line with the classical treatment as parallel translation was not introduced until the early part of the 20th century. It also eases the treatment quite a bit.

Throughout we study abstract surfaces, but note that many calculations are much easier if we think of the surfaces as sitting in $\mathbb{R}^3$.

7.1. Geodesics

Definition 7.1.1. A curve $\mathbf{q}$ on a surface M is called a *geodesic* if the tangential part of the acceleration vanishes $\ddot{\mathbf{q}}^{\mathrm{I}} = 0$, or specifically

$$\begin{aligned} \frac{d^2u}{dt^2} + \Gamma^u(\dot{\mathbf{q}}, \dot{\mathbf{q}}) &= 0, \\ \frac{d^2v}{dt^2} + \Gamma^v(\dot{\mathbf{q}}, \dot{\mathbf{q}}) &= 0. \end{aligned}$$

When $M \subset \mathbb{R}^3$ this is equivalent to saying that $\ddot{\mathbf{q}}$ is normal to the surface or that $\ddot{\mathbf{q}} = \ddot{\mathbf{q}}^{\mathrm{II}} = \mathbf{N}\mathrm{II}(\dot{\mathbf{q}}, \dot{\mathbf{q}})$.

Proposition 7.1.2. *A geodesic has constant speed.*

Proof. Let $\mathbf{q}(t)$ be a geodesic. We compute the derivative of the square of the speed:

$$\frac{d}{dt}\mathrm{I}(\dot{\mathbf{q}}, \dot{\mathbf{q}}) = \frac{d}{dt}(\dot{\mathbf{q}} \cdot \dot{\mathbf{q}}) = 2\ddot{\mathbf{q}} \cdot \dot{\mathbf{q}} = 2\mathrm{II}(\dot{\mathbf{q}}, \dot{\mathbf{q}})\,\mathbf{N} \cdot \dot{\mathbf{q}} = 0$$

since $\mathbf{N}$ and $\dot{\mathbf{q}}$ are perpendicular. Thus $\mathbf{q}$ has constant speed.

There is also a purely intrinsic proof that works for abstract surfaces. Since it is convenient to do this proof in a more general context it will be delayed until the end of the next section. □

Next we address existence of geodesics.

Theorem 7.1.3. *Given a point $p = \mathbf{q}(u_0, v_0)$ and a tangent vector $V = V^u \frac{\partial \mathbf{q}}{\partial u}(u_0, v_0) + V^v \frac{\partial \mathbf{q}}{\partial v}(u_0, v_0) \in T_pM$ there is a unique geodesic $\mathbf{q}(t) = \mathbf{q}(u(t), v(t))$*

defined on some small interval $t \in (-\varepsilon, \varepsilon)$ with the initial values

$$\begin{aligned}\mathbf{q}(0) &= p,\\ \dot{\mathbf{q}}(0) &= V.\end{aligned}$$

Proof. The existence and uniqueness part is a very general statement about solutions to differential equations (see theorem A.5.1). In this case we note that in the (u, v) parameters we must solve a system of second order equations

$$\begin{aligned}\frac{d^2u}{dt^2} &= -\begin{bmatrix} \frac{du}{dt} & \frac{dv}{dt} \end{bmatrix}\begin{bmatrix} \Gamma^u_{uu} & \Gamma^u_{uv} \\ \Gamma^u_{vu} & \Gamma^u_{vv} \end{bmatrix}\begin{bmatrix} \frac{du}{dt} \\ \frac{dv}{dt} \end{bmatrix},\\ \frac{d^2v}{dt^2} &= -\begin{bmatrix} \frac{du}{dt} & \frac{dv}{dt} \end{bmatrix}\begin{bmatrix} \Gamma^v_{uu} & \Gamma^v_{uv} \\ \Gamma^v_{vu} & \Gamma^v_{vv} \end{bmatrix}\begin{bmatrix} \frac{du}{dt} \\ \frac{dv}{dt} \end{bmatrix},\end{aligned}$$

with the initial values

$$\begin{aligned}(u(0), v(0)) &= (u_0, v_0),\\ (\dot{u}(0), \dot{v}(0)) &= (V^u, V^v).\end{aligned}$$

As long as the Christoffel symbols are sufficiently smooth there is a unique solution to such a system of equations given the initial values. The domain $(-\varepsilon, \varepsilon)$ on which such a solution exists is quite hard to determine. It'll depend on the domain of parameters U, the initial values, and Christoffel symbols. □

This theorem allows us to find all geodesics on spheres and in the plane without calculation.

Example 7.1.4. In $\mathbb{R}^2$ straight lines $\mathbf{q}(t) = p + vt$ are clearly geodesics. Since these solve all possible initial problems there are no other geodesics.

Example 7.1.5. On S^2 we claim that the great circles

$$\begin{aligned}\mathbf{q}(t) &= q\cos(|v|\,t) + \frac{v}{|v|}\sin(|v|\,t),\\ q &\in S^2,\\ q\cdot v &= 0\end{aligned}$$

are geodesics. Note that this is a curve on S^2, and that $\mathbf{q}(0) = q$, $\dot{\mathbf{q}}(0) = v$. The acceleration as computed in $\mathbb{R}^3$ is given by

$$\ddot{\mathbf{q}}(t) = -q\,|v|^2\cos(|v|\,t) - v\,|v|\sin(|v|\,t) = -|v|^2\,\mathbf{q}(t)$$

and is consequently normal to the sphere. In particular $\ddot{\mathbf{q}}^{\mathrm{I}} = 0$. This means that we have also solved all initial value problems on the sphere.

Depending on our parametrization (u, v)-geodesics can be pictured in many ways. We'll study a few models or parametrizations of the sphere where geodesics take on some familiar shapes and can be described directly by equations rather than in parametrized form.

Unit Sphere Model: Consider the sphere where great circles and hence geodesics are described by the two equations:

$$\begin{aligned}ax + by + cz &= 0,\\ x^2 + y^2 + z^2 &= 1.\end{aligned}$$

Given a specific geodesic $\mathbf{q}(t) = q\cos(|v|\,t) + \frac{v}{|v|}\sin(|v|\,t)$ we can use $(a, b, c) = q \times v$.

Elliptic Model: If we use the Monge patch $\left(u, v, \sqrt{1-u^2-v^2}\right)$ on the upper hemisphere, i.e., project to the (x, y)-plane along the z-axis, then the equations of the geodesics become

$$\left(a^2+c^2\right) u^2 + 2abuv + \left(b^2+c^2\right) v^2 = c^2.$$

These are the equations of ellipses whose axes go through the origin and are inscribed in the unit circle. This is how you draw great circles on the sphere!

Beltrami Model: If we use the parametrization

$$\frac{1}{\sqrt{1+u^2+v^2}}\left(u, v, 1\right)$$

on the upper hemisphere, i.e., $\frac{x}{z} = u$, $\frac{y}{z} = v$, then these equations simply become straight lines in (u, v) coordinates:

$$au + bv + c = 0.$$

This reparametrization was also discussed in section 4.5 exercise 8, where it was called the Beltrami projection. It is simply the projection of the upper hemisphere along radial lines to the tangent plane $\{z = 1\}$ at the North pole.

Conformal Model: The radial projection that was used for the Beltrami model is an example of a perspective projection, i.e., a projection along radial lines from a point to a plane that does not pass through this point. The stereographic parametrization from section 4.5 exercise 5 is projection along lines through $(0, 0, 1)$ to the (x, y)-plane. In this model the upper hemisphere is parametrized as

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

One can show that this is a conformal or isothermal parametrization. The geodesics are either straight lines through the origin:

$$au + bv = 0$$

or when $c \neq 0$ we can normalize so that $c = 1$ in which case the geodesics become circles

$$\left(u+a\right)^2 + \left(v+b\right)^2 = 1 + a^2 + b^2.$$

Next we consider hyperbolic space.

Imaginary Unit Sphere Model: We defined hyperbolic space $H \subset \mathbb{R}^{2,1}$ in example 6.2.2 as the the imaginary unit sphere with $z > 0$, specifically it is the rotationally symmetric surface

$$x^2 + y^2 - z^2 = -1,\ z \geq 1$$

with a metric that is inherited from the space-time inner product structure. Observe that the tangent space can be characterized as

$$T_qM = \left\{v \in \mathbb{R}^{2,1} \mid v \cdot q = 0\right\}.$$

This means that the normal is be given by $\mathbf{N}\left(q\right) = q$. In analogy with the sphere we consider the curves

$$\begin{aligned}
\mathbf{q}\left(t\right) &= q\cosh\left(\left|v\right| t\right) + \frac{v}{\left|v\right|}\sinh\left(\left|v\right| t\right), \\
q &\in H, \\
v &\in T_qH.
\end{aligned}$$

Since $q \cdot v = 0$ this is a curve on H with $\mathbf{q}(0) = q$, $\dot{\mathbf{q}}(0) = v$. Note also that it lies in the plane spanned by q and v.

The acceleration as computed in $\mathbb{R}^{2,1}$ is given by

$$\ddot{\mathbf{q}}(t) = q\,|v|^2 \cosh(|v|\,t) + v\,|v| \sinh(|v|\,t) = |v|^2\,\mathbf{q}(t).$$

In particular, it has no tangential component and thus has vanishing intrinsic acceleration (see also remark 6.3.3).

If we use $(a, b, c) = q \times v$, then we also obtain the equation form:

$$\begin{aligned} ax + by + cz &= 0, \\ x^2 + y^2 - z^2 &= -1,\ z \geq 1. \end{aligned}$$

Note that for these planes to intersect the surface it is necessary to assume that:

$$c^2 < a^2 + b^2.$$

Hyperbolic Model: This is the orthogonal projection onto the (x, y)-plane. The parametrization is a Monge patch and is given by $\left(u, v, \sqrt{1 + u^2 + v^2}\right)$. The geodesics will be straight lines through the origin when $c = 0$ and hyperbolas whose asymptotes are lines through the origin when $0 < c^2 < a^2 + b^2$:

$$\left(a^2 - c^2\right) u^2 + 2abuv + \left(b^2 - c^2\right) v^2 = c^2.$$

Recall that the level sets to quadratic equations:

$$\alpha x^2 + 2\beta xy + \gamma y^2 = R^2$$

are ellipses centered at the origin when $\alpha\gamma - \beta^2 > 0$ and hyperbolas with asymptotes that pass through the origin when $\alpha\gamma - \beta^2 < 0$.

Beltrami Model: The Beltrami model comes from a perspective projection along radial lines through the origin to the plane $z = 1$. It gives us the parametrization

$$\frac{1}{\sqrt{1 - u^2 - v^2}}\,(u, v, 1)\,,\ u^2 + v^2 < 1.$$

And the geodesics are straight lines:

$$au + bv + c = 0.$$

Conformal Models: Stereographic projection along radial lines through $(0, 0, 1)$ to the (x, y)-plane gives the *Poincaré model*. The parametrization is given by:

$$\frac{1}{1 - u^2 - v^2}\left(2u, 2v, -\left(1 + u^2 + v^2\right)\right),\ u^2 + v^2 < 1.$$

It is also called the unit disc model since the open disc is the domain for the parameters. One can show that this parametrization is conformal or isothermal. The geodesics are either straight lines through the origin

$$au + bv = 0$$

or when $c \neq 0$ and we scale so that $c = 1$ circles centered outside the unit disc:

$$(u - a)^2 + (v - b)^2 = a^2 + b^2 - 1,\ a^2 + b^2 > 1.$$

The upper half plane model comes from a conformal transformation of the the upper half plane to the unit disc (see section 4.5 exercise 6). This map is given by

$$F(x,y) = \frac{1}{x^2+(y+1)^2}\left(2x, 1-x^2-y^2\right).$$

The geodesics will again be lines and circles but F does not necessarily take lines to lines. The lines are all vertical:

$$x = 0, \text{ when } c = 0,\, b = 0,$$

or

$$x = 1/a, \text{ when } c = 1,\, b = -1,$$

and the circles have centers along the x-axis

$$\left(x - \frac{a}{b}\right)^2 + y^2 = 1 + \frac{a^2}{b^2}, \text{ when } c = 0,$$

or

$$\left(x - \frac{a}{b+1}\right)^2 + y^2 = \frac{a^2+b^2-1}{(b+1)^2}, \text{ when } c = 1.$$

It is interesting to note that for the sphere only the unit sphere model actually covers the entire sphere. In contrast, all of the models for hyperbolic space are equivalent in the sense that they are models for all of hyperbolic space, not just part of it.

Definition 7.1.6. An abstract surface is said to be *geodesically complete* if all geodesics exist for all time $t \in \mathbb{R}$. It is said to be *geodesically complete at a point*, if all geodesics through that point are defined for all time.

Example 7.1.7. The unit sphere, all of the above models for hyperbolic space, and all planes are geodesically complete.

As we have seen, it is often simpler to find the unparametrized form of the geodesics, i.e., in a given parametrization they are easier to find as an equation or as functions $u(v)$ or $v(u)$. There is in fact a tricky characterization of geodesics that does not refer to the arc-length parameter. The idea is that a regular curve can be reparametrized to be a geodesic if and only if its tangential acceleration $\ddot{\mathbf{q}}^{\mathrm{I}}$ is tangent to the curve.

Lemma 7.1.8. *A regular curve $\mathbf{q}(t) = \mathbf{q}(u(t), v(t))$ can be reparametrized as a geodesic if and only if*

$$\frac{dv}{dt}\left(\frac{d^2u}{dt^2} + \Gamma^u(\dot{\mathbf{q}}, \dot{\mathbf{q}})\right) = \frac{du}{dt}\left(\frac{d^2v}{dt^2} + \Gamma^v(\dot{\mathbf{q}}, \dot{\mathbf{q}})\right).$$

Proof. First observe that this formula holds iff $\lambda(t)\dot{\mathbf{q}}(t) = \ddot{\mathbf{q}}^{\mathrm{I}}(t)$ for some function λ.

If we reparametrize the curve, then the velocity satisfies: $\dot{\mathbf{q}}(t)=\frac{ds}{dt}\dot{\mathbf{q}}(s)$. For the acceleration we calculate in coordinates:

$$\begin{aligned}\frac{d^2u}{dt^2}+\Gamma^u(\dot{\mathbf{q}},\dot{\mathbf{q}}) &= \frac{d^2s}{dt^2}\frac{du}{ds}+\left(\frac{ds}{dt}\right)^2\frac{d^2u}{ds^2}+\Gamma^u\left(\frac{ds}{dt}\frac{d\mathbf{q}}{ds},\frac{ds}{dt}\frac{d\mathbf{q}}{ds}\right)\\ &= \frac{d^2s}{dt^2}\frac{du}{ds}+\left(\frac{ds}{dt}\right)^2\frac{d^2u}{ds^2}+\left(\frac{ds}{dt}\right)^2\Gamma^u\left(\frac{d\mathbf{q}}{ds},\frac{d\mathbf{q}}{ds}\right)\\ &= \frac{d^2s}{dt^2}\frac{du}{ds}+\left(\frac{ds}{dt}\right)^2\left(\frac{d^2u}{ds^2}+\Gamma^u\left(\frac{d\mathbf{q}}{ds},\frac{d\mathbf{q}}{ds}\right)\right).\end{aligned}$$

Similarly

$$\frac{d^2v}{dt^2}+\Gamma^v(\dot{\mathbf{q}},\dot{\mathbf{q}})=\frac{d^2s}{dt^2}\frac{dv}{ds}+\left(\frac{ds}{dt}\right)^2\left(\frac{d^2v}{ds^2}+\Gamma^v\left(\frac{d\mathbf{q}}{ds},\frac{d\mathbf{q}}{ds}\right)\right).$$

It follows that

$$\ddot{\mathbf{q}}^{\mathrm{I}}(t)=\frac{d^2s}{dt^2}\dot{\mathbf{q}}(s)+\left(\frac{ds}{dt}\right)^2\ddot{\mathbf{q}}^{\mathrm{I}}(s).$$

This shows first of all that, if $\mathbf{q}(s)$ is a geodesic, then $\ddot{\mathbf{q}}^{\mathrm{I}}(t)=\frac{d^2s}{dt^2}\dot{\mathbf{q}}(s)$ as claimed. Conversely assume that $\lambda(t)\dot{\mathbf{q}}(t)=\ddot{\mathbf{q}}^{\mathrm{I}}(t)$. Then

$$\lambda(s)\frac{ds}{dt}\dot{\mathbf{q}}(s)=\frac{d^2s}{dt^2}\dot{\mathbf{q}}(s)+\left(\frac{ds}{dt}\right)^2\ddot{\mathbf{q}}^{\mathrm{I}}(s).$$

So $\ddot{\mathbf{q}}^{\mathrm{I}}(s)=\mu(s)\dot{\mathbf{q}}(s)$ for some function μ. If we assume that s is the arclength parameter, then we also know that

$$\begin{aligned}0 &= \mathrm{I}\left(\ddot{\mathbf{q}}^{\mathrm{I}}(s),\dot{\mathbf{q}}(s)\right)\\ &= \mu(s).\end{aligned}$$

This shows that $\ddot{\mathbf{q}}^{\mathrm{I}}(s)=0$. □

Exercises.

(1) Let $\mathbf{q}(t)$ be a unit speed curve on a surface with normal $\mathbf{N}$. Show that it is a geodesic if and only if
$$\det[\dot{\mathbf{q}},\ddot{\mathbf{q}},\mathbf{N}]=0.$$
(2) Let $\mathbf{q}(t)$ be a unit speed curve on a surface. Show that
$$|\kappa_g|=\left|\ddot{\mathbf{q}}^{\mathrm{I}}\right|.$$
(3) Consider a unit speed curve $\mathbf{q}(t)$ on a surface of revolution
$$\mathbf{q}(u,\mu)=(r(u)\cos\mu,r(u)\sin\mu,z(u)),$$
where the profile curve $(r(u),z(u))$ is unit speed. Let $\theta(t)$ denote the angle with the meridians.
   (a) (Clairaut) Show that $r\sin\theta$ is constant along $\mathbf{q}(t)$ if it is a geodesic.
   (b) We say that $\mathbf{q}(t)$ is a loxodrome if θ is constant. Show that if all geodesics are loxodromes then the surface is a cylinder.

(4) Let $\mathbf{q}(t)$ be a unit speed geodesic on a surface in space. Show that
$$\begin{aligned}0 &= \kappa_g,\\ \kappa &= \kappa_n,\\ \tau &= \tau_g,\end{aligned}$$

where κ and τ are the curvature and torsion of $\mathbf{q}(t)$ as a space curve.

(5) Show that in the conformal model of the unit sphere the geodesics that pass through $(u, v) = (1, 0)$ all have center on the v-axis. Show that all initial value problems can be solved.

(6) Show that if a unit speed curve on a surface also lies in a plane that is perpendicular to the surface, then it is a geodesic.

(7) Show that geodesics satisfy a second order equation of the type

$$\frac{d^2v}{du^2} = A\left(\frac{dv}{du}\right)^3 + B\left(\frac{dv}{du}\right)^2 + C\frac{dv}{du} + D$$

and calculate the functions A, B, C, D in terms of the appropriate Christoffel symbols and metric coefficients.

(8) (Beltrami) Assume that $\mathbf{q}(u, v)$ is a parametrized surface with the property that all geodesics are lines in the domain U, i.e.,

$$au + bv + c = 0,\ (a, b) \neq (0, 0).$$

(a) Show that

$$\begin{aligned} \Gamma_{uu}^{v} &= \Gamma_{vv}^{u} = 0, \\ \Gamma_{uu}^{u} &= 2\Gamma_{uv}^{v}, \\ \Gamma_{vv}^{v} &= 2\Gamma_{uv}^{u}. \end{aligned}$$

Hint: Use lemma 7.1.8 and parametrize the curve by u or v.

(b) Use the Gauss equations

$$\begin{aligned} g_{vv}K &= R_{uvv}^{u}, \\ g_{uu}K &= R_{vuu}^{v}, \\ g_{vu}K &= -R_{uvv}^{v}, \\ g_{vu}K &= -R_{vuu}^{u}, \end{aligned}$$

together with the definitions of R_{ijk}^{l} to show that

$$0 = \begin{bmatrix} \frac{\partial K}{\partial v} & -\frac{\partial K}{\partial u} \end{bmatrix} \begin{bmatrix} g_{uu} & g_{uv} \\ g_{vu} & g_{vv} \end{bmatrix}.$$

(c) Conclude that the Gauss curvature is constant.

7.2. Mixed Partials

We need to generalize the intrinsic acceleration to also include mixed partial derivatives. The formulas obtained in section 6.3 will guide us.

Instead of just having a curve $\mathbf{q}(t) = \mathbf{q}(u(t), v(t))$ within a parametrization we assume that we have a family of curves $\mathbf{q}(s, t) = \mathbf{q}(u(s, t), v(s, t))$ such that for each s there is a curve parametrized by t. We shall generally assume that $(s, t) \in (-\epsilon, \epsilon) \times [a, b]$. In this case such a family of curves is called a *variation* of the base curve $\mathbf{q}(t) = \mathbf{q}(0, t)$. Note that $\mathbf{q}(s, t)$ does not have to be a valid parametrization of the surface.

To ease the notation we will use the conventions $\mathbf{q}^w(s, t) = w(s, t)$ so that we can write $\partial_s \mathbf{q}^w = \frac{\partial w}{\partial s}$, $\partial_t \partial_s \mathbf{q}^w = \frac{\partial^2 w}{\partial t \partial s}$, etc, and also use $\partial_t \partial_s \mathbf{q}^i$ with i in place of w.

We also define

$$\Gamma^w(X,Y) = \sum_{i,j=u,v} \Gamma^w_{ij} X^i Y^j = \begin{bmatrix} X^u & X^v \end{bmatrix} \begin{bmatrix} \Gamma^w_{uu} & \Gamma^w_{uv} \\ \Gamma^w_{vu} & \Gamma^w_{vv} \end{bmatrix} \begin{bmatrix} Y^u \\ Y^v \end{bmatrix}.$$

Keeping t or s fixed we already have that

$$\begin{aligned} \left(\frac{\partial^2 \mathbf{q}}{\partial s^2}\right)^{\mathrm{I}}(s,t) &= \left(\partial_s^2 u + \Gamma^u(\partial_s \mathbf{q}, \partial_s \mathbf{q})\right)\partial_u \mathbf{q} + \left(\partial_s^2 v + \Gamma^v(\partial_s \mathbf{q}, \partial_s \mathbf{q})\right)\partial_v \mathbf{q} \\ &= \sum_{i=u,v} \left(\partial_s^2 \mathbf{q}^i + \Gamma^i(\partial_s \mathbf{q}, \partial_s \mathbf{q})\right)\partial_i \mathbf{q}. \end{aligned}$$

and

$$\left(\frac{\partial^2 \mathbf{q}}{\partial t^2}\right)^{\mathrm{I}}(s,t) = \sum_{i=u,v} \left(\partial_t^2 \mathbf{q}^i + \Gamma^i(\partial_t \mathbf{q}, \partial_t \mathbf{q})\right)\partial_i \mathbf{q}.$$

Moreover, when the surface lies in $\mathbb{R}^3$, then these intrinsic second partials are in fact the tangential components of the second partials in $\mathbb{R}^3$.

The intrinsic mixed partial is similarly defined as

$$\left(\frac{\partial^2 \mathbf{q}}{\partial s \partial t}\right)^{\mathrm{I}}(s,t) = \sum_{i=u,v} \left(\partial_s \partial_t \mathbf{q}^i + \Gamma^i(\partial_s \mathbf{q}, \partial_t \mathbf{q})\right)\partial_i \mathbf{q}.$$

This mixed partial also commutes commutes since

$$\frac{\partial^2 w}{\partial s \partial t} = \frac{\partial^2 w}{\partial t \partial s}$$

and

$$\Gamma^w\left(\frac{\partial \mathbf{q}}{\partial s}, \frac{\partial \mathbf{q}}{\partial t}\right) = \Gamma^w\left(\frac{\partial \mathbf{q}}{\partial t}, \frac{\partial \mathbf{q}}{\partial s}\right).$$

We can also show that all possible product formulas for taking derivatives hold:

$$\begin{aligned} \partial_s \mathrm{I}(\partial_s \mathbf{q}, \partial_t \mathbf{q}) &= \mathrm{I}\left(\left(\partial_s^2 \mathbf{q}\right)^{\mathrm{I}}, \partial_t \mathbf{q}\right) + \mathrm{I}\left(\partial_s \mathbf{q}, \left(\partial_s \partial_t \mathbf{q}\right)^{\mathrm{I}}\right), \\ \partial_s \mathrm{I}(\partial_t \mathbf{q}, \partial_t \mathbf{q}) &= 2\mathrm{I}\left(\partial_t \mathbf{q}, \left(\partial_s \partial_t \mathbf{q}\right)^{\mathrm{I}}\right), \\ \partial_s \mathrm{I}(\partial_s \mathbf{q}, \partial_s \mathbf{q}) &= 2\mathrm{I}\left(\left(\partial_s^2 \mathbf{q}\right)^{\mathrm{I}}, \partial_s \mathbf{q}\right), \\ \partial_t \mathrm{I}(\partial_s \mathbf{q}, \partial_t \mathbf{q}) &= \mathrm{I}\left(\left(\partial_t \partial_s \mathbf{q}\right)^{\mathrm{I}}, \partial_t \mathbf{q}\right) + \mathrm{I}\left(\partial_s \mathbf{q}, \left(\partial_t^2 \mathbf{q}\right)^{\mathrm{I}}\right), \\ \partial_t \mathrm{I}(\partial_s \mathbf{q}, \partial_s \mathbf{q}) &= 2\mathrm{I}\left(\partial_s \mathbf{q}, \left(\partial_t \partial_s \mathbf{q}\right)^{\mathrm{I}}\right), \\ \partial_t \mathrm{I}(\partial_t \mathbf{q}, \partial_t \mathbf{q}) &= 2\mathrm{I}\left(\partial_t \mathbf{q}, \left(\partial_t^2 \mathbf{q}\right)^{\mathrm{I}}\right). \end{aligned}$$

The proofs are all similar so we concentrate on the first. The essential idea is that we have the product formula

$$\partial_s g_{ij} = \Gamma_{sij} + \Gamma_{sji}$$

directly from the abstract definition of the Christoffel symbols as in section 6.2.

$$\begin{aligned}
&\partial_s \mathrm{I}\left(\partial_s \mathbf{q}, \partial_t \mathbf{q}\right) \\
&= \partial_s\left(g_{ij} \partial_s \mathbf{q}^i \partial_t \mathbf{q}^j\right) \\
&= \partial_s\left(g_{ij}\right) \partial_s \mathbf{q}^i \partial_t \mathbf{q}^j+g_{ij} \partial_s^2 \mathbf{q}^i \partial_t \mathbf{q}^j+g_{ij} \partial_s \mathbf{q}^i \partial_s \partial_t \mathbf{q}^j \\
&= \left(\partial_k g_{ij}\right) \partial_s \mathbf{q}^k \partial_s \mathbf{q}^i \partial_t \mathbf{q}^j+g_{ij} \partial_s^2 \mathbf{q}^i \partial_t \mathbf{q}^j+g_{ij} \partial_s \mathbf{q}^i \partial_s \partial_t \mathbf{q}^j \\
&= \left(\Gamma_{kij}+\Gamma_{kji}\right) \partial_s \mathbf{q}^k \partial_s \mathbf{q}^i \partial_t \mathbf{q}^j+g_{ij} \partial_s^2 \mathbf{q}^i \partial_t \mathbf{q}^j+g_{ij} \partial_s \mathbf{q}^i \partial_s \partial_t \mathbf{q}^j \\
&= \Gamma_{kij} \partial_s \mathbf{q}^k \partial_s \mathbf{q}^i \partial_t \mathbf{q}^j+g_{ij} \partial_s^2 \mathbf{q}^i \partial_t \mathbf{q}^j \\
&\quad +\partial_s \mathbf{q}^i \Gamma_{kji} \partial_s \mathbf{q}^k \partial_t \mathbf{q}^j+g_{ij} \partial_s \mathbf{q}^i \partial_s \partial_t \mathbf{q}^j \\
&= g_{lj} \Gamma_{ki}^l \partial_s \mathbf{q}^k \partial_s \mathbf{q}^i \partial_t \mathbf{q}^j+g_{ij} \partial_s^2 \mathbf{q}^i \partial_t \mathbf{q}^j \\
&\quad +g_{il} \partial_s \mathbf{q}^i \Gamma_{kj}^l \partial_s \mathbf{q}^k \partial_t \mathbf{q}^j+g_{ij} \partial_s \mathbf{q}^i \partial_s \partial_t \mathbf{q}^j \\
&= g_{ij} \Gamma_{kl}^i \partial_s \mathbf{q}^k \partial_s \mathbf{q}^l \partial_t \mathbf{q}^j+g_{ij} \partial_s^2 \mathbf{q}^i \partial_t \mathbf{q}^j \\
&\quad +g_{ij} \partial_s \mathbf{q}^i \Gamma_{kl}^j \partial_s \mathbf{q}^k \partial_t \mathbf{q}^l+g_{ij} \partial_s \mathbf{q}^i \partial_s \partial_t \mathbf{q}^j \\
&= g_{ij}\left(\Gamma^i\left(\partial_s \mathbf{q}, \partial_s \mathbf{q}\right)+\partial_s^2 \mathbf{q}^i\right) \partial_t \mathbf{q}^j \\
&\quad +g_{ij} \partial_s \mathbf{q}^i\left(\Gamma_{kl}^j \partial_s \mathbf{q}^k \partial_t \mathbf{q}^l+\partial_s \partial_t \mathbf{q}^j\right) \\
&= \mathrm{I}\left(\left(\partial_s^2 \mathbf{q}\right)^{\mathrm{I}}, \partial_t \mathbf{q}\right)+\mathrm{I}\left(\partial_s \mathbf{q},\left(\partial_s \partial_t \mathbf{q}\right)^{\mathrm{I}}\right).
\end{aligned}$$

Finally we should also justify why these second partial derivatives do not depend on the initial (u, v)-parametrization. This could be done via a notationally nasty change of parameters or by a more general formula that doesn't depend a parametrization. This general formula, however, also has a defect in that it involves a new variable r so that $w = w(r, s, t)$:

$$2\mathrm{I}\left(\left(\partial_s \partial_t \mathbf{q}\right)^{\mathrm{I}}, \partial_r \mathbf{q}\right)=\partial_s \mathrm{I}\left(\partial_t \mathbf{q}, \partial_r \mathbf{q}\right)+\partial_t \mathrm{I}\left(\partial_s \mathbf{q}, \partial_r \mathbf{q}\right)-\partial_r \mathrm{I}\left(\partial_s \mathbf{q}, \partial_t \mathbf{q}\right).$$

Here the right hand side can be calculated independently of a (u, v)-parametrization. Since we can think of the r-variable as being anything we please, this implicitly calculates $\left(\partial_s \partial_t \mathbf{q}\right)^{\mathrm{I}}$. The proof of this identity comes from from using the product rule on each on the terms on the right hand side and using that the intrinsic mixed partials commute:

$$\begin{aligned}
&\partial_s \mathrm{I}\left(\partial_t \mathbf{q}, \partial_r \mathbf{q}\right)+\partial_t \mathrm{I}\left(\partial_s \mathbf{q}, \partial_r \mathbf{q}\right)-\partial_r \mathrm{I}\left(\partial_s \mathbf{q}, \partial_t \mathbf{q}\right) \\
&\quad= \mathrm{I}\left(\left(\partial_s \partial_t \mathbf{q}\right)^{\mathrm{I}}, \partial_r \mathbf{q}\right)+\mathrm{I}\left(\partial_t \mathbf{q},\left(\partial_s \partial_r \mathbf{q}\right)^{\mathrm{I}}\right) \\
&\quad\quad +\mathrm{I}\left(\left(\partial_t \partial_s \mathbf{q}\right)^{\mathrm{I}}, \partial_r \mathbf{q}\right)+\mathrm{I}\left(\partial_s \mathbf{q},\left(\partial_t \partial_r \mathbf{q}\right)^{\mathrm{I}}\right) \\
&\quad\quad -\left(\mathrm{I}\left(\left(\partial_r \partial_s \mathbf{q}\right)^{\mathrm{I}}, \partial_t \mathbf{q}\right)+\mathrm{I}\left(\partial_s \mathbf{q},\left(\partial_r \partial_t \mathbf{q}\right)^{\mathrm{I}}\right)\right) \\
&\quad= 2\mathrm{I}\left(\left(\partial_s \partial_t \mathbf{q}\right)^{\mathrm{I}}, \partial_r \mathbf{q}\right).
\end{aligned}$$

Proposition 7.2.1. *Let $\mathbf{q}(t)$ be a curve on a surface M. $\mathbf{q}$ has constant speed if and only if its intrinsic acceleration is perpendicular to the speed.*

Proof. The proof is now a simple calculation using the product rule for intrinsic second derivatives:

$$\frac{d}{dt} \mathrm{I}(\dot{\mathbf{q}}, \dot{\mathbf{q}})=2\mathrm{I}\left(\ddot{\mathbf{q}}^{\mathrm{I}}, \dot{\mathbf{q}}\right).$$

□

7.3. Shortest Curves

The goal is to show that the shortest curves are geodesics.

In the last section we considered variations $\mathbf{q}(s,t) = \mathbf{q}(u(s,t), v(s,t))$ where $(s,t) \in (-\epsilon, \epsilon) \times [a,b]$. The variational field of $\mathbf{q}(t) = \mathbf{q}(0,t)$ is given by the tangent vectors $V(t) = \frac{\partial \mathbf{q}}{\partial s}(0,t)$ along the curve. The first proposition shows that any such field $V(t) \in T_{\mathbf{q}(t)}M$ comes from a variation.

Proposition 7.3.1. *For any curve $\mathbf{q}(t)$, $t \in [a,b]$ and tangent field $V(t) \in T_{\mathbf{q}(t)}M$, there is a variation whose variational field is $V(t)$.*

Proof. For each $V(t)$ let $s \mapsto \mathbf{q}(s,t)$ be the unique geodesic with $\mathbf{q}(0,t) = \mathbf{q}(t)$ and $\frac{\partial \mathbf{q}}{\partial s}(0,t) = V(t)$. The fact that $[a,b]$ is compact shows that we can find $\epsilon > 0$ so that $\mathbf{q}(s,t)$ is defined on $(-\epsilon,\epsilon) \times [a,b]$.

The fact that the geodesics depend smoothly on the initial values shows that the variation is a smooth as $\mathbf{q}(t)$ and $V(t)$. In particular, if $\mathbf{q}(t)$ is only piecewise smooth, then the variation will also consist of piecewise smooth curves that break at exactly the same points. □

Definition 7.3.2. The *length of a curve* is defined as

$$L(\mathbf{q}) = \int_a^b |\dot{\mathbf{q}}| \, dt$$

and the (kinetic) *energy* as

$$E(\mathbf{q}) = \frac{1}{2} \int_a^b |\dot{\mathbf{q}}|^2 \, dt.$$

We know that the length of a curve does not change if we parametrize it. This is very far from true for the energy. You might even have noticed this yourself in terms of gas consumption when driving. Stop and go city driving consumes far more gas, than the more steady driving on an empty stretch of road on the country side. On the other hand this feature of the energy has the advantage that minima or stationary points for the energy functional come with a fixed parametrization.

Lemma 7.3.3 (*First Variation Formula*). *Consider a smooth variation $\mathbf{q}(s,t)$, $(s,t) \in (-\epsilon,\epsilon) \times [0,1]$, with base curve $\mathbf{q}(t) = \mathbf{q}(0,t)$, then*

$$\frac{d}{ds} \frac{1}{2} \int_0^1 I(\dot{\mathbf{q}}, \dot{\mathbf{q}}) \, dt = I(\partial_s \mathbf{q}, \partial_t \mathbf{q})|_0^1 - \int_0^1 I\left(\partial_s \mathbf{q}, \ddot{\mathbf{q}}^I\right) dt.$$

If $0 = a_0 < a_1 < \cdots < a_n = 1$ and the variation is smooth when restricted to $(-\epsilon,\epsilon) \times [a_{i-1}, a_i]$, then

$$\frac{d}{ds} \frac{1}{2} \int_0^1 I(\dot{\mathbf{q}}, \dot{\mathbf{q}}) \, dt = \sum_{i=1}^n I(\partial_s \mathbf{q}, \partial_t \mathbf{q})|_{a_{i-1}}^{a_i} - \int_0^1 I\left(\partial_s \mathbf{q}, \ddot{\mathbf{q}}^I\right) dt.$$

Proof. The calculation is straightforward in the smooth case:

$$\begin{aligned}
\frac{d}{ds}\frac{1}{2}\int_0^1 \mathrm{I}(\dot{\mathbf{q}},\dot{\mathbf{q}})\,dt &= \int_0^1 \mathrm{I}\left((\partial_s\partial_t\mathbf{q})^{\mathrm{I}},\partial_t\mathbf{q}\right)dt \\
&= \int_0^1 \left(\partial_t\mathrm{I}(\partial_s\mathbf{q},\partial_t\mathbf{q}) - \mathrm{I}\left(\partial_s\mathbf{q},\left(\partial_t^2\mathbf{q}\right)^{\mathrm{I}}\right)\right)dt \\
&= \mathrm{I}(\partial_s\mathbf{q},\partial_t\mathbf{q})|_0^1 - \int_0^1 \mathrm{I}\left(\partial_s\mathbf{q},\left(\partial_t^2\mathbf{q}\right)^{\mathrm{I}}\right)dt \\
&= \mathrm{I}(\partial_s\mathbf{q},\partial_t\mathbf{q})|_0^1 - \int_0^1 \mathrm{I}\left(\partial_s\mathbf{q},\ddot{\mathbf{q}}^{\mathrm{I}}\right)dt.
\end{aligned}$$

When the variation is only piecewise smooth, then we can break it up into smooth parts and add the contributions. □

We define $\Omega_{p,q}$ as the space of piecewise smooth curves between points $p,q\in M$ parametrized on $[0,1]$.

Theorem 7.3.4. *If a piecewise curve on a surface is stationary for the energy functional on $\Omega_{p,q}$, then it is a geodesic.*

Proof. We consider a piecewise smooth variation $\mathbf{q}(s,t)$ where the base curve $\mathbf{q}(t)=\mathbf{q}(0,t)$ corresponds to $s=0$. For simplicity assume that there is only one break point at a. Computing the energy of the curves $t\to\mathbf{q}(s,t)$ gives a function of s. The derivative with respect to s can be calculated as

$$\frac{d}{ds}\frac{1}{2}\int_0^1 \mathrm{I}(\dot{\mathbf{q}},\dot{\mathbf{q}})\,dt = \mathrm{I}(\partial_s\mathbf{q},\partial_t\mathbf{q})|_0^a + \mathrm{I}(\partial_s\mathbf{q},\partial_t\mathbf{q})|_a^1 - \int_0^1 \mathrm{I}\left(\partial_s\mathbf{q},\ddot{\mathbf{q}}^{\mathrm{I}}\right)dt.$$

When all the curves lie in $\Omega_{p,q}$ they have the same end points at $t=0,1$, i.e., $\mathbf{q}(s,0)=p$ and $\mathbf{q}(s,1)=q$ for all s. Such a variation is also called a *proper variation*. Thus, $\frac{\partial\mathbf{q}}{\partial s}(0,t)=0$ at $t=0,1$ and the formula simplifies to

$$\frac{d}{ds}\frac{1}{2}\int_0^1 \mathrm{I}(\dot{\mathbf{q}},\dot{\mathbf{q}})\,dt = \mathrm{I}\left(\partial_s\mathbf{q}(a),\frac{\partial\mathbf{q}}{\partial t^-}(a)-\frac{\partial_{t^+}\mathbf{q}}{\partial t^+}(a)\right) - \int_0^1 \mathrm{I}\left(\partial_s\mathbf{q},\ddot{\mathbf{q}}^{\mathrm{I}}\right)dt.$$

By assumption $s=0$ is a stationary point for $\frac{1}{2}\int_0^1 \mathrm{I}(\dot{\mathbf{q}},\dot{\mathbf{q}})\,dt$ so

$$0=\mathrm{I}\left(\partial_s\mathbf{q}(a),\frac{\partial\mathbf{q}}{\partial t^-}(a)-\frac{\partial\mathbf{q}}{\partial t^+}(a)\right) - \int_0^1 \mathrm{I}\left(\partial_s\mathbf{q},\ddot{\mathbf{q}}^{\mathrm{I}}\right)dt.$$

First select the variation so that $\partial_s\mathbf{q}(0,t)$ is proportional to the tangential acceleration $\ddot{\mathbf{q}}^{\mathrm{I}}$, i.e., $\partial_s\mathbf{q}(0,t)=\mu(t)\,\ddot{\mathbf{q}}^{\mathrm{I}}$, where $\mu(a)=0$. Then we obtain

$$0=-\int_0^1 \mu(t)\left|\ddot{\mathbf{q}}^{\mathrm{I}}\right|^2 dt.$$

Since μ can be chosen to be positive on $(0,a)\cup(a,1)$ this shows that $\ddot{\mathbf{q}}^{\mathrm{I}}=0$ on $(0,a)\cup(a,1)$. This shows that each of the two parts of $\mathbf{q}(t)$ on $[0,a]$ and $[a,1]$ are geodesics.

Next select a variation where

$$\partial_s\mathbf{q}(0,a)=\frac{\partial\mathbf{q}}{\partial t^-}(a)-\frac{\partial\mathbf{q}}{\partial t^+}(a).$$

In this case

$$0=\mathrm{I}\left(\partial_s\mathbf{q}(a),\frac{\partial\mathbf{q}}{\partial t^-}(a)-\frac{\partial\mathbf{q}}{\partial t^+}(a)\right)=\mathrm{I}\left(\frac{\partial\mathbf{q}}{\partial t^-}(a)-\frac{\partial\mathbf{q}}{\partial t^+}(a),\frac{\partial\mathbf{q}}{\partial t^-}(a)-\frac{\partial\mathbf{q}}{\partial t^+}(a)\right)$$

so it follows that

$$\frac{\partial \mathbf{q}}{\partial t^-}(a) = \frac{\partial \mathbf{q}}{\partial t^+}(a).$$

Uniqueness of geodesics, then shows that the two parts of $\mathbf{q}(t)$ fit together to form a smooth geodesic on $[0,1]$.

Finally any curve of minimal energy is necessarily stationary since the derivative always vanishes at a minimum for a function. □

Now that we have identified the minima for the energy functional we show that they are also minima for the length functional.

LEMMA 7.3.5. *A minimum for the energy functional is also a minimum for the length functional.*

PROOF. We start by observing that the Cauchy-Schwarz inequality for the inner product of functions defined by

$$(f,g) = \int_a^b f(t)\,g(t)\,dt$$

implies that:

$$\int_a^b |\dot{\mathbf{q}}|\,dt \leq \sqrt{\int_a^b 1^2 dt}\sqrt{\int_a^b |\dot{\mathbf{q}}|^2\,dt} = \sqrt{b-a}\sqrt{\int_a^b |\dot{\mathbf{q}}|^2\,dt},$$

where equality occurs if $|\dot{\mathbf{q}}|$ is constant multiple of 1, i.e., $\mathbf{q}$ has constant speed. When the right hand side is minimized we just saw that $\mathbf{q}$ has zero acceleration and consequently constant speed. Let $\mathbf{q}_{\min}$ be a minimum for the energy in $\Omega_{p,q}$ and $\mathbf{q}$ any other curve in $\Omega_{p,q}$. We further assume that $\mathbf{q}$ has constant speed as reparametrizing the curve won't change its length. We now have

$$\begin{aligned}
\int_0^1 |\dot{\mathbf{q}}_{\min}|\,dt &\leq \sqrt{\int_0^1 |\dot{\mathbf{q}}_{\min}|^2\,dt} \\
&\leq \sqrt{\int_0^1 |\dot{\mathbf{q}}|^2\,dt} \\
&= \int_0^1 |\dot{\mathbf{q}}|\,dt
\end{aligned}$$

which shows the claim. □

COROLLARY 7.3.6. *If a piecewise smooth curve has constant speed and is a minimum for the length functional, then it is a minimum for the energy and a geodesic.*

PROOF. If $\mathbf{q}_{\min}$ is a constant speed minimum for the length functional and $\mathbf{q} \in \Omega_{p,q}$, then

$$\begin{aligned}
\int_0^1 |\dot{\mathbf{q}}_{\min}|^2\,dt &= \left(\int_0^1 |\dot{\mathbf{q}}_{\min}|\,dt\right)^2 \\
&\leq \left(\int_0^1 |\dot{\mathbf{q}}|\,dt\right)^2 \\
&\leq \int_0^1 |\dot{\mathbf{q}}|^2\,dt.
\end{aligned}$$

This shows that $\mathbf{q}_{\min}$ also minimizes the energy functional and by theorem 7.3.4 that it must be a geodesic. □

REMARK 7.3.7. Note that minima for the length functional are not forced to be geodesics unless they are assumed to have constant speed!

7.4. Short Geodesics

We start by introducing geodesic coordinates along a curve. We then proceed to do the same construction around a point. This construction is similar but complicated by the fact that our base curve is a fixed point. In Euclidean space this corresponds to the singularity at the origin when switching from Cartesian to polar coordinates.

PROPOSITION 7.4.1. *Every surface admits geodesic coordinates around every point.*

PROOF. Start by choosing a unit speed curve $\mathbf{q}(v)$, $v \in [a,b]$ such that the specified point $q = \mathbf{q}(v_0)$ for some $v_0 \in (a,b)$. Next select a consistent choice of unit normal vector $\mathbf{S}(v)$ to this curve inside the surface as a variational field. Then let $u \mapsto \mathbf{q}(u,v)$ be the unique unit speed geodesic with $\mathbf{q}(0,v) = \mathbf{q}(v)$ and $\partial_u \mathbf{q}(0,v) = \mathbf{S}(v)$ to obtain a variation on $(-\epsilon,\epsilon) \times [a,b]$.

Since $u \mapsto \mathbf{q}(u,v)$ is unit speed we have $\mathrm{I}(\partial_u \mathbf{q}, \partial_u \mathbf{q}) = 1$. Next consider the inner product $\mathrm{I}(\partial_u \mathbf{q}, \partial_v \mathbf{q})$. Since $\partial_u \mathbf{q}(0,v) = \mathbf{S}(v)$ is perpendicular to $\partial_v \mathbf{q}(0,v) = \partial_v \mathbf{q}(v)$ this inner product vanishes for all parameters $(0,v)$. If we differentiate the inner product with respect to u and use the product rule twice we obtain

$$\begin{aligned}
\partial_u \mathrm{I}(\partial_u \mathbf{q}, \partial_v \mathbf{q}) &= \mathrm{I}\left(\left(\partial_u^2 \mathbf{q}\right)^{\mathrm{I}}, \partial_v \mathbf{q}\right) + \mathrm{I}\left(\partial_u \mathbf{q}, \left(\partial_u \partial_v \mathbf{q}\right)^{\mathrm{I}}\right) \\
&= \mathrm{I}\left(\partial_u \mathbf{q}, \left(\partial_u \partial_v \mathbf{q}\right)^{\mathrm{I}}\right) \\
&= \mathrm{I}\left(\partial_u \mathbf{q}, \left(\partial_v \partial_u \mathbf{q}\right)^{\mathrm{I}}\right) \\
&= \frac{1}{2}\left(\mathrm{I}\left(\partial_u \mathbf{q}, \left(\partial_v \partial_u \mathbf{q}\right)^{\mathrm{I}}\right) + \mathrm{I}\left(\left(\partial_v \partial_u \mathbf{q}\right)^{\mathrm{I}}, \partial_u \mathbf{q}\right)\right) \\
&= \frac{1}{2}\left(\partial_v \mathrm{I}(\partial_u \mathbf{q}, \partial_u \mathbf{q})\right) \\
&= \frac{1}{2}(\partial_v 1) \\
&= 0.
\end{aligned}$$

This shows that $\mathrm{I}(\partial_u \mathbf{q}, \partial_v \mathbf{q})$ is constant along u-curves and vanishes at $u = 0$. Thus it vanishes everywhere.

Finally define $g_{vv} = \mathrm{I}(\partial_v \mathbf{q}, \partial_v \mathbf{q})$. It now just remains to note that $g_{vv}(0,v) = 1$ and $g_{vv}(u,v)$ is continuous. Thus we can, after possibly decreasing ϵ, assume that $g_{vv} > 0$ on all of the region $(-\epsilon,\epsilon) \times [a,b]$. This shows that the velocity fields $\partial_u \mathbf{q}$ and $\partial_v \mathbf{q}$ never vanish and are always orthogonal. Thus they give the desired parametrization. We can then further restrict the domain around $(0, v_0)$ if we wish to obtain a coordinate system where the parametrization is a local diffeomorphism. □

We now fix a point $p \in M$. For a tangent vector $X \in T_p M$, let $\mathbf{q}_X$ be the unique geodesic with $\mathbf{q}(0) = p$ and $\dot{\mathbf{q}}(0) = X$, and $[0, b_X)$ the non-negative part of the maximal interval on which $\mathbf{q}$ is defined. Notice that uniqueness of geodesics implies the *homogeneity property*: $\mathbf{q}_{\alpha X}(t) = \mathbf{q}_X(\alpha t)$ for all $\alpha > 0$ and $t < b_{\alpha X}$. In particular, $b_{\alpha X} = \alpha^{-1} b_X$. Let $O_p \subset T_pM$ be the set of vectors X such that $1 < b_X$. In other words $\mathbf{q}_X(t)$ is defined on $[0,1]$.

DEFINITION 7.4.2. The *exponential map* at p, $\exp_p : O_p \to M$, is defined by

$$\exp_p(X) = \mathbf{q}_X(1).$$

The homogeneity property $\mathbf{q}_X(t) = \mathbf{q}_{tX}(1)$ shows that $\exp_p(tX) = \mathbf{q}_X(t)$. Therefore, it is natural to think of $\exp_p(X)$ in a polar coordinate representation, where from p one goes "distance" $|X|$ in the direction of $\frac{X}{|X|}$. This gives the point $\exp_p(X)$, since $\mathbf{q}_{\frac{X}{|X|}}(|X|) = \mathbf{q}_X(1)$.

It is an important property that $\exp_p$ is in fact a local diffeomorphism around $0 \in T_pM$.

PROPOSITION 7.4.3. *For each $p \in M$ there exists $\epsilon > 0$ so that $B(0,\epsilon) \subset O_p \subset T_pM$ and the differential $D\exp_p$ is nonsingular at the origin. Consequently, $\exp_p$ is a local diffeomorphism.*

PROOF. By theorem A.5.1 there exists $\epsilon > 0$ such that $\mathbf{q}_X(t)$ is defined on $[0,2\epsilon)$ for all unit vectors $X \in T_pM$. The homogeneity then shows that $B(0,\epsilon) \subset O_p$.That the differential is non-singular also follows from the homogeneity property of geodesics. For a fixed vector $X \in T_pM$ we just saw that

$$\exp_p(tX) = \mathbf{q}_X(t)$$

and thus

$$\begin{aligned}\left(D\exp_p\right)(X) &= \frac{d}{dt}|_{t=0}\exp_p(tX)\\ &= \dot{\mathbf{q}}_X(0)\\ &= X.\end{aligned}$$

This shows that the differential is the identity map and in particular non-singular. The second statement follows from the inverse function theorem. □

We can now introduce Gauss's version of geodesic polar coordinates.

LEMMA 7.4.4 (Gauss Lemma). *Around any point $p \in M$ it is possible to introduce polar geodesic coordinate parameters $\mathbf{q}(r,\theta)$ where the r-parameter curves are the unit speed geodesics emanating from p and*

$$[I] = \begin{bmatrix} 1 & 0 \\ 0 & g_{\theta\theta} \end{bmatrix}.$$

PROOF. Pick $\epsilon > 0$ such that $\exp_p : B(0,\varepsilon) \to B = \exp_p(B(0,\varepsilon))$ is a diffeomorphism. Then $r(q) = |\exp_p^{-1}(q)|$ is well-defined for all $q \in B$. Note that r is simply the Euclidean distance function from the origin on $B(0,\varepsilon) \subset T_pM$ in exponential coordinates. This function can be continuously extended to $\bar{B}$ by defining $r(\partial B) = \varepsilon$. Select an orthonormal basis E_1, E_2 for T_pM and introduce Cartesian coordinates (x,y) on T_pM. These parameters are then also used on B via the exponential map $\mathbf{q}(x,y) = \exp_p(xE_1 + yE_2)$. We define the polar coordinates by

$$x = r\cos\theta,\ y = r\sin\theta$$

and note that

$$\begin{aligned} r &= \sqrt{x^2+y^2}, \\ \partial_r \mathbf{q} &= \frac{x}{r}\partial_x \mathbf{q} + \frac{y}{r}\partial_y \mathbf{q}, \\ \partial_\theta \mathbf{q} &= -y\partial_x \mathbf{q} + x\partial_y \mathbf{q}. \end{aligned}$$

Observe that $\partial_r \mathbf{q}$ is not defined at p, while $\partial_\theta \mathbf{q}$ is defined on all of B even though the angle θ is not. We now need to check what the first fundamental form looks like in polar coordinates. First note that the r-parameter curves by definition have velocity $\partial_r \mathbf{q}$. On the other hand via the exponential map they correspond to unit speed radial lines rX, where $|X| = 1$. This means that they are of the form $\exp_p(rX) = \mathbf{q}_X(r)$ and are unit speed geodesics. This shows that $g_{rr} = \mathrm{I}(\partial_r \mathbf{q}, \partial_r \mathbf{q}) = 1$. To show that $g_{r\theta} = 0$ we first calculate its derivative

$$\begin{aligned} \partial_r \mathrm{I}(\partial_r \mathbf{q}, \partial_\theta \mathbf{q}) &= \mathrm{I}\left(\left(\partial_r^2 \mathbf{q}\right)^{\mathrm{I}}, \partial_\theta \mathbf{q}\right) + \mathrm{I}\left(\partial_r \mathbf{q}, \left(\partial_r\partial_\theta \mathbf{q}\right)^{\mathrm{I}}\right) \\ &= 0 + \mathrm{I}\left(\partial_r \mathbf{q}, \left(\partial_\theta\partial_r \mathbf{q}\right)^{\mathrm{I}}\right) \\ &= \frac{1}{2}\partial_\theta \mathrm{I}(\partial_r \mathbf{q}, \partial_r \mathbf{q}) \\ &= 0. \end{aligned}$$

Thus $\mathrm{I}(\partial_r \mathbf{q}, \partial_\theta \mathbf{q})$ is constant along geodesics emanating from p. To show that it vanishes it is tempting to simply evaluate at p since $\partial_\theta \mathbf{q}$ vanishes there. However, $\partial_r \mathbf{q}$ is undefined so we use a limit argument. First observe that

$$\begin{aligned} |\mathrm{I}(\partial_r \mathbf{q}, \partial_\theta \mathbf{q})| &\leq |\partial_r \mathbf{q}|\,|\partial_\theta \mathbf{q}| \\ &= |\partial_\theta \mathbf{q}| \\ &\leq |x|\,|\partial_y \mathbf{q}| + |y|\,|\partial_x \mathbf{q}| \\ &\leq r\left(|\partial_x \mathbf{q}| + |\partial_y \mathbf{q}|\right). \end{aligned}$$

Continuity of $D\exp_p$ shows that $\partial_x \mathbf{q}, \partial_y \mathbf{q}$ are bounded near p. Thus $\mathrm{I}(\partial_r \mathbf{q}, \partial_\theta \mathbf{q}) \to 0$ as $r \to 0$. This forces $\mathrm{I}(\partial_r \mathbf{q}, \partial_\theta \mathbf{q}) = 0$.

Finally we can just define $g_{\theta\theta} = \mathrm{I}(\partial_\theta \mathbf{q}, \partial_\theta \mathbf{q})$ and note that it is positive as $\partial_\theta \mathbf{q}$ only vanishes at p. □

THEOREM 7.4.5. *Let M be a surface, $p \in M$, and $\varepsilon > 0$ chosen such that*

$$\exp_p : B(0,\varepsilon) \to B \subset M$$

is a diffeomorphism onto its image $B \subset M$. It follows that the geodesic $\mathbf{q}_X(t) = \exp_p(tX)$, $t \in [0,1]$ is the one and only minimal geodesic in M from p to $q = \exp_p X$.

PROOF. The proof is analogous to the specific situation on the round sphere covered in example 1.2.9.

To see that $\mathbf{q}_X(t)$ is the one and only shortest curve in M, we must show that any other curve from p to q has length $> |X|$. Suppose we have a curve $\mathbf{q} : [0,b] \to M$ from p to q. If $a \in [0,b]$ is the largest value so that $\mathbf{q}(a) = p$, then $\mathbf{q}|_{[a,b]}$ is a shorter curve from p to q. Next let $b_0 \in (a,b)$ be the first value for which $\mathbf{q}(t_0) \notin B$ if such points exist, otherwise $b_0 = b$. The curve $\mathbf{q}|_{(a,b_0)}$ now lies entirely in $B - \{p\}$ and is shorter than the original curve. Its length is easily estimated from below

$$\begin{aligned} L\left(\mathbf{q}|_{(a,b_0)}\right) &= \int_a^{b_0} |\dot{\mathbf{q}}|\, dt \\ &= \int_a^{b_0} |\partial_r \mathbf{q}| \cdot |\dot{\mathbf{q}}|\, dt \\ &\geq \int_a^{b_0} \mathrm{I}\left(\partial_r \mathbf{q}, \dot{\mathbf{q}}\right) dt \\ &= \int_a^{b_0} \mathrm{I}\left(\partial_r \mathbf{q}, \frac{dr\left(\mathbf{q}\left(t\right)\right)}{dt}\partial_r \mathbf{q} + \frac{d\theta\left(\mathbf{q}\left(t\right)\right)}{dt}\partial_\theta \mathbf{q}\right) dt \\ &= \int_a^{b_0} \frac{dr\left(\mathbf{q}\left(t\right)\right)}{dt} dt \\ &= r\left(\mathbf{q}\left(b_0\right)\right) - r\left(\mathbf{q}\left(a\right)\right) \\ &= r\left(\mathbf{q}\left(b_0\right)\right), \end{aligned}$$

where we used that $r(p) = 0$. If $\mathbf{q}\left(b_0\right) \in \partial B$, then $\mathbf{q}$ is not a segment from p to q as it has length $\geq \varepsilon > |X|$. If $b = b_0$, then $L\left(\mathbf{q}|_{(a,b)}\right) \geq r\left(\mathbf{q}\left(b\right)\right) = |X|$ and equality can only hold if $\dot{\mathbf{q}}\left(t\right)$ is proportional to $\partial_r \mathbf{q}$ for all $t \in (a, b]$. This shows the short geodesic is a minimal geodesic and that any other curve of the same length must be a reparametrization of this short geodesic. □

7.5. Distance and Completeness

Definition 7.5.1. The distance between two points in a surface M is defined by attempting to minimize the length of curves between the points:

$$|pq| = \inf\left\{L\left(\mathbf{q}\right) \mid \mathbf{q} \in \Omega_{pq}\right\}.$$

This distance satisfies the usual properties of a distance:

(1) $|pq| > 0$ unless $p = q$,
(2) $|pq| = |qp|$,
(3) $|pq| \leq |px| + |xq|$.

2 and 3 are also immediate from the definition. It is also clear that $|pq| \geq 0$. Finally, if $|pq| = 0$, then $q \in B = \exp_p\left(B\left(0, \epsilon\right)\right)$ as in theorem 7.4.5. In this case $|pq|$ is a minimum realized by the short geodesic in B joining p and q. Thus $p = q$.

Definition 7.5.2. We define the open ball, closed ball and distance sphere around a point $p \in M$ as:

$$\begin{aligned} B\left(p, r\right) &= \left\{x \in M \mid |px| < r\right\}, \\ \bar{B}\left(p, r\right) &= \left\{x \in M \mid |px| \leq r\right\}, \\ S\left(p, r\right) &= \left\{x \in M \mid |px| = r\right\}. \end{aligned}$$

The next corollary is almost an immediate consequence of theorem 7.4.5 and its proof now that we have introduced the concept of distance.

Corollary 7.5.3. *If $p \in M$ and $\varepsilon > 0$ is such that $\exp_p : B\left(0, \varepsilon\right) \to B$ is defined and a diffeomorphism, then for each $\delta \leq \varepsilon$,*

$$\exp_p(B(0, \delta)) = B(p, \delta),$$

and for each $\delta < \epsilon$

$$\exp_p(\bar{B}(0,\delta)) = \bar{B}(p,\delta).$$

In particular, it follows that $p_i \to p$ *if and only if* $|pp_i| \to 0$.

PROOF. We first have to show that $B(p,\varepsilon) = B$. We already have $B \subset B(p,\varepsilon)$. Conversely if $q \in B(p,\varepsilon)$, then it is joined to p by a curve $\mathbf{q}(t) \in \Omega_{pq}$ of length $< \varepsilon$. The proof of theorem 7.4.5 now shows that any curve starting at p that leaves B has length $\geq \epsilon$. This means that $\mathbf{q}(t)$ lies in O and $q \in O$. This argument can now be repeated for each $\delta < \epsilon$. This in turn also shows that $\exp_p(\bar{B}(0,\delta)) = \bar{B}(p,\delta)$ when $\delta < \epsilon$.

Finally, note that by our definition of convergence any sequence p_i that converges to p eventually must lie within the exponential parametrization of $B(p,\delta)$. The same clearly also holds if $|pp_i| \to 0$. Since this is true for all $\delta > 0$ the claim follows. $\square$

We are now ready to connect the concept of geodesic completeness with the existence of shortest curves on a larger scale.

THEOREM 7.5.4. *(Hopf-Rinow, 1931) If a surface* M *is geodesically complete at* p*, then any point* $q \in M$ *is joined to* p *by a minimal geodesic of length* $|pq|$.

PROOF. Consider p, q and choose $\epsilon > 0$ such that any point in $\bar{B}(p,\epsilon)$ can be joined to p by a unique minimal geodesic (see corollary 7.5.3). This shows that $\bar{B}(p,\epsilon)$ is homeomorphic to a disc with boundary $S(p,\epsilon)$. In particular $S(p,\epsilon)$ is compact. This shows that there exists a $q_0 \in S(p,\epsilon)$ closest to q. For this q_0 we claim that $|pq_0| + |q_0q| = |pq|$. Otherwise there would be a unit speed curve $\gamma \in \Omega_{p,q}$ with $L(\gamma) < |pq_0| + |q_0q|$. Choose t so that $\gamma(t) \in S(p,\epsilon)$. Since $t + |\gamma(t)\, q| \leq L(\gamma) < |pq_0| + |q_0q|$ it follows that $|\gamma(t)\, q| < |q_0q|$ contradicting the choice of q_0. Now let $\mathbf{q}(t)$ be the unit speed geodesic with $\mathbf{q}(0) = p$, $\mathbf{q}(\epsilon) = q_0$, and

$$A = \{t \in [0, |pq|] \mid |pq| = t + |\mathbf{q}(t)\, q|\}.$$

Clearly $0 \in A$. Also $\epsilon \in A$ since $\mathbf{q}(\epsilon) = q_0$. Note that if $t \in A$, then

$$|pq| = t + |\mathbf{q}(t)\, q| \geq |p\mathbf{q}(t)| + |\mathbf{q}(t)\, q| \geq |pq|,$$

which implies that $t = |p\mathbf{q}(t)|$. We first claim that if $t_0 \in A$, then $[0,t_0] \subset A$. Let $t < t_0$ and note that

$$\begin{aligned} |pq| &\leq |p\mathbf{q}(t)| + |\mathbf{q}(t)\, q| \\ &\leq |p\mathbf{q}(t)| + |\mathbf{q}(t)\, \mathbf{q}(t_0)| + |\mathbf{q}(t_0)\, q| \\ &\leq t + t_0 - t + |\mathbf{q}(t_0)\, q| \\ &\leq t_0 + |\mathbf{q}(t_0)\, q| \\ &= |pq|. \end{aligned}$$

This implies that $|p\mathbf{q}(t)| + |\mathbf{q}(t)\, q| = |pq|$ and $t = |p\mathbf{q}(t)|$, showing together that $t \in A$.

Since $t \mapsto |\mathbf{q}(t)\, q|$ is continuous it follows that A is closed.

Finally if $t_0 \in A$, then $t_0 + \delta \in A$ for sufficiently small $\delta > 0$. Select $\delta > 0$ so that any point in $\bar{B}(\mathbf{q}(t_0), \delta)$ can be joined to $\mathbf{q}(t_0)$ by a minimal geodesic. Then select $q_1 \in S(\mathbf{q}(t_0), \delta)$ closest to q. We now have

$$\begin{aligned} |pq| &= t_0 + |\mathbf{q}(t_0)\, q| \\ &= t_0 + |\mathbf{q}(t_0)\, q_1| + |q_1 q| \\ &= t_0 + \delta + |q_1 q| \\ &\geq |pq_1| + |q_1 q| \\ &\geq |pq|. \end{aligned}$$

It follows that $|pq_1| = t_0 + \delta$ from which we conclude that the piecewise smooth geodesic that goes from p to $\mathbf{q}(t_0)$ and then from $\mathbf{q}(t_0)$ to q_1 has length $|pq_1|$. Consequently it is a smooth geodesic and $q_1 = \mathbf{q}(t_0 + \delta)$. It then follows from $|pq| = t_0 + \delta + |q_1 q|$ that $\mathbf{q}(t_0 + \delta) \in A$. □

This in turns shows that several different completeness criteria are all equivalent.

Theorem 7.5.5. *(Hopf-Rinow, 1931) The following statements are equivalent for a surface M:*

1. *M is geodesically complete, i.e., all geodesics are defined for all time.*
2. *M is geodesically complete at p, i.e., all geodesics through p are defined for all time.*
3. *M satisfies the Heine-Borel property, i.e., every closed bounded set is compact.*
4. *M is metrically complete.*

Proof. (1)⇒(2) is trivial. (3)⇒(4) follows from the fact that Cauchy sequences are bounded.

For (4)⇒(1): If we have a unit speed geodesic $\mathbf{q} : [0, b) \to M$, then $|\mathbf{q}(t)\, \mathbf{q}(s)| \leq |t - s|$. So if $b < \infty$, it follows that $|\mathbf{q}(t)\, \mathbf{q}(s)| \to 0$ as $t, s \to b$. This shows that $\mathbf{q}(t)$ is a Cauchy sequence as $t \to b$ and by (4) must converge to a point p. In particular, $\mathbf{q}(t)$ lies in a compact set $\bar{B}(p, \delta)$ as $t \to b$. The derivative is also bounded, so it follows from theorem A.5.1 that starting at any time t_0 where $\mathbf{q}(t_0) \in \bar{B}(p, \delta)$ the geodesic exists on an interval $(-\epsilon + t_0, t_0 + \epsilon)$ where ϵ is independent of t_0. When $t_0 + \epsilon > b$ we'll have found an extension of the geodesic. This shows that the any geodesic must be defined on $[0, \infty)$.

Finally the traditionally difficult part (2)⇒(3) is an easy consequence of theorem 7.5.4. We show that $\exp_p\left(\bar{B}(0, r)\right) = \bar{B}(p, r)$ for all $r > 0$. It is clear that any point in $\exp_p\left(\bar{B}(0, r)\right)$ is joined to p by a geodesic of length $\leq r$. Thus $\exp_p\left(\bar{B}(0, r)\right) \subset \bar{B}(p, r)$. Conversely we just proved in theorem 7.5.4 that any point in $\bar{B}(p, r)$ is joined to p by a geodesic of length $\leq r$. But any such geodesic is of the form $\mathbf{q}_X(t)$ with $\mathbf{q}_X(0) = p$, $t \in [0, 1]$, and $|X| \leq r$. This shows that $\mathbf{q}_X(1) \in \exp_p\left(\bar{B}(0, r)\right)$. We now have that all of the closed balls $\bar{B}(p, r)$ are compact as they are the image of a closed ball in $\mathbb{R}^2$. Since any bounded subset of M lies in such a ball $\bar{B}(p, r)$ the Heine-Borel property follows. □

7.6. Isometries

So far we've mostly discussed how quantities remain invariant if we change parameters at a given point. Here we shall exploit more systematically what isometries can do to help us with finding and calculating geometric invariants. Recall that an isometry is simply a map that preserves the first fundamental forms. Thus isometries preserve all intrinsic notions. Isometries are also often referred to as symmetries, especially when they are maps from a surface to it self.

Corollary 7.6.1. *An isometry maps geodesics to geodesics, preserves Gauss curvature, and preserves the length of curves.*

Proof. Let $\mathbf{q}(t)$ be a geodesic and F an isometry. The geodesic equation depends only on the first fundamental form. By definition isometries preserve the first fundamental form, thus $F(\mathbf{q}(t))$ must also be a geodesic.

Next assume that F is an isometry such that $F(p) = q$. Again F preserves the first fundamental form so the Gauss curvatures must again be the same.

Finally when $\mathbf{q}(t)$ is a curve we have

$$\begin{aligned} L(F \circ \mathbf{q}) &= \int_a^b \left| \frac{d}{dt} (F(\mathbf{q}(t))) \right| dt \\ &= \int_a^b |DF(\dot{\mathbf{q}}(t))| \, dt \\ &= \int_a^b |\dot{\mathbf{q}}(t)| \, dt \\ &= L(\mathbf{q}). \end{aligned}$$

□

Corollary 7.6.2. *An isometry is distance decreasing. Moreover, if it is a bijection then it is distance preserving.*

Proof. Since isometries preserve length of curves it is clear from the definition of distance that they are distance decreasing. In case F is also a bijection it follows that F^{-1} exists and is also an isometry. Thus both F and F^{-1} are distance decreasing. This shows that they are distance preserving. □

Basic examples of isometries are rotations around the z axis for surfaces of revolution around the z axis, or mirror symmetries in meridians on a surface of revolution. The sphere has an even larger number of isometries as it is a surface of revolution around any line through the origin. The plane also has rotational and mirror symmetries, but in addition translations.

It is possible to construct isometries that do not preserve the second fundamental form. The simplest example is to imagine a flat tarp or blanket, here all points have vanishing second fundamental form and also there are isometries between all points. Now lift one side of the tarp. Part of it will still be flat on the ground, while the part that's lifted off the ground is curved. The first fundamental form has not changed but the curved part will now have nonzero entries in the second fundamental form.

It is not always possible to directly determine all isometries. But as with geodesics there are some uniqueness results that will help.

Theorem 7.6.3. *If F and G are isometries that satisfy $F(p) = G(p)$ and $DF(p) = DG(p)$, then $F = G$ in a neighborhood of p.*

Proof. We just saw that isometries preserve geodesics. So if $\mathbf{q}(t)$ is a geodesic with $\mathbf{q}(0) = p$, then $F(\mathbf{q}(t))$ and $G(\mathbf{q}(t))$ are both geodesics. Moreover they have the same initial values

$$\begin{aligned} F(\mathbf{q}(0)) &= F(p), \\ G(\mathbf{q}(0)) &= G(p), \\ \frac{d}{dt}F(\mathbf{q}(t))|_{t=0} &= DF(\dot{\mathbf{q}}(0)), \\ \frac{d}{dt}G(\mathbf{q}(t))|_{t=0} &= DG(\dot{\mathbf{q}}(0)). \end{aligned}$$

This means that $F(\mathbf{q}(t)) = G(\mathbf{q}(t))$. By varying the initial velocity of $\dot{\mathbf{q}}(0)$ we can reach all points in a neighborhood of p. $\square$

Often the best method for finding isometries is to make educated guesses based on what the metric looks like. One general guideline for creating isometries is the observation that if the first fundamental form doesn't depend on a specific variable such as v, then translations in that variable will generate isometries. This is exemplified by surfaces of revolution where the metric doesn't depend on μ. Translations in μ are the same as rotations by a fixed angle and we know that such transformations are isometries. Note that reflections in such a parameter where v is mapped to $v_0 - v$ will also be isometries in such a case.

Example 7.6.4. The linear orthogonal transformations $O(3)$ of $\mathbb{R}^3$ preserve the spheres centered at the origin. Moreover, with these transformations it is possible to solve all possible initial value problems as in theorem 7.6.3. To see this last statement we concentrate on the unit sphere. An orthonormal basis e_1, e_2 for T_pS^2 will give us an orthonormal basis e_1, e_2, p for $\mathbb{R}^3$. Let f_1, f_2, q be another orthonormal basis, i.e., f_1, f_2 is an orthonormal basis for T_qS^2. We then have two orthogonal matrices

$$\begin{bmatrix} f_1 & f_2 & q \end{bmatrix}, \begin{bmatrix} e_1 & e_2 & p \end{bmatrix} \in O(3).$$

We define $O \in O(3)$ by

$$O = \begin{bmatrix} f_1 & f_2 & q \end{bmatrix}\begin{bmatrix} e_1 & e_2 & p \end{bmatrix}^{-1}.$$

Thus

$$\begin{aligned} \begin{bmatrix} O(e_1) & O(e_2) & O(p) \end{bmatrix} &= O\begin{bmatrix} e_1 & e_2 & p \end{bmatrix} \\ &= \begin{bmatrix} f_1 & f_2 & q \end{bmatrix}\begin{bmatrix} e_1 & e_2 & p \end{bmatrix}^{-1}\begin{bmatrix} e_1 & e_2 & p \end{bmatrix} \\ &= \begin{bmatrix} f_1 & f_2 & q \end{bmatrix}. \end{aligned}$$

In other words $O(p) = q$, $O(e_1) = f_1$, and $O(e_2) = f_2$. This shows that we can solve all initial value problems.

Example 7.6.5. The isometries of $\mathbb{R}^2$ are all of the form $F(x) = Ox + q$, where $O \in O(2)$ represents the differential $O = DF(0)$ and $q \in \mathbb{R}^2$ the initial point $q = F(0)$. Theorem 7.6.3 again shows that there are no more isometries.

Example 7.6.6. The linear transformations that preserve the space-time inner product on $\mathbb{R}^{2,1}$ are denoted $O(2,1)$. They are characterized by being of the form

$O = \begin{bmatrix} e_1 & e_2 & e_3 \end{bmatrix}$, where $e_i \cdot e_j = 0$ when $i \neq j$, $|e_1|^2 = |e_2|^2 = 1$, and $|e_3|^2 = -1$. Note that

$$O \begin{bmatrix} x \\ y \\ z \end{bmatrix} = xe_1 + ye_2 + ze_3$$

and that

$$|xe_1 + ye_2 + ze_3|^2 = x^2 + y^2 - z^2.$$

This means that these transformations preserve the two sheeted hyperboloid $x^2 + y^2 - z^2 = -1$. Any given O either preserves each of the two sheets or flips the two sheets. The first case happens when O preserves H and the set of these transformations is denoted $O^+(2,1)$. We can determine when $O \in O^+(2,1)$ by checking that the 33 entry in O is positive as that means that $(0,0,1)$ is mapped to a point in H. The key observation is that any orthonormal basis e_1, e_2 for T_pH will give us an element $\begin{bmatrix} e_1 & e_2 & p \end{bmatrix} \in O^+(2,1)$. Consequently, we can, as in the sphere case, create the desired transformation using

$$O = \begin{bmatrix} f_1 & f_2 & q \end{bmatrix} \begin{bmatrix} e_1 & e_2 & p \end{bmatrix}^{-1}.$$

Here is a slightly more surprising relationship between geodesics and isometries.

THEOREM 7.6.7. *Let F be a nontrivial isometry and $\mathbf{q}(t)$ a unit speed curve such that $F(\mathbf{q}(t)) = \mathbf{q}(t)$ for all t, then $\mathbf{q}(t)$ is a geodesic.*

PROOF. Since F is an isometry and it preserves $\mathbf{q}$ we must also have that it preserves its velocity and tangential acceleration

$$\begin{aligned} DF(\dot{\mathbf{q}}(t)) &= \dot{\mathbf{q}}(t), \\ DF\left(\ddot{\mathbf{q}}^{\mathrm{I}}(t)\right) &= \ddot{\mathbf{q}}^{\mathrm{I}}(t). \end{aligned}$$

As $\mathbf{q}$ is unit speed we have $\dot{\mathbf{q}} \cdot \ddot{\mathbf{q}}^{\mathrm{I}} = 0$. If $\ddot{\mathbf{q}}^{\mathrm{I}}(t) \neq 0$, then DF preserves $\mathbf{q}(t)$ as well as the basis $\dot{\mathbf{q}}(t), \ddot{\mathbf{q}}^{\mathrm{I}}(t)$ for the tangent space at $\mathbf{q}(t)$. By the uniqueness result above this shows that F is the identity map as that map is always an isometry that fixes any point and basis. But this contradicts that F is nontrivial. □

Note that circles in the plane are preserved by rotations, but they are not fixed, nor are they geodesics. The picture we should have in mind for such an isometry and geodesic is a mirror symmetry in a line, or a mirror symmetry in a great circle on the sphere.

7.7. Constant Curvature

We've already seen many models of surfaces with constant curvature and in some cases we explicitly showed how they could be reparametrized to be isometric. This is no accident and can be done more abstractly. The goal will be to give a canonical local structure for surfaces with constant Gauss curvature. This will be done in the form of a canonical parametrization.

THEOREM 7.7.1. *(Gauss, 1827) If an abstract surface has vanishing Gauss curvature, then it admits Cartesian coordinates.*

PROOF. We use geodesic coordinates along a unit speed geodesic as in proposition 7.4.1. Thus $v \mapsto \mathbf{q}(0,v)$ is a unit speed geodesic and all of the u-curves are unit speed geodesics. The first fundamental form is

$$[\mathrm{I}] = \begin{bmatrix} 1 & 0 \\ 0 & g_{vv} \end{bmatrix}.$$

Assuming $K = 0$, the formula for the Gauss curvature

$$K = -\frac{\partial_u^2 \sqrt{g_{vv}}}{\sqrt{g_{vv}}}$$

shows that

$$\sqrt{g_{vv}(u,v)} = \sqrt{g_{vv}(0,v)} + u \cdot \left(\partial_u \sqrt{g_{vv}}\right)(0,v).$$

We also have the initial condition:

$$\sqrt{g_{vv}(0,v)} = \left|\frac{\partial \mathbf{q}}{\partial v}\right| = 1.$$

The explicit form in u, v parameters for the curve is simply $\mathbf{q}(t) = \mathbf{q}(0,t)$ so all second derivatives vanish and the velocity is pointing in the v direction. Thus the geodesic equations tell us

$$\begin{aligned} 0 &= 0 + \begin{bmatrix} 0 & 1 \end{bmatrix} \begin{bmatrix} \Gamma_{uu}^u & \Gamma_{uv}^u \\ \Gamma_{vu}^u & \Gamma_{vv}^u \end{bmatrix} \begin{bmatrix} 0 \\ 1 \end{bmatrix} \\ &= \Gamma_{vv}^u(0,v) \\ &= \Gamma_{vvu}(0,v) \\ &= \left(\partial_u \sqrt{g_{vv}}\right)(0,v). \end{aligned}$$

This shows that $\sqrt{g_{vv}(u,v)} = 1$ and hence that we have Cartesian coordinates in a neighborhood of a geodesic. □

THEOREM 7.7.2. *(Minding, 1839) If two abstract surfaces have constant Gauss curvature K, then they are locally isometric to each other.*

PROOF. It suffices to show that if a surface has constant curvature K, then it has a parametrization around every point where the first fundamental form only depends on K.

As before we fix a geodesic coordinate system $\mathbf{q}(u,v)$ where all u-curves are unit speed geodesics and $\mathbf{q}(0,v)$ is a unit speed geodesic. The first fundamental form is

$$[\mathrm{I}] = \begin{bmatrix} 1 & 0 \\ 0 & g_{vv} \end{bmatrix},$$

where as in the proof above:

$$\begin{aligned} \sqrt{g_{vv}(0,v)} &= 1, \\ \left(\partial_u \sqrt{g_{vv}}\right)(0,v) &= 0, \end{aligned}$$

and

$$K = -\frac{\partial_u^2 \sqrt{g_{vv}}}{\sqrt{g_{vv}}}.$$

The last equation dictates how $\sqrt{g_{vv}}$ changes along u curves and the two previous equations are the initial values. When $K=0$ we saw that $\sqrt{g_{vv}}=1$, otherwise

$$\sqrt{g_{vv}(u,v)}=\begin{cases}\cos\left(\sqrt{K}u\right), & K>0,\\ \cosh\left(\sqrt{-K}u\right), & K<0.\end{cases}$$

$\square$

THEOREM 7.7.3. *Any complete simply connected surface M with constant curvature k is bijectively isometric to S_k^2.*

PROOF. We know from theorem 7.7.2 that given $x\in M$ sufficiently small balls $B(x,r)\subset M$ are isometric to balls $B(\bar{x},r)\subset S_k^2$. Furthermore, if $q\in B(x,r)$, $\bar{q}\in S_k^2$, and $L:T_qM\to T_{\bar{q}}S_k^2$ is a linear isometry, then there is a unique bijective isometry $F:B(x,r)\to B(F(x),r)\subset S_k^2$, where $F(q)=\bar{q}$ and $DF|_q=L$. Note that when $k\leq 0$, all metric balls in S_k^2 are convex, while when $k>0$ we need their radius to be $<\frac{\pi}{2\sqrt{k}}$ for this to be true. For the remainder of the proof assume that all metric balls are chosen to be isometric to convex balls in the space form. So for small radii the metrics balls are either disjoint or have connected intersection.

The construction of $F:M\to S_k^2$ proceeds basically in the same way one does analytic continuation on simply connected domains. Fix base points $p\in M$, $\bar{p}\in S_k^2$ and a linear isometry $L:T_pM\to T_{\bar{p}}S_k^2$. Next, let $x\in M$ be an arbitrary point. If $c\in\Omega_{p,x}$ is a curve from p to x in M, then we can cover c by a string of balls $B(p_i,r)$, $i=0,...,k$, where $p=p_0$, $x=p_k$, and $B(p_{i-1},r)\cap B(p_i,r)\neq\emptyset$. Define $F_0:B(p_0,r)\to S_k^2$ so that $F(p)=\bar{p}$ and $DF_0|_{p_0}=L$. Then define $F_i:B(p_i,r)\to S_k^2$ successively to make it agree with F_{i-1} on $B(p_{i-1},r)\cap B(p_i,r)$ (this just requires their values and differentials agree at one point). Define a function $G:\Omega_{p,x}\to S_k^2$ by $G(c)=F_k(x)$. We have to check that it is well-defined in the sense that it doesn't depend on our specific way of covering the curve. This is easily done by selecting a different covering and then showing that the set of values in $[0,1]$ where the two choices agree is both open and closed.

If $\bar{c}\in\Omega_{p,x}$ is sufficiently close to c, then it lies inside a fixed covering of c, but then it is clear that $G(c)=G(\bar{c})$. This implies that G is locally constant. In particular, G has the same value on all curves in $\Omega_{p,x}$ that are homotopic to each other. Simple-connectivity simply means that all curves are homotopic to each other so G is constant on $\Omega_{p,x}$. This means that $F(x)$ becomes well-defined and a Riemannian isometry.

If M is geodesically complete at a point p, then any point $x\in M$ lies on a unit speed geodesic $\mathbf{q}(t):[0,\infty)\to M$ so that $\mathbf{q}(0)=p$. The map F will take this to a unit speed geodesic from $\bar{p}$. Now any point in S_k^2 lies on a unit speed geodesic that starts at $\bar{p}$, so this shows that F is onto.

If $F(x)=F(y)$, then we have two unit speed geodesics emanating from $\bar{p}$ that intersect at $F(x)=F(y)$. When $k\leq 0$ this is impossible unless the geodesics agree. Thus F is both onto and one-to-one when $k\leq 0$.

In case $k>0$ two unit speed geodesics in S_k^2 that start at $\bar{p}$ can only intersect at the antipodal point $-\bar{p}$. So if we have two different unit speed geodesics $\mathbf{q}_1,\mathbf{q}_2:[0,\infty)\to M$ with $\mathbf{q}_i(0)=p$. Then $F\circ\mathbf{q}_i(t)$ are different unit speed geodesics emanating from $\bar{p}$ that intersect when $t=n\pi/\sqrt{k}$, $n=1,2,3....$ In particular, $F:B(p,\pi/\sqrt{k})\to S_k^2-\{-\bar{p}\}$ is one-to-one and $F(S(p,\pi/\sqrt{k}))=\{-\bar{p}\}$. Then $F^{-1}:S_k^2-\{-\bar{p}\}\to B(p,\pi/\sqrt{k})$ is a well-defined isometry that maps points close to $\bar{p}$ to points that are close to $S\left(p, \pi/\sqrt{k}\right)$. Since points that are close to $\bar{p}$ are also close to each other it must follow that $S\left(p, \pi/\sqrt{k}\right)$ consists of a single point q. This shows that all geodesics that start at p go through q. We can then conclude that $\bar{B}\left(p, \pi/\sqrt{k}\right) = M$ and that $F : M \to S_k^2$ is one-to-one. $\square$

7.8. Comparison Results

In this section we prove several classical results for surfaces where the Gauss curvature is either bounded from below or above. Such results are often referred to as comparison results since they are obtained by a comparison with a corresponding constant curvature geometry.

We start by analyzing the second derivative of energy for some very specific variations.

LEMMA 7.8.1. *(Jacobi, 1842) Let $\mathbf{q}(u,v)$ be geodesic coordinates where all u-curves are geodesics along a unit speed geodesic $\mathbf{q}(0,v)$. Consider a variation: $u = su(t)$ and $v = t$, i.e., $\mathbf{q}(s,t) = \mathbf{q}(su(t), t)$, then*

$$\frac{d^2E}{ds^2}|_{s=0} = \int_a^b \left(\dot{u}^2 - Ku^2\right) dt.$$

PROOF. We write the velocity out in coordinates

$$\frac{\partial \mathbf{q}}{\partial t} = s\dot{u}\partial_u \mathbf{q} + \partial_v \mathbf{q}$$

and obtain

$$\mathrm{I}\left(\frac{\partial \mathbf{q}}{\partial t}, \frac{\partial \mathbf{q}}{\partial t}\right) = s^2\dot{u}^2 + g_{vv}.$$

For fixed s the energy of $t \mapsto \mathbf{q}(su(t), t)$ is given by

$$E(s) = \frac{1}{2}\int_a^b \left(s^2\dot{u}^2 + g_{vv}\right) dt,$$

Keeping in mind that $g_{vv} = g_{vv}(su(t), t)$ the derivatives are easily calculated:

$$\frac{dE}{ds} = \int_a^b \left(s\dot{u}^2 + \frac{1}{2}u\partial_u g_{vv}\right) dt,$$

$$\frac{d^2E}{ds^2} = \int_a^b \left(\dot{u}^2 + \frac{1}{2}u^2\partial_u^2 g_{vv}\right) dt.$$

This is related to the Gauss curvature through the formula

$$K = -\frac{1}{2}\left(\frac{\partial_u^2 g_{vv}}{g_{vv}} - \left(\frac{\partial_u g_{vv}}{g_{vv}}\right)^2\right).$$

Since $\mathbf{q}(0,v)$ is a unit speed curve we have $g_{vv}(0,v) = 1$. The derivative is calculated as follows

$$\begin{aligned}
\partial_u g_{vv} &= 2\mathrm{I}\left((\partial_u\partial_v \mathbf{q})^{\mathrm{I}}, \partial_v \mathbf{q}\right) \\
&= 2\mathrm{I}\left((\partial_v\partial_u \mathbf{q})^{\mathrm{I}}, \partial_v \mathbf{q}\right) \\
&= 2\partial_v \mathrm{I}(\partial_u \mathbf{q}, \partial_v \mathbf{q}) - 2\mathrm{I}\left(\partial_u \mathbf{q}, \left(\partial_v^2 \mathbf{q}\right)^{\mathrm{I}}\right) \\
&= 2\partial_v g_{uv} - 2\mathrm{I}\left(\partial_u \mathbf{q}, \left(\partial_v^2 \mathbf{q}\right)^{\mathrm{I}}\right) \\
&= -2\mathrm{I}\left(\partial_u \mathbf{q}, \left(\partial_v^2 \mathbf{q}\right)^{\mathrm{I}}\right).
\end{aligned}$$

This vanishes when $u = 0$ since $\mathbf{q}(0,v)$ is a geodesic. The result now follows. □

Corollary 7.8.2. *(Bonnet, 1855) If $K \geq R^{-2} > 0$, then no geodesic of length $> \pi R$ is minimal.*

Proof. We can assume that the geodesic doesn't intersect itself (if it does it is clearly not minimal) and construct geodesic coordinates where $\mathbf{q}(0,v)$ is the given geodesic parametrized by arclength on $[0, L]$. Then select a variation as in lemma 7.8.1 of the form $u(t) = \sin(t\pi/L)$. This will yield a proper variation with the second derivative of energy satisfying

$$\begin{aligned}
\frac{d^2E}{ds^2}|_{s=0} &= \int_0^L \left(\dot{u}^2 - Ku^2\right) dt \\
&\leq \int_0^L \left(\left(\frac{\pi}{L}\right)^2 \cos^2(t\pi/L) - R^{-2}\sin^2(t\pi/L)\right) dt \\
&= \left(\frac{\pi}{L}\right)^2 \int_0^L \cos^2(t\pi/L)\, dt - R^{-2}\int_0^L \sin^2(t\pi/L)\, dt \\
&= \left(\left(\frac{\pi}{L}\right)^2 - R^{-2}\right)\frac{L}{2}.
\end{aligned}$$

This is strictly negative when $L > \pi R$ showing that the geodesic is a local maximum for the energy. Since the variation is fixed at the end points there will be nearby curves of strictly smaller energy with the same end points. Corollary 7.3.6 then shows that it can't be a minimum for the length functional. □

Corollary 7.8.3. *(Hopf-Rinow, 1931) If a complete surface satisfies $K \geq R^{-2} > 0$, then all distances are $\leq \pi R$ and must in particular be a closed surface.*

Theorem 7.8.4. *If a closed surface has positive curvature, then any two closed geodesics intersect.*

Proof. Assume otherwise and obtain a shortest geodesic between the two closed geodesics. This geodesic is perpendicular to both of the closed geodesics. In particular if we let it be the $\mathbf{q}(0,v)$ curve in a geodesic parametrization, then the curves $\mathbf{q}(u,0)$ and $\mathbf{q}(u,L)$ are our two closed geodesics. Now consider the variation where $s = u$ and $t = v$, then the second variation is given by

$$\frac{d^2E}{ds^2}|_{s=0} = \int_a^b \left(\dot{u}^2 - Ku^2\right) dt = \int_a^b -Ku^2 dt < 0.$$

This shows that the curves $v \mapsto \mathbf{q}(u, v)$ are shorter than L. As they are also curves between the two closed geodesics this contradicts that our original curve was the shortest such curve. $\square$

THEOREM 7.8.5. *(Mangoldt, 1881 Hadamard, 1889?) A complete surface M with $K \leq 0$ admits a global parametrization $\mathbf{q}(u, v)$ where $(u, v) \in \mathbb{R}^2$. If on $\mathbb{R}^2$ we introduce the first fundamental form from M, then we obtain a complete metric on $\mathbb{R}^2$ with $K \leq 0$ where all geodesics are minimal.*

PROOF. The parametrization is given by the exponential map. Identify a fixed tangent space T_pM with $\mathbb{R}^2$ via a choice of orthonormal basis E_1, E_2 and introduce Cartesian (x, y) as well as polar coordinates (r, θ). We can use $\mathbf{q}(r, \theta) = \exp_p(r\cos\theta E_1 + r\sin\theta E_2)$ as a potential parametrization on M. Even when it isn't a parametrization as in lemma 7.4.4 we note that it is a geodesic variation with the radial lines as unit speed geodesics. We have the velocity fields $\partial_r\mathbf{q}, \partial_\theta\mathbf{q}$ for the r- and θ-curves which for each (r, θ) give us tangent vectors in $T_{\mathbf{q}(r,\theta)}M$. Since the r-curves are unit speed geodesics we have $|\partial_r\mathbf{q}| = 1$ everywhere. We can also show that $\mathrm{I}(\partial_r\mathbf{q}, \partial_\theta\mathbf{q}) = 0$. First note that it vanishes at $r = 0$ since $\partial_\theta\mathbf{q}(0, \theta) = 0$. Next see that $\mathrm{I}(\partial_r\mathbf{q}, \partial_\theta\mathbf{q}) = 0$ is constant since

$$\begin{aligned}
\partial_r \mathrm{I}(\partial_r\mathbf{q}, \partial_\theta\mathbf{q}) &= \mathrm{I}\left(\left(\partial_r^2\mathbf{q}\right)^{\mathrm{I}}, \partial_\theta\mathbf{q}\right) + \mathrm{I}\left(\partial_r\mathbf{q}, (\partial_r\partial_\theta\mathbf{q})^{\mathrm{I}}\right) \\
&= \mathrm{I}\left(\partial_r\mathbf{q}, (\partial_\theta\partial_r\mathbf{q})^{\mathrm{I}}\right) \\
&= \frac{1}{2}\partial_\theta \mathrm{I}(\partial_r\mathbf{q}, \partial_r\mathbf{q}) \\
&= 0.
\end{aligned}$$

Thus $\mathrm{I}(\partial_r\mathbf{q}, \partial_\theta\mathbf{q}) = 0$ everywhere. It follows that $D\exp_p$ is nonsingular at a point (r, θ) precisely when $\mathrm{I}(\partial_\theta\mathbf{q}, \partial_\theta\mathbf{q}) > 0$ at (r, θ).

Define a first fundamental form on $\mathbb{R}^2$ by

$$\begin{bmatrix} g_{rr} & g_{r\theta} \\ g_{\theta r} & g_{\theta\theta} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & g_{\theta\theta} \end{bmatrix},$$

where

$$g_{\theta\theta} = \left|D\exp_p(-yE_1 + xE_2)\right|^2 = \mathrm{I}(\partial_\theta\mathbf{q}, \partial_\theta\mathbf{q}).$$

When $D\exp_p$ is nonsingular this corresponds precisely to the first fundamental form of M in this parametrization.

By continuity $\exp_p : T_pM \to M$ is nonsingular on some open set O that contains the origin. Let $B(0, R) \subset O$ be the largest ball inside O. We claim that $R = \infty$ and note that if $R < \infty$ then the closure $\bar{B}(0, R)$ cannot be contained in O. On $B(0, R)$ the (r, θ)-coordinates are geodesic polar coordinates with respect to

$$\begin{bmatrix} 1 & 0 \\ 0 & g_{\theta\theta} \end{bmatrix}.$$

Since they correspond to the first fundamental form on M the Gauss curvature satisfies

$$0 \leq K = -\frac{\partial_r^2\sqrt{g_{\theta\theta}}}{\sqrt{g_{\theta\theta}}}.$$

In particular, $\partial_r^2\sqrt{g_{\theta\theta}} \geq 0$. The initial values at $r = 0$ for $\sqrt{g_{\theta\theta}}$ are:

$$\sqrt{g_{\theta\theta}(0, \theta)} = 0,\ \left(\partial_r\sqrt{g_{\theta\theta}}\right)(0, \theta) = 1.$$

This follows directly by using that $D\exp_p$ is the identity at $r=0$

$$\begin{aligned}\sqrt{g_{\theta\theta}(r,\theta)} &= \left|D\exp_p(-yE_1+xE_2)\right| \\ &= |-yE_1+xE_2|+O\left(r^2\right) \\ &= r+O\left(r^2\right).\end{aligned}$$

Together with $\partial_r^2\sqrt{g_{\theta\theta}}\geq 0$ this this shows that $\sqrt{g_{\theta\theta}(r,\theta)}\geq r$. If we let $r\to R$ this shows that $g_{\theta\theta}(R,\theta)$ does not vanish when $R<\infty$, in particular, $B(0,R)\subset O$ can't be maximal unless $R=\infty$.

This gives us the desired global parametrization on M with a first fundamental form on $\mathbb{R}^2$ that has $K\leq 0$. This will also help us establish the second part of the result. In fact, no metric on $\mathbb{R}^2$ with $K\leq 0$ can have geodesics that intersect at more than one point as that would violate Gauss-Bonnet. Consider two geodesics $\mathbf{q}_1(t)$ and $\mathbf{q}_2(t)$ with $\mathbf{q}_i(0)=p$. By lemma 7.4.4 they can't intersect near p. Therefore, if they intersect at some later point, then there will a point $q\neq p$ closest to p where they intersect. In this case we can after reparametrizing assume that $\mathbf{q}_i(1)=q$ and that when restricted to $t\in[0,1]$ there are no other intersections between the geodesics. Now create a triangle by using p, q, and, say $\mathbf{q}_1(1/2)$, as vertices. This triangle has angle sum $>\pi$ as one angle is π. This however, violates the Gauss-Bonnet theorem as the whole triangle is a closed simple curve of rotation index 2π when oriented appropriately. Specifically, as the geodesic curvature vanishes the Gauss-Bonnet theorem 6.5.2 tells us

$$0\geq\int_{\mathbf{q}(R)}KdA=2\pi-\sum\theta_i,$$

where θ_i are the exterior angles at the three vertices. Since they are complementary to the interior angles α,β,γ we have

$$0\geq\int_{\mathbf{q}(R)}KdA=2\pi-\sum\theta_i=-\pi+\alpha+\beta+\gamma.$$

$\square$

REMARK 7.8.6. There are different proofs of the latter part that do not appeal to the Gauss-Bonnet theorem.

CHAPTER 8

Riemannian Geometry

As with abstract surfaces we simply define what the dot products of the tangent fields should be:

$$[\mathrm{I}] = \begin{bmatrix} \frac{\partial \mathbf{q}}{\partial u^1} & \cdots & \frac{\partial \mathbf{q}}{\partial u^n} \end{bmatrix}^t \begin{bmatrix} \frac{\partial \mathbf{q}}{\partial u^1} & \cdots & \frac{\partial \mathbf{q}}{\partial u^n} \end{bmatrix} = \begin{bmatrix} g_{11} & \cdots & g_{1n} \\ \vdots & \ddots & \vdots \\ g_{n1} & \cdots & g_{nn} \end{bmatrix}$$

The notation $\frac{\partial \mathbf{q}}{\partial u^i} = \partial_i \mathbf{q}$ for the tangent field that corresponds to the velocity of the u^i curves is borrowed from our view of what happens on a surface.

We have the very general formula for how vectors are expanded

$$\begin{aligned} V &= \begin{bmatrix} U_1 & \cdots & U_n \end{bmatrix} \left(\begin{bmatrix} U_1 & \cdots & U_n \end{bmatrix}^t \begin{bmatrix} U_1 & \cdots & U_n \end{bmatrix} \right)^{-1} \begin{bmatrix} U_1 & \cdots & U_n \end{bmatrix}^t V \\ &= \begin{bmatrix} U_1 & \cdots & U_n \end{bmatrix} \begin{bmatrix} U_1 \cdot U_1 & \cdots & U_1 \cdot U_n \\ \vdots & \ddots & \vdots \\ U_n \cdot U_1 & \cdots & U_n \cdot U_n \end{bmatrix}^{-1} \begin{bmatrix} U_1 \cdot V \\ \vdots \\ U_n \cdot V \end{bmatrix} \end{aligned}$$

provided we know how to compute dot products of the basis vectors and dots products of V with the basis vectors. So we will now assume that were are given a symmetric matrix $[\mathrm{I}] = [g_{ij}]$ of function on some domain $U \subset \mathbb{R}^n$ that uses u^i as parameters. We shall further assume that this first fundamental form has non-vanishing determinant so that we can calculate the inverse $[\mathrm{I}]^{-1} = \left[g^{ij}\right]$. We shall then think of $g_{ij} = \mathrm{I}(\partial_i \mathbf{q}, \partial_j \mathbf{q})$ as describing the inner product of the coordinate vector fields and $\mathbf{q}$ as a point on the space we are investigating. When dealing with surfaces we also used that this defined an inner product. For the moment we will not need this condition.

We can define the Christoffel symbols in relation to the tangent fields when we know the dot products of those tangent fields:

$$\begin{aligned} \Gamma_{ijk} &= \frac{1}{2} \left(\partial_j g_{ki} + \partial_i g_{kj} - \partial_k g_{ij} \right), \\ \Gamma_{ij}^k &= \sum_l g^{kl} \Gamma_{ijl}. \end{aligned}$$

Proposition 8.0.7. *The metric and Christoffel symbols are also related by*

$$\begin{aligned} \partial_k g_{ij} &= \Gamma_{kij} + \Gamma_{kji} \\ \partial_k g^{ij} &= -\sum_l g^{il} \Gamma_{kl}^j + g^{jl} \Gamma_{kl}^i \end{aligned}$$

PROOF. The first formula follows directly from the definition

$$\begin{aligned}\Gamma_{kij}+\Gamma_{kji} &= \frac{1}{2}\left(\partial_k g_{ij}+\partial_i g_{kj}-\partial_j g_{ki}\right)\\ &\quad +\frac{1}{2}\left(\partial_k g_{ji}+\partial_j g_{ki}-\partial_i g_{kj}\right)\\ &= \partial_k g_{ji}.\end{aligned}$$

For the second we first have to calculate the derivative of the inverse of a matrix. Symbolically this is done as follows. If $I_n=\left[\delta^i_j\right]$ denotes the identity matrix then

$$\begin{aligned}I_n &= [\mathrm{I}]\,[\mathrm{I}]^{-1}\\ \delta^j_i &= g_{ik}g^{kj}\end{aligned}$$

so

$$\begin{aligned}0=\partial_s I_n &= \left(\partial_s[\mathrm{I}]\right)[\mathrm{I}]^{-1}+[\mathrm{I}]\,\partial_s[\mathrm{I}]^{-1}\\ 0=\partial_s\delta^j_i &= \sum_l \partial_s g_{il}g^{lj}+\sum_k g_{ik}\partial_s g^{kj}\end{aligned}$$

showing that

$$\begin{aligned}\partial_s[\mathrm{I}]^{-1} &= -[\mathrm{I}]^{-1}\left(\partial_s[\mathrm{I}]\right)[\mathrm{I}]^{-1}\\ \partial_s g^{kj} &= -\sum_{i,l} g^{ki}\partial_s g_{il}g^{lj}\end{aligned}$$

We can now use the first formula to prove the second

$$\begin{aligned}\partial_k g^{ij} &= -\sum_{s,t} g^{is}\partial_k g_{st}g^{tj}\\ &= -\sum_{s,t} g^{is}\left(\Gamma_{kst}+\Gamma_{kts}\right)g^{tj}\\ &= -\sum_{s,t} g^{is}\Gamma_{kst}g^{tj}-\sum_{s,t} g^{is}\Gamma_{kts}g^{tj}\\ &= -\sum_s g^{is}\Gamma^j_{ks}-\sum_t \Gamma^i_{kt}g^{tj}\\ &= -\sum_l g^{il}\Gamma^j_{kl}+g^{jl}\Gamma^i_{kl}\end{aligned}$$

□

While we have not yet specified where $\mathbf{q}$ is placed we can still attempt to define second partials intrinsically. This means that we imitate what happened for surfaces but assume that there is no normal vector.

To start with we should have

$$\partial^2_{ij}\mathbf{q}\cdot\partial_k\mathbf{q}=\Gamma_{ijk}$$

leading to

$$\begin{aligned}
\partial^2_{ij}\mathbf{q} &= \begin{bmatrix} \partial_1\mathbf{q} & \cdots & \partial_n\mathbf{q} \end{bmatrix} [\mathrm{I}]^{-1} \begin{bmatrix} \Gamma_{ij1} & \cdots & \Gamma_{ijn} \end{bmatrix}^t \\
&= \begin{bmatrix} \partial_1\mathbf{q} & \cdots & \partial_n\mathbf{q} \end{bmatrix} \begin{bmatrix} \Gamma^1_{ij} \\ \vdots \\ \Gamma^n_{ij} \end{bmatrix}
\end{aligned}$$

Note that the symmetry of the metric and Christoffel symbols now tell us that we still have

$$\partial^2_{ij}\mathbf{q} = \partial^2_{ji}\mathbf{q}$$

This will allow us to define intrinsic acceleration and hence geodesics. It'll also allow us to show that the stationary curves for energy are geodesics. If in addition the metric is positive definite, i.e., $\mathrm{I}(V,V) > 0$ unless $V = 0$, then we can define the length of vectors and consider arc-length of curves. It will then also be true that short geodesics minimize arc-length.

To define curvature we collect the Gauss formulas

$$\begin{aligned}
\partial_i \begin{bmatrix} \partial_1\mathbf{q} & \cdots & \partial_n\mathbf{q} \end{bmatrix} &= \begin{bmatrix} \partial_1\mathbf{q} & \cdots & \partial_n\mathbf{q} \end{bmatrix} \begin{bmatrix} \Gamma^1_{i1} & \cdots & \Gamma^1_{in} \\ \vdots & \ddots & \vdots \\ \Gamma^n_{i1} & \cdots & \Gamma^n_{in} \end{bmatrix} \\
&= \begin{bmatrix} \partial_1\mathbf{q} & \cdots & \partial_n\mathbf{q} \end{bmatrix} [\Gamma_i]
\end{aligned}$$

and form the expression

$$\partial_i [\Gamma_j] - \partial_j [\Gamma_i] + [\Gamma_i][\Gamma_j] - [\Gamma_j][\Gamma_i]$$

that we used to define the curvatures involved in the Gauss equations.

This time we don't have a Gauss curvature, but we can define the Riemann curvature as the k, l entry in this expression:

$$\begin{aligned}
[R_{ij}] &= \partial_i [\Gamma_j] - \partial_j [\Gamma_i] + [\Gamma_i][\Gamma_j] - [\Gamma_j][\Gamma_i], \\
R^l_{ijk} &= \partial_i\Gamma^l_{jk} - \partial_j\Gamma^l_{ik} + \begin{bmatrix} \Gamma^l_{i1} & \cdots & \Gamma^l_{in} \end{bmatrix} \begin{bmatrix} \Gamma^1_{jk} \\ \vdots \\ \Gamma^n_{jk} \end{bmatrix} - \begin{bmatrix} \Gamma^l_{j1} & \cdots & \Gamma^l_{jn} \end{bmatrix} \begin{bmatrix} \Gamma^1_{ik} \\ \vdots \\ \Gamma^n_{ik} \end{bmatrix}
\end{aligned}$$

This expression shows how certain third order partials might not commute as this means that

$$\partial^3_{ijk}\mathbf{q} - \partial^3_{jik}\mathbf{q} = \begin{bmatrix} \partial_1\mathbf{q} & \cdots & \partial_n\mathbf{q} \end{bmatrix} \begin{bmatrix} R^1_{ijk} \\ \vdots \\ R^n_{ijk} \end{bmatrix}$$

But recall that since second order partials do commute we have

$$\partial^3_{ijk}\mathbf{q} = \partial^3_{ikj}\mathbf{q}$$

So we see that third order partials commute if and only if the Riemann curvature vanishes. This can be used to establish the difficult existence part of the next result.

THEOREM 8.0.8. [Riemann] *The Riemann curvature vanishes if and only if there are Cartesian coordinates around any point.*

Proof. The easy direction is to assume that Cartesian coordinates exist. Certainly this shows that the curvatures vanish when we use Cartesian coordinates, but this does not guarantee that they also vanish in some arbitrary coordinate system. For that we need to figure out how the curvature terms change when we change coordinates. A long tedious calculation shows that if the new coordinates are called v^i and the curvature in these coordinates $\tilde{R}^l_{ijk}$, then

$$\tilde{R}^l_{ijk} = \frac{\partial u^\alpha}{\partial v^i}\frac{\partial u^\beta}{\partial v^j}\frac{\partial u^\gamma}{\partial v^k}\frac{\partial v^l}{\partial u^\delta}R^\delta_{\alpha\beta\gamma}.$$

Thus we see that if the all curvatures vanish in one coordinate system, then they vanish in all coordinate systems.

Conversely, to find Cartesian coordinates we set up a system of differential equations

$$\begin{aligned} \partial_i \mathbf{q} &= U_i \\ \partial_i \begin{bmatrix} U_1 & \cdots & U_n \end{bmatrix} &= \begin{bmatrix} U_1 & \cdots & U_n \end{bmatrix} [\Gamma_i] \end{aligned}$$

whose integrability conditions are a consequence of having vanishing curvature. We select a point $u_0 \in U$ in our given parametrization and assume that we are looking for a map $\mathbf{q} : U \to \mathbb{R}^n$ where $\mathbf{q}(u_0) = 0$ and $U_i(u_0) = u_i$ a suitable basis for $\mathbb{R}^n$.

The integrability conditions for the first set of equations are

$$\partial_i U_j = \partial_j U_i$$

which from the second set of equations mean that

$$\begin{bmatrix} U_1 & \cdots & U_n \end{bmatrix} \begin{bmatrix} \Gamma^1_{ij} \\ \vdots \\ \Gamma^n_{ij} \end{bmatrix} = \begin{bmatrix} U_1 & \cdots & U_n \end{bmatrix} \begin{bmatrix} \Gamma^1_{ji} \\ \vdots \\ \Gamma^n_{ji} \end{bmatrix}$$

These conditions holds since $\Gamma^k_{ij} = \Gamma^k_{ji}$.

For the second set of equations the integrability conditions are given by

$$\partial_i \left(\begin{bmatrix} U_1 & \cdots & U_n \end{bmatrix} [\Gamma_j] \right) = \partial_j \left(\begin{bmatrix} U_1 & \cdots & U_n \end{bmatrix} [\Gamma_i] \right)$$

which we know reduce to

$$\partial_i [\Gamma_j] + [\Gamma_i][\Gamma_j] = \partial_j [\Gamma_i] + [\Gamma_j][\Gamma_i]$$

These conditions hold because we assume that $[R_{ij}] = 0$.

This means that we can solve these equations on some neighborhood of $u_0 \in U$ with the specified initial conditions. We then have to show that the new parametrization is Cartesian. The new parameters are given by the coordinates for $\mathbf{q}$, i.e.,

$$\left(x^1, ..., x^n\right) = \left(\mathbf{q}^1, ..., \mathbf{q}^n\right) = \mathbf{q}\left(u^1, ..., u^n\right)$$

Thus

$$\partial_j \mathbf{q}^k = \frac{\partial \mathbf{q}^k}{\partial u^j} = U^k_j$$

and

$$\partial_{ij}\mathbf{q}^k = \partial_i U^k_j = U^k_l \Gamma^l_{ij} = \partial_l \mathbf{q}^k \Gamma^l_{ij}$$

The new first fundamental form is then given by

$$\begin{aligned}\tilde{g}_{kl} &= \frac{\partial u^i}{\partial x^k} g_{ij} \frac{\partial u^j}{\partial x^l} \\ [\tilde{g}_{kl}] &= \left[\frac{\partial u^i}{\partial x^k}\right][g_{ij}]\left[\frac{\partial u^j}{\partial x^l}\right]\end{aligned}$$

Unfortunately we don't know what the matrix $\left[\frac{\partial u^i}{\partial x^k}\right]$ is. It is given as the inverse of $\left[\frac{\partial x^k}{\partial u^i}\right]$ which in turn is the matrix $\begin{bmatrix} U_1 & \cdots & U_n \end{bmatrix}$ by our first equations. This means that we have

$$\tilde{g}^{kl} = \frac{\partial x^k}{\partial u^i} g^{ij} \frac{\partial u^l}{\partial x^j} = \partial_i \mathbf{q}^k g^{ij} \partial_j \mathbf{q}^l$$

We can now calculate the derivative of this as

$$\begin{aligned}\partial_s \tilde{g}^{kl} &= \partial_{si} \mathbf{q}^k g^{ij} \partial_j \mathbf{q}^l + \partial_i \mathbf{q}^k g^{ij} \partial_{sj} \mathbf{q}^l \\ &\quad + \partial_i \mathbf{q}^k \partial_s g^{ij} \partial_j \mathbf{q}^l \\ &= \partial_t \mathbf{q}^k \Gamma^t_{si} g^{ij} \partial_j \mathbf{q}^l + \partial_i \mathbf{q}^k g^{ij} \partial_t \mathbf{q}^l \Gamma^t_{sj} \\ &\quad - \partial_i \mathbf{q}^k \left(g^{it}\Gamma^j_{st} + g^{jt}\Gamma^i_{st}\right) \partial_j \mathbf{q}^l \\ &= \partial_t \mathbf{q}^k \Gamma^t_{si} g^{ij} \partial_j \mathbf{q}^l - \partial_i \mathbf{q}^k g^{jt} \Gamma^i_{st} \partial_j \mathbf{q}^l \\ &\quad + \partial_i \mathbf{q}^k g^{ij} \partial_t \mathbf{q}^l \Gamma^t_{sj} - \partial_i \mathbf{q}^k g^{it} \Gamma^j_{st} \partial_j \mathbf{q}^l \\ &= 0 + 0\end{aligned}$$

showing that the new metric coefficients are constant. We can then specify the basis u_i so that the new metric becomes Cartesian at u_0 and hence Cartesian everywhere since the metric coefficients are constant. □

APPENDIX A

Vector Calculus

A.1. Vector and Matrix Notation

Given a basis e, f for a two-dimensional vector space we expand vectors using matrix multiplication

$$v = v^e e + v^f f = \begin{bmatrix} e & f \end{bmatrix} \begin{bmatrix} v^e \\ v^f \end{bmatrix}$$

The matrix representation $[L]$ for a linear map/transformation L can be found from

$$\begin{aligned} \begin{bmatrix} L(e) & L(f) \end{bmatrix} &= \begin{bmatrix} e & f \end{bmatrix} [L] \\ &= \begin{bmatrix} e & f \end{bmatrix} \begin{bmatrix} L_e^e & L_f^e \\ L_e^f & L_f^f \end{bmatrix} \end{aligned}$$

Next we relate matrix multiplication and the dot product in $\mathbb{R}^3$. We think of vectors as being columns or 3×1 matrices. Keeping that in mind and using transposition of matrices we immediately obtain:

$$\begin{aligned} X^t Y &= X \cdot Y, \\ X^t \begin{bmatrix} X_2 & Y_2 \end{bmatrix} &= \begin{bmatrix} X \cdot X_2 & X \cdot Y_2 \end{bmatrix} \\ \begin{bmatrix} X_1 & Y_1 \end{bmatrix}^t X &= \begin{bmatrix} X_1 \cdot X \\ Y_1 \cdot X \end{bmatrix} \\ \begin{bmatrix} X_1 & Y_1 \end{bmatrix}^t \begin{bmatrix} X_2 & Y_2 \end{bmatrix} &= \begin{bmatrix} X_1 \cdot X_2 & X_1 \cdot Y_2 \\ Y_1 \cdot X_2 & Y_1 \cdot Y_2 \end{bmatrix}, \\ \begin{bmatrix} X_1 & Y_1 & Z_1 \end{bmatrix}^t \begin{bmatrix} X_2 & Y_2 & Z_2 \end{bmatrix} &= \begin{bmatrix} X_1 \cdot X_2 & X_1 \cdot Y_2 & X_1 \cdot Z_2 \\ Y_1 \cdot X_2 & Y_1 \cdot Y_2 & Y_1 \cdot Z_2 \\ Z_1 \cdot X_2 & Z_1 \cdot Y_2 & Z_1 \cdot Z_2 \end{bmatrix} \end{aligned}$$

These formulas can be used to calculate the coefficients of a vector with respect to a general basis. Recall first that if E_1, E_2 is an orthonormal basis for $\mathbb{R}^2$, then

$$\begin{aligned} X &= (X \cdot E_1) E_1 + (X \cdot E_2) E_2 \\ &= \begin{bmatrix} E_1 & E_2 \end{bmatrix} \begin{bmatrix} E_1 & E_2 \end{bmatrix}^t X \end{aligned}$$

So the coefficients for X are simply the dot products with the basis vectors. More generally we have

Theorem A.1.1. *Let U, V be a basis for $\mathbb{R}^2$, then*

$$\begin{aligned} X &= \begin{bmatrix} U & V \end{bmatrix} \left(\begin{bmatrix} U & V \end{bmatrix}^t \begin{bmatrix} U & V \end{bmatrix} \right)^{-1} \begin{bmatrix} U & V \end{bmatrix}^t X \\ &= \begin{bmatrix} U & V \end{bmatrix} \left(\begin{bmatrix} U & V \end{bmatrix}^t \begin{bmatrix} U & V \end{bmatrix} \right)^{-1} \begin{bmatrix} U \cdot X \\ V \cdot X \end{bmatrix} \end{aligned}$$

PROOF. First write

$$X = \begin{bmatrix} U & V \end{bmatrix} \begin{bmatrix} X^u \\ X^v \end{bmatrix}$$

The goal is to find a formula for the coefficients X^u, X^v in terms of the dot products $X \cdot U, X \cdot V$. To that end we notice

$$\begin{aligned} \begin{bmatrix} U \cdot X \\ V \cdot X \end{bmatrix} &= \begin{bmatrix} U & V \end{bmatrix}^t X \\ &= \begin{bmatrix} U & V \end{bmatrix}^t \begin{bmatrix} U & V \end{bmatrix} \begin{bmatrix} X^u \\ X^v \end{bmatrix} \end{aligned}$$

Showing directly that

$$\begin{bmatrix} X^u \\ X^v \end{bmatrix} = \left(\begin{bmatrix} U & V \end{bmatrix}^t \begin{bmatrix} U & V \end{bmatrix} \right)^{-1} \begin{bmatrix} U \cdot X \\ V \cdot X \end{bmatrix}$$

and consequently

$$X = \begin{bmatrix} U & V \end{bmatrix} \left(\begin{bmatrix} U & V \end{bmatrix}^t \begin{bmatrix} U & V \end{bmatrix} \right)^{-1} \begin{bmatrix} U \cdot X \\ V \cdot X \end{bmatrix}$$

□

REMARK A.1.2. There is a similar formula in $\mathbb{R}^3$ which is a bit longer. In practice we shall only need it in the case where the third basis vector is perpendicular to the first two. Also note that if U, V are orthonormal then

$$\begin{bmatrix} U & V \end{bmatrix}^t \begin{bmatrix} U & V \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

and we recover the standard formula for the expansion of a vector in an orthonormal basis.

THEOREM A.1.3. *A real symmetric matrix, or symmetric linear operator on a finite dimensional Euclidean space, has an orthonormal basis of eigenvectors.*

PROOF. First observe that if we have two eigenvectors

$$Av = \lambda v, \; Aw = \mu w$$

where $\lambda \neq \mu$, then

$$\begin{aligned} (\lambda - \mu) \left(v^t w \right) &= (\lambda v)^t w - v^t (\mu w) \\ &= (Av)^t w - v^t (Aw) \\ &= v^t A^t w - v^t A w \\ &= v^t A w - v^t A w \\ &= 0 \end{aligned}$$

so it must follow that $v \perp w$.

This shows that the eigenspaces are all perpendicular to each other. Thus we are reduced to showing that such matrices only have real eigenvalues. There are many fascinating proofs of this. We give a fairly down to earth proof in the cases that are relevant to us.

For a 2×2 matrix

$$A = \begin{bmatrix} a & b \\ b & d \end{bmatrix}$$

the characteristic polynomial is

$$\lambda^2 - (a+d)\,\lambda + ad - b^2$$

so the discriminant is

$$\Delta = (a+d)^2 - 4\left(ad - b^2\right) = (a-d)^2 + 4b^2 \geq 0$$

This shows that the roots must be real.

For a 3×3 matrix the characteristic polynomial is cubic. The intermediate value theorem then guarantees at least one real root. If we make a change of basis to another orthonormal basis where the first basis vector is an eigenvector then the new matrix will still be symmetric and look like

$$\begin{bmatrix} \lambda_1 & 0 & 0 \\ 0 & a & b \\ 0 & b & d \end{bmatrix}$$

The characteristic polynomial then looks like

$$(\lambda - \lambda_1)\left(\lambda^2 - (a+d)\,\lambda + ad - b^2\right)$$

where we see as before that $\lambda^2 - (a+d)\,\lambda + ad - b^2$ has two real roots. $\square$

A.2. Geometry

Here are a few geometric formulas that use vector notation:

- The length or size of a vector X is denoted:
$$|X| = \sqrt{X^t \cdot X}$$
- The distance from X to a point P:
$$|X - P|$$
- The projection of a vector X onto another vector N:
$$\frac{X \cdot N}{|N|^2} N$$
- The signed distance from P to a plane that goes through X_0 and has normal N, i.e., given by $(X - X_0) \cdot N = 0$:
$$\frac{(P - X_0) \cdot N}{|N|}$$
the actual distance is the absolute value of the signed distance. This formula also works for the (signed) distance from a point to a line in $\mathbb{R}^2$.
- The distance from P to a line with direction N that passes through X_0:
$$\left|(P - X_0) - \frac{(P - X_0) \cdot N}{|N|^2} N\right| = \sqrt{|P - X_0|^2 - \frac{|(P - X_0) \cdot N|^2}{|N|^2}}$$
- The area of a parallelogram spanned by two vectors X, Y is
$$\sqrt{\det\left(\begin{bmatrix} X & Y \end{bmatrix}^t \begin{bmatrix} X & Y \end{bmatrix}\right)}$$
- If $X, Y \in \mathbb{R}^2$ there is also a signed area given by
$$\det \begin{bmatrix} X & Y \end{bmatrix}$$

- If $X, Y \in \mathbb{R}^3$ the area is also given by
$$|X \times Y|$$
- The volume of a parallelepiped spanned by X, Y, Z is
$$\sqrt{\det\left(\begin{bmatrix} X & Y & Z \end{bmatrix}^t \begin{bmatrix} X & Y & Z \end{bmatrix}\right)}$$
- If $X, Y, Z \in \mathbb{R}^3$ the signed volume is given by
$$\begin{aligned} \det \begin{bmatrix} X & Y & Z \end{bmatrix} &= X \cdot (Y \times Z) \\ &= X^t (Y \times Z) \end{aligned}$$
- The

A.3. Geometry of Space-Time

We collect a few of the special features of space-time $\mathbb{R}^{2,1}$ where we use the inner product
$$X \cdot Y = X^x Y^x + X^y Y^y - X^z Y^z.$$

A.4. Differentiation and Integration

A.4.1. Directional Derivatives. If h is a function on $\mathbb{R}^3$ and $X = (P, Q, R)$ then
$$\begin{aligned} D_X h &= P\frac{\partial h}{\partial x} + Q\frac{\partial h}{\partial y} + R\frac{\partial h}{\partial z} \\ &= (\nabla h) \cdot X \\ &= [\nabla h]^t [X] \\ &= \begin{bmatrix} \frac{\partial h}{\partial x} & \frac{\partial h}{\partial y} & \frac{\partial h}{\partial z} \end{bmatrix} [X] \end{aligned}$$
and for a vector field V we get
$$D_X V = \begin{bmatrix} \frac{\partial V}{\partial x} & \frac{\partial V}{\partial y} & \frac{\partial V}{\partial z} \end{bmatrix} [X].$$
We can also calculate directional derivatives by selecting a curve such that $\dot{c}(0) = X$. Along the curve the chain rule says:
$$\frac{d(V \circ c)}{dt} = \begin{bmatrix} \frac{\partial V}{\partial x} & \frac{\partial V}{\partial y} & \frac{\partial V}{\partial z} \end{bmatrix} \left[\frac{dc}{dt}\right] = D_{\dot{c}} V$$
Thus
$$D_X V = \frac{d(V \circ c)}{dt}(0)$$

A.4.2. Chain Rules. Consider a vector function $V : \mathbb{R}^3 \to \mathbb{R}^n$ and a curve $c : I \to \mathbb{R}^3$. That the curves goes in to space and the vector function is defined on the same space is important, but that it has dimension 3 is not. Note also that the vector function can have values in a higher or lower dimensional space.

The chain rule for calculating the derivative of the composition $V \circ c$ is:
$$\frac{d(V \circ c)}{dt} = \begin{bmatrix} \frac{\partial V}{\partial x} & \frac{\partial V}{\partial y} & \frac{\partial V}{\partial z} \end{bmatrix} \left[\frac{dc}{dt}\right]$$

There is a very convenient short cut for writing such chain rules if we keep in mind that they simply involve matrix notation. Write

$$X = \begin{bmatrix} x \\ y \\ z \end{bmatrix}$$

and

$$c(t) = X(t)$$

Then this chain rule can be written as

$$\frac{d(V \circ c)}{dt} = \frac{\partial V}{\partial X}\frac{dX}{dt}$$

were we think of

$$\frac{\partial V}{\partial X} = \frac{\partial V}{\partial (x, y, z)} = \begin{bmatrix} \frac{\partial V}{\partial x} & \frac{\partial V}{\partial y} & \frac{\partial V}{\partial z} \end{bmatrix}$$

and

$$\frac{dX}{dt} = \frac{d}{dt}\begin{bmatrix} x \\ y \\ z \end{bmatrix}$$

It is also sometimes convenient to have X be a function of several variables, say, $X(u, v)$. In that case we obtain

$$\frac{\partial V(X(u, v))}{\partial u} = \frac{\partial V}{\partial X}\frac{\partial X}{\partial u}$$

as partial derivatives are simple regular derivatives in one variable when all other variables are fixed.

A.4.3. Local Invertibility. Mention Inverse and Implicit Function Theorems. Lagrange multipliers.

A.4.4. Integration. Change of variables. Green's, divergence, and Stokes' thms. Use Green Thm to prove the change of variable formula, and similarly with Stokes.

A.5. Differential Equations

The basic existence and uniqueness theorem for systems of first order equations is contained in the following statement. The first part is standard and can be found in most text books. The second part about the assertion of smoothness in relation to the initial value is very important, but is somewhat trickier to establish.

Theorem A.5.1. *Given a smooth function $F : \mathbb{R} \times \mathbb{R}^n \to \mathbb{R}^n$ the initial value problem*

$$\frac{d}{dt}x = F(t, x), \; x(0) = x_0$$

has a solution

$$x(t) = \begin{bmatrix} x^1(t) \\ \vdots \\ x^n(t) \end{bmatrix}$$

that is unique on some possibly small interval $(-\epsilon, \epsilon)$. When $|x_0| \leq R$, we can pick ϵ independently of x_0. Moreover this solution is smooth in both t and the initial *value* x_0*. In case* $|F(t,x)| \leq M + C|x|$ *for constants* $M, C \geq 1$ *we can choose* $\epsilon = \infty$*.*

Proof. The proof is quite long and consists of several different proof. The existence and uniqueness is relatively standard. The long term existence is less standard so we supply a proof below. The smoothness on initial values is also standard but not covered in many texts (see , however, MM for a good proof).

Long term existence:.................................... □

The above result was strictly about ODEs (ordinary differential equations), but it can be used to prove certain results about PDEs (partial differential equations) as well.

We consider a system

$$\begin{aligned} \frac{\partial}{\partial u}x &= P(u,v,x) \\ \frac{\partial}{\partial v}x &= Q(u,v,x) \\ x(0,0) &= x_0 \end{aligned}$$

where $x(u,v)$ is now a function of two variables with values in $\mathbb{R}^n$.

The standard situation from multivariable calculus is:

Theorem A.5.2. (Clairaut's Theorem) *When* $P = P(u,v)$ *and* $Q = Q(u,v)$ *do not depend on* x *a solution to*

$$\begin{aligned} \frac{\partial}{\partial u}x &= P(u,v) \\ \frac{\partial}{\partial v}x &= Q(u,v) \\ x(0,0) &= x_0 \end{aligned}$$

can be found if and only if the system is exact*, i.e.,*

$$\frac{\partial}{\partial u}Q = \frac{\partial}{\partial v}P.$$

This solution will be defined on all of $\mathbb{R}^2$ *provided* $P, Q : \mathbb{R}^2 \to \mathbb{R}$*.*

Proof. If such a solution exists, then it follows that

$$\frac{\partial}{\partial u}Q = \frac{\partial^2 x}{\partial u \partial v} = \frac{\partial^2 x}{\partial v \partial u} = \frac{\partial}{\partial v}P.$$

Conversely start by defining $x_1(u)$ as

$$x_1(u) = x_0 + \int_0^u P(s,0)\,ds.$$

Next define the function $x(u,v)$ for a fixed u by

$$x(u,v) = x_1(u) + \int_0^v Q(u,t)\,dt.$$

This gives us

$$\frac{\partial x}{\partial v} = Q,\ x(0,0) = x_0.$$

Thus it remains to check that

$$\frac{\partial x}{\partial u} = P.$$

Note however that when $v = 0$ we have

$$\frac{\partial x}{\partial u}(u,0) = \frac{dx_1}{du}(u) = P(u,0).$$

More generally the v-derivatives satisfy

$$\begin{aligned}
\frac{\partial^2 x}{\partial v \partial u} &= \frac{\partial^2 x}{\partial u \partial v} \\
&= \frac{\partial Q}{\partial u} \\
&= \frac{\partial P}{\partial v}.
\end{aligned}$$

So it follows that

$$\frac{\partial}{\partial v}\left(\frac{\partial x}{\partial u} - P\right) = 0.$$

For fixed u this shows that

$$v \mapsto \frac{\partial x}{\partial u} - P$$

is constant. Since $\left(\frac{\partial x}{\partial u} - P\right)(u,0) = 0$ this implies that $\frac{\partial x}{\partial u} = P$. □

This result can be extended to the more general situation as follows. When computing the derivative of $P(u,v,x(u,v))$ with respect to v it is clearly necessary to use the chain rule

$$\frac{\partial}{\partial v}(P(u,v,x(u,v))) = \frac{\partial P}{\partial v} + \frac{\partial P}{\partial x}\frac{\partial x}{\partial v} = \frac{\partial P}{\partial v} + \frac{\partial P}{\partial x}Q$$

where $\frac{\partial P}{\partial v}$ is the partial derivative of P keeping v and x fixed. Similarly

$$\frac{\partial}{\partial u}(Q(u,v,x)) = \frac{\partial Q}{\partial u} + \frac{\partial Q}{\partial x}\frac{\partial x}{\partial u} = \frac{\partial Q}{\partial u} + \frac{\partial Q}{\partial x}P$$

so if a solution exists the functions P and Q must satisfy the condition

$$\frac{\partial P}{\partial v} + \frac{\partial P}{\partial x}Q = \frac{\partial Q}{\partial u} + \frac{\partial Q}{\partial x}P.$$

This is called the *integrability condition* for the system. Conversely we have

THEOREM A.5.3. *Assume $P, Q : \mathbb{R}^2 \times \mathbb{R}^n \to \mathbb{R}^n$ are two smooth functions that satisfy the integrability condition*

$$\frac{\partial P}{\partial v} + \frac{\partial P}{\partial x}Q = \frac{\partial Q}{\partial u} + \frac{\partial Q}{\partial x}P.$$

The solution

$$x(u,v) = \begin{bmatrix} x^1(u,v) \\ \vdots \\ x^n(u,v) \end{bmatrix}$$

for the initial value problem

$$\begin{aligned}
\frac{\partial}{\partial u}x &= P(u,v,x), \\
\frac{\partial}{\partial v}x &= Q(u,v,x), \\
x(0,0) &= x_0
\end{aligned}$$

exists and is unique on some possibly small domain $(-\epsilon,\epsilon)^2$. *When* $|P|,|Q| \leq M + C|x|$ *for constants* $M, C \geq 1$ *the solution exists on all of* $\mathbb{R}^2$.

Proof. We invoke theorem A.5.3 to define x_1 as the unique solution to

$$\frac{d}{du}x_1(u) = P(u,0,x_1(u)),\ x_1(0) = x_0.$$

Next use theorem A.5.3 to define the function $x(u,v)$ for a fixed u as the solution to

$$\frac{d}{dv}x(u,v) = Q(u,v,x(u,v)),\ x(u,0) = x_1(u)$$

as well as to check that $x(u,v)$ is smooth in both variables. This gives us

$$\frac{\partial x}{\partial v} = Q,\ x(0,0) = x_0.$$

Thus it remains to check that

$$\frac{\partial x}{\partial u} = P.$$

Note however that when $v = 0$ we have

$$\frac{\partial x}{\partial u}(u,0) = \frac{dx_1}{du}(u) = P(u,0,x(u,0)).$$

More generally the v-derivatives satisfy

$$\begin{aligned}
\frac{\partial^2 x}{\partial v \partial u} &= \frac{\partial^2 x}{\partial u \partial v} \\
&= \frac{\partial}{\partial u}(Q(u,v,x)) \\
&= \frac{\partial Q}{\partial u} + \frac{\partial Q}{\partial x}\frac{\partial x}{\partial u} \\
&= \frac{\partial P}{\partial v} + \frac{\partial P}{\partial x}Q - \frac{\partial Q}{\partial x}P + \frac{\partial Q}{\partial x}\frac{\partial x}{\partial u}
\end{aligned}$$

and

$$\frac{\partial}{\partial v}P(u,v,x) = \frac{\partial P}{\partial v} + \frac{\partial P}{\partial x}\frac{\partial x}{\partial v} = \frac{\partial P}{\partial v} + \frac{\partial P}{\partial x}Q.$$

So it follows that

$$\frac{\partial}{\partial v}\left(\frac{\partial x}{\partial u} - P\right) = \frac{\partial Q}{\partial x}\left(\frac{\partial x}{\partial u} - P\right).$$

For fixed u this is a differential equation in $\frac{\partial x}{\partial u} - P$. Now $\left(\frac{\partial x}{\partial u} - P\right)(u,0) = 0$ and the zero function clearly solves this equation so it follows that

$$\frac{\partial x}{\partial u} - P = 0$$

for all v. As u was arbitrary this shows the claim.

In case P, Q are bounded we can invoke theorem A.5.1 to see that x is also defined for all $(u,v) \in \mathbb{R}^2$. □

Remark A.5.4. It is not difficult to expand this result to systems of m equations if x has m variables.

The most important case for us is when $x = X$ is a row matrix of vector functions

$$X = \begin{bmatrix} U_1 & \cdots & U_m \end{bmatrix},$$

where $U_i : \Omega \to V$ are defined on some domain $\Omega \subset \mathbb{R}^n$ and the vector space V is m-dimensional. We will generally assume that for each $p \in \Omega$ the vectors $U_1(p), ..., U_m(p)$ form a basis for V. This implies that the derivatives of these vector functions are linear combinations of this basis. Thus we obtain a system

$$\frac{\partial}{\partial u^i} \begin{bmatrix} U_1 & \cdots & U_m \end{bmatrix} = \begin{bmatrix} U_1 & \cdots & U_m \end{bmatrix} [D_i],$$

where $[D_i]$ is an $m \times m$ matrix whose columns represent the coefficients of the vectors on the left hand side

$$\frac{\partial U_j}{\partial u^i} = d_{ij}^1 U_1 + \cdots + d_{ij}^m U_m = \begin{bmatrix} U_1 & \cdots & U_m \end{bmatrix} \begin{bmatrix} d_{ij}^1 \\ \vdots \\ d_{ij}^m \end{bmatrix}.$$

In this way each of the entries are functions on the domain $d_{ij}^k : \Omega \to \mathbb{R}$.

The necessary integrability conditions now become

$$\frac{\partial^2}{\partial u^i \partial u^j} \begin{bmatrix} U_1 & \cdots & U_m \end{bmatrix} = \frac{\partial^2}{\partial u^j \partial u^i} \begin{bmatrix} U_1 & \cdots & U_m \end{bmatrix}.$$

As

$$\begin{aligned}
\frac{\partial^2}{\partial u^i \partial u^j} \begin{bmatrix} U_1 & \cdots & U_m \end{bmatrix} &= \frac{\partial}{\partial u^i} \left(\frac{\partial}{\partial u^j} \begin{bmatrix} U_1 & \cdots & U_m \end{bmatrix} \right) \\
&= \frac{\partial}{\partial u^i} \left(\begin{bmatrix} U_1 & \cdots & U_m \end{bmatrix} [D_j] \right) \\
&= \left(\frac{\partial}{\partial u^i} \begin{bmatrix} U_1 & \cdots & U_m \end{bmatrix} \right) [D_j] + \begin{bmatrix} U_1 & \cdots & U_m \end{bmatrix} \frac{\partial}{\partial u^i} [D_j] \\
&= \begin{bmatrix} U_1 & \cdots & U_m \end{bmatrix} [D_i] [D_j] + \begin{bmatrix} U_1 & \cdots & U_m \end{bmatrix} \frac{\partial}{\partial u^i} [D_j] \\
&= \begin{bmatrix} U_1 & \cdots & U_m \end{bmatrix} \left([D_i] [D_j] + \frac{\partial}{\partial u^i} [D_j] \right)
\end{aligned}$$

and $U_1, ..., U_m$ form a basis the integrability conditions become

$$[D_i] [D_j] + \frac{\partial}{\partial u^i} [D_j] = [D_j] [D_i] + \frac{\partial}{\partial u^j} [D_i].$$

Depending on the individual context it might be possible to calculate $[D_i]$ without first finding the partial derivatives

$$\frac{\partial U_k}{\partial u^i}$$

but we can't expect this to always happen. Note, however, that if V comes with an inner product, then the product rule implies that

$$\frac{\partial (U_k \cdot U_l)}{\partial u^i} = \frac{\partial U_k}{\partial u^i} \cdot U_l + U_k \cdot \frac{\partial U_l}{\partial u^i}.$$

This means in matrix form that

$$\begin{aligned}
&\frac{\partial}{\partial u^i}\left(\begin{bmatrix} U_1 & \cdots & U_m \end{bmatrix}^t \begin{bmatrix} U_1 & \cdots & U_m \end{bmatrix}\right) \\
= \quad & \left(\frac{\partial}{\partial u^i}\begin{bmatrix} U_1 & \cdots & U_m \end{bmatrix}^t\right)\begin{bmatrix} U_1 & \cdots & U_m \end{bmatrix} + \begin{bmatrix} U_1 & \cdots & U_m \end{bmatrix}^t \frac{\partial}{\partial u^i}\begin{bmatrix} U_1 & \cdots & U_m \end{bmatrix} \\
= \quad & [D_i]^t \begin{bmatrix} U_1 & \cdots & U_m \end{bmatrix}^t \begin{bmatrix} U_1 & \cdots & U_m \end{bmatrix} + \begin{bmatrix} U_1 & \cdots & U_m \end{bmatrix}^t \begin{bmatrix} U_1 & \cdots & U_m \end{bmatrix} [D_i] .
\end{aligned}$$

Or more condensed

$$\frac{\partial}{\partial u^i}\left(X^t X\right) = [D_i]^t X^t X + X^t X [D_i] .$$

If we additionally assume that $d_{ij}^k = d_{ji}^k$, then we obtain the surprising formula:

$$d_{ij}^k = g^{kl}\left(\frac{\partial g_{li}}{\partial u^j} + \frac{\partial g_{lj}}{\partial u^i} - \frac{\partial g_{ij}}{\partial u^l}\right).$$

APPENDIX B

Special Coordinate Representations

The purpose of this appendix is to collect properties and formulas that are specific to the type of parametrization that is being used. These are used in several places in the text and also appear as exercises.

B.1. Cartesian and Oblique Coordinates

Cartesian coordinates on a surface is a parametrization where

$$[\mathrm{I}] = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

Oblique coordinates more generally come from a parametrization where

$$[\mathrm{I}] = \begin{bmatrix} a & b \\ b & d \end{bmatrix}$$

for constants a, b, d with $a, d > 0$ and $ad - b^2 > 0$.

Note that the Christoffel symbols all vanish if we have a parametrization where the metric coefficients are constant. In particular, the rather nasty formula we developed in the proof of Theorema Egregium shows that the Gauss curvature vanishes. This immediately tells us that Cartesian or oblique coordinates cannot exist if the Gauss curvature doesn't vanish. When we have defined geodesic coordinates below we'll also be able to show that even abstract surfaces with zero Gauss curvature admit Cartesian coordinates.

B.2. Surfaces of Revolution

Many features of surfaces show themselves for surfaces of revolution. While this is certainly a special class of surfaces it is broad enough to give a rich family examples.

We consider

$$\mathbf{q}(t, \mu) = (r(t)\cos\mu, r(t)\sin\mu, z(t)).$$

It is often convenient to select or reparametrized (r, z) so that it is a unit speed curve. In this case we use the parametrization

$$\begin{aligned} \mathbf{q}(s, \mu) &= (r(s)\cos\mu, r(s)\sin\mu, h(s)), \\ (r')^2 + (h')^2 &= 1 \end{aligned}$$

We get the unit sphere by using $r = \sin s$ and $h = \cos s$.

We get a cone, cylinder or plane, by considering $r = (\alpha t + \beta)$ and $h = \gamma t$. When $\gamma = 0$ these are simply polar coordinates in the $\mathbf{q}$, y plane. When $\alpha = 0$ we get a cylinder, while if both α and γ are nontrivial we get a cone. When $\alpha^2 + \gamma^2 = 1$ we have a parametrization by arclength.

The basis is given by

$$\begin{aligned}
\frac{\partial \mathbf{q}}{\partial t} &= \left(\dot{r}\cos\mu, \dot{r}\sin\mu, \dot{h}\right), \\
\frac{\partial \mathbf{q}}{\partial \mu} &= \left(-r\sin\mu, r\cos\mu, 0\right), \\
\mathbf{N} &= \frac{\left(-\dot{h}\cos\mu, -\dot{h}\sin\mu, \dot{r}\right)}{\sqrt{\dot{h}^2+\dot{r}^2}}
\end{aligned}$$

and first fundamental form by

$$\begin{aligned}
g_{tt} &= \dot{h}^2+\dot{r}^2, \\
g_{\mu\mu} &= r^2 \\
g_{t\mu} &= 0
\end{aligned}$$

Note that the cylinder has the same first fundamental form as the plane if we use Cartesian coordinates in the plane. The cone also allows for Cartesian coordinates, but they are less easy to construct directly. This is not so surprising as we just saw that it took different types of coordinates for the cylinder and the plane to recognize that they admitted Cartesian coordinates. Pictorially one can put Cartesian coordinates on the cone by slicing it open along a meridian and then unfolding it to be flat. Think of unfolding a lamp shade or the Cartesian grid on a waffle cone.

Taking a surface of revolution using the arclength parameter s, we see that

$$\begin{aligned}
\frac{\partial \mathbf{N}}{\partial s} &= \frac{\partial}{\partial s}\left(-h'\cos\mu, -h'\sin\mu, r'\right) \\
&= \left(-h''\cos\mu, -h''\sin\mu, r''\right) \\
\frac{\partial \mathbf{N}}{\partial \mu} &= \frac{\partial}{\partial \mu}\left(-h'\cos\mu, -h'\sin\mu, r'\right) \\
&= \left(h'\sin\mu, -h'\cos\mu, 0\right)
\end{aligned}$$

The Weingarten map is now found by expanding these two vectors. For the last equation this is simply

$$\begin{aligned}
\frac{\partial \mathbf{N}}{\partial \mu} &= \left(h'\sin\mu, -h'\cos\mu, 0\right) \\
&= -\frac{h'}{r}\left(-r\sin\mu, r\cos\mu, 0\right) \\
&= -\frac{h'}{r}\frac{\partial \mathbf{q}}{\partial \mu}
\end{aligned}$$

Thus we have

$$\begin{aligned}
L_\mu^s &= L_s^\mu = 0, \\
L_\mu^\mu &= \frac{h'}{r}
\end{aligned}$$

This leaves us with finding L_s^s. Since $\frac{\partial \mathbf{q}}{\partial s}$ is a unit vector this is simply

$$\begin{aligned} L_s^s &= -\frac{\partial \mathbf{N}}{\partial s} \cdot \frac{\partial \mathbf{q}}{\partial s} \\ &= \left(h'' \cos \mu, h'' \sin \mu, -r''\right) \cdot \left(r' \cos \mu, r' \sin \mu, h'\right) \\ &= h'' r' - r'' h' \end{aligned}$$

Thus

$$\begin{aligned} K &= \left(h'' r' - r'' h'\right) \frac{h'}{r} \\ H &= \frac{h'}{r} + h'' r' - r'' h' \end{aligned}$$

In the case of cylinder, plane, and cone we note that K vanishes, but H only vanishes when it is a plane. This means that we have a selection of surfaces all with Cartesian coordinates with different H.

We can in general simplify the Gauss curvature by noting that

$$\begin{aligned} 1 &= \left(r'\right)^2 + \left(h'\right)^2 \\ 0 &= \left(\left(r'\right)^2 + \left(h'\right)^2\right)' = 2r' r'' + 2h' h'' \end{aligned}$$

Thus yielding

$$\begin{aligned} K &= \left(r'' \frac{\left(r'\right)^2}{h'} - r'' h'\right) \frac{h'}{r} \\ &= \frac{r''}{r}\left(-\left(r'\right)^2 - \left(h'\right)^2\right) \\ &= -\frac{r''}{r} \\ &= -\frac{\frac{\partial^2}{\partial s^2}\left(\sqrt{g_{rr}}\right)}{\sqrt{g_{rr}}} \end{aligned}$$

This makes it particularly easy to calculate the Gauss curvature and also to construct examples with a given curvature function. It also shows that the Gauss curvature can be computed directly from the first fundamental form! For instance if we want $K = -1$, then we can just use $r(s) = \exp(-s)$ for $s > 0$ and then adjust $h(s)$ for $s \in (0, \infty)$ such that

$$1 = \left(r'\right)^2 + \left(h'\right)^2$$

If we introduce a new parameter $t = \exp(s) > 1$, then we obtain a new parametrization of the same surface

$$\begin{aligned} \mathbf{q}(t, \mu) &= \mathbf{q}(\ln(t), \mu) \\ &= (\exp(-\ln t) \cos \mu, \exp(-\ln t) \sin \mu, h(\ln t)) \\ &= \left(\frac{1}{t} \cos \mu, \frac{1}{t} \sin \mu, h(\ln t)\right) \end{aligned}$$

To find the first fundamental form of this surface we have to calculate

$$\begin{aligned}
\frac{d}{dt}h(\ln t) &= \frac{dh}{ds}\frac{1}{t} \\
&= \sqrt{1-(r')^2}\frac{1}{t} \\
&= \sqrt{1-(-\exp(-s))^2}\frac{1}{t} \\
&= \sqrt{1-\exp(-2\ln t)}\frac{1}{t} \\
&= \sqrt{1-\frac{1}{t^2}}\frac{1}{t}
\end{aligned}$$

Thus

$$\mathrm{I} = \begin{bmatrix} \frac{1}{t^4}+\left(1-\frac{1}{t^2}\right)\frac{1}{t^2} & 0 \\ 0 & \frac{1}{t^2} \end{bmatrix} = \begin{bmatrix} \frac{1}{t^2} & 0 \\ 0 & \frac{1}{t^2} \end{bmatrix}$$

This is exactly what the first fundamental form for the upper half plane looks like. But the domains for the two are quite different. What we have achieved is a local representation of part of the upper half plane.

Exercises.

(1) Show that geodesics on a surface of revolution satisfy Clairaut's condition: $r\sin\phi$ is constant, where ϕ is the angle the geodesic forms with the meridians.

B.3. Monge Patches

This is more complicated than the previous case, but that is only to be expected as all surfaces admit Monge patches. We consider $\mathbf{q}(u,v) = (u,v,f(u,v))$. Thus

$$\begin{aligned}
\frac{\partial \mathbf{q}}{\partial u} &= \left(1,0,\frac{\partial f}{\partial u}\right), \\
\frac{\partial \mathbf{q}}{\partial v} &= \left(0,1,\frac{\partial f}{\partial v}\right)
\end{aligned}$$

$$\mathbf{N} = -\frac{\left(\frac{\partial f}{\partial u},\frac{\partial f}{\partial v},-1\right)}{\sqrt{1+\left(\frac{\partial f}{\partial u}\right)^2+\left(\frac{\partial f}{\partial v}\right)^2}}$$

$$
\begin{aligned}
g_{uu} &= 1+\left(\frac{\partial f}{\partial u}\right)^2, \\
g_{vv} &= 1+\left(\frac{\partial f}{\partial v}\right)^2, \\
g_{uv} &= \frac{\partial f}{\partial u}\frac{\partial f}{\partial v}, \\
[\mathrm{I}] &= \left[\begin{array}{cc} 1+\left(\frac{\partial f}{\partial u}\right)^2 & \frac{\partial f}{\partial u}\frac{\partial f}{\partial v} \\ \frac{\partial f}{\partial u}\frac{\partial f}{\partial v} & 1+\left(\frac{\partial f}{\partial v}\right)^2 \end{array}\right] \\
\det[\mathrm{I}] &= 1+\left(\frac{\partial f}{\partial u}\right)^2+\left(\frac{\partial f}{\partial v}\right)^2
\end{aligned}
$$

$$\frac{\partial^2 \mathbf{q}}{\partial w_1 \partial w_2} = \left(0, 0, \frac{\partial^2 f}{\partial w_1 \partial w_2}\right)$$

So we immediately get

$$\Gamma_{w_1 w_2 w_3} = \frac{\partial^2 f}{\partial w_1 \partial w_2}\frac{\partial f}{\partial w_3}$$

$$L_{w_1 w_2} = \frac{\frac{\partial^2 f}{\partial w_1 \partial w_2}}{\sqrt{1+\left(\frac{\partial f}{\partial u}\right)^2+\left(\frac{\partial f}{\partial v}\right)^2}}$$

The Gauss curvature is then the determinant of

$$L = \left[\begin{array}{cc} L_u^u & L_v^u \\ L_u^v & L_v^v \end{array}\right] = \left[\begin{array}{cc} g^{uu} & g^{uv} \\ g^{vu} & g^{vv} \end{array}\right]\left[\begin{array}{cc} L_{uu} & L_{uv} \\ L_{vu} & L_{vv} \end{array}\right]$$

$$
\begin{aligned}
K &= \frac{1}{\det[\mathrm{I}]}\det\left[\begin{array}{cc} L_{uu} & L_{uv} \\ L_{vu} & L_{vv} \end{array}\right] \\
&= \frac{\frac{\partial^2 f}{\partial u^2}\frac{\partial^2 f}{\partial v^2}-\left(\frac{\partial^2 f}{\partial u \partial v}\right)^2}{\det[\mathrm{I}]^2}
\end{aligned}
$$

We note that

$$
\begin{aligned}
[\mathrm{I}]^{-1} &= \frac{1}{\det[\mathrm{I}]}\left[\begin{array}{cc} 1+\left(\frac{\partial f}{\partial v}\right)^2 & -\frac{\partial f}{\partial u}\frac{\partial f}{\partial v} \\ -\frac{\partial f}{\partial u}\frac{\partial f}{\partial v} & 1+\left(\frac{\partial f}{\partial u}\right)^2 \end{array}\right], \\
[\mathrm{II}] &= \frac{1}{\sqrt{\det[\mathrm{I}]}}\left[\begin{array}{cc} \frac{\partial^2 f}{\partial u^2} & \frac{\partial^2 f}{\partial u \partial v} \\ \frac{\partial^2 f}{\partial u \partial v} & \frac{\partial^2 f}{\partial v^2} \end{array}\right]
\end{aligned}
$$

and the Weingarten map

$$\begin{aligned}
[L] &= [\mathrm{I}]^{-1}[\mathrm{II}] \\
&= \frac{1}{(\det[\mathrm{I}])^{\frac{3}{2}}}\begin{bmatrix} 1+\left(\frac{\partial f}{\partial v}\right)^2 & -\frac{\partial f}{\partial u}\frac{\partial f}{\partial v} \\ -\frac{\partial f}{\partial u}\frac{\partial f}{\partial v} & 1+\left(\frac{\partial f}{\partial u}\right)^2 \end{bmatrix}\begin{bmatrix} \frac{\partial^2 f}{\partial u^2} & \frac{\partial^2 f}{\partial u \partial v} \\ \frac{\partial^2 f}{\partial u \partial v} & \frac{\partial^2 f}{\partial v^2} \end{bmatrix} \\
&= \frac{1}{(\det[\mathrm{I}])^{\frac{3}{2}}}\begin{bmatrix} \left(1+\left(\frac{\partial f}{\partial v}\right)^2\right)\frac{\partial^2 f}{\partial u^2} - \frac{\partial f}{\partial u}\frac{\partial f}{\partial v}\frac{\partial^2 f}{\partial u \partial v} & \left(1+\left(\frac{\partial f}{\partial v}\right)^2\right)\frac{\partial^2 f}{\partial u \partial v} - \frac{\partial f}{\partial u}\frac{\partial f}{\partial v}\frac{\partial^2 f}{\partial v^2} \\ \left(1+\left(\frac{\partial f}{\partial u}\right)^2\right)\frac{\partial^2 f}{\partial u \partial v} - \frac{\partial f}{\partial u}\frac{\partial f}{\partial v}\frac{\partial^2 f}{\partial u^2} & \left(1+\left(\frac{\partial f}{\partial u}\right)^2\right)\frac{\partial^2 f}{\partial v^2} - \frac{\partial f}{\partial u}\frac{\partial f}{\partial v}\frac{\partial^2 f}{\partial u \partial v} \end{bmatrix}
\end{aligned}$$

This gives us a general example where the Weingarten map might not be a symmetric matrix.

B.4. Surfaces Given by an Equation

This is again very general. Note that any Monge patch $(u, v, f(u, v))$ also yields a function $F(x, y, z) = z - f(x, y)$ such that the zero level of F is precisely the Monge patch. This case is also complicated by the fact that while the normal is easy to find, it is proportional to the gradient of F, we don't have a basis for the tangent space without resorting to a Monge patch. This is troublesome, but not insurmountable as we can solve for the derivatives of F. Assume that near some point p we know $\frac{\partial F}{\partial z} \neq 0$, then we can use x, y as coordinates. Our coordinates vector fields look like

$$\begin{aligned}
\frac{\partial \mathbf{q}}{\partial u} &= \left(1, 0, \frac{\partial f}{\partial u}\right), \\
\frac{\partial \mathbf{q}}{\partial v} &= \left(0, 1, \frac{\partial f}{\partial v}\right)
\end{aligned}$$

where

$$\frac{\partial f}{\partial w} = -\frac{\frac{\partial F}{\partial w}}{\frac{\partial F}{\partial z}}$$

Thus we actually get some explicit formulas

$$\begin{aligned}
\frac{\partial \mathbf{q}}{\partial u} &= \left(1, 0, -\frac{\frac{\partial F}{\partial u}}{\frac{\partial F}{\partial z}}\right), \\
\frac{\partial \mathbf{q}}{\partial v} &= \left(0, 1, -\frac{\frac{\partial F}{\partial v}}{\frac{\partial F}{\partial z}}\right).
\end{aligned}$$

We can however describe the second fundamental form without resorting to coordinates. We consider a surface given by an equation

$$F(x, y, z) = C.$$

The normal can be calculated directly as

$$\mathbf{N} = \frac{\nabla F}{|\nabla F|}.$$

This shows first of all that we have a simple equation defining the tangent space at each point p

$$T_p M = \left\{ Y \in \mathbb{R}^3 \mid Y \cdot \nabla F(p) = 0 \right\}.$$

Next we make the claim that

$$\begin{aligned}\mathrm{II}(X,Y) &= -\frac{1}{|\nabla F|}\mathrm{I}(D_X\nabla F, Y)\\ &= -\frac{1}{|\nabla F|}Y\cdot D_X\nabla F,\end{aligned}$$

where D_X is the directional derivative. We can only evaluate II on tangent vectors, but $Y\cdot D_X\nabla F$ clearly makes sense for all vectors. This has the advantage that we can even use Cartesian coordinates in $\mathbb{R}^3$ for our tangent vectors. First we show that

$$L(X) = -D_X\mathbf{N}.$$

Select a parametrization $\mathbf{q}(u,v)$ such that

$$\frac{\frac{\partial \mathbf{q}}{\partial u}\times\frac{\partial \mathbf{q}}{\partial v}}{\left|\frac{\partial \mathbf{q}}{\partial u}\times\frac{\partial \mathbf{q}}{\partial v}\right|} = \frac{\nabla F}{|\nabla F|}.$$

The Weingarten equations then tell us that

$$L\left(\frac{\partial \mathbf{q}}{\partial w}\right) = -\frac{\partial \mathbf{N}}{\partial w} = -D_{\frac{\partial \mathbf{q}}{\partial w}}\mathbf{N}.$$

We can now return to the second fundamental form. Let Y be another tangent vector then, $Y\cdot\nabla F = 0$ so

$$\begin{aligned}-\mathrm{II}(X,Y) &= -\mathrm{I}(L(X), Y)\\ &= Y\cdot D_X\mathbf{N}\\ &= Y\cdot\left(D_X\frac{1}{|\nabla F|}\right)\nabla F + Y\cdot\frac{1}{|\nabla F|}D_X\nabla F\\ &= Y\cdot\frac{1}{|\nabla F|}D_X\nabla F.\end{aligned}$$

Note that even when X is tangent it does not necessarily follow that $D_X\nabla F$ is also tangent to the surface.

In case $\frac{\partial F}{\partial z}\neq 0$ we get a relatively simple orthogonal basis for the tangent space. In case $\frac{\partial F}{\partial x} = \frac{\partial F}{\partial y} = 0$ we can simply use

$$X = (1,0,0),\ Y = (0,1,0)$$

otherwise we obtain an orthogonal basis by using

$$\begin{aligned}X &= \left(-\frac{\partial F}{\partial y}, \frac{\partial F}{\partial x}, 0\right),\\ Y &= \left(\frac{\partial F}{\partial z}\frac{\partial F}{\partial x}, \frac{\partial F}{\partial z}\frac{\partial F}{\partial y}, -\left(\left(\frac{\partial F}{\partial x}\right)^2 + \left(\frac{\partial F}{\partial y}\right)^2\right)\right).\end{aligned}$$

With that basis the Weingarten map can then be calculated as

$$\begin{aligned}[L] &= [\mathrm{I}]^{-1}[\mathrm{II}]\\ &= \begin{bmatrix}|X|^{-2} & 0\\ 0 & |Y|^{-2}\end{bmatrix}\begin{bmatrix}\mathrm{II}(X,X) & \mathrm{II}(X,Y)\\ \mathrm{II}(X,Y) & \mathrm{II}(Y,Y)\end{bmatrix}.\end{aligned}$$

To calculate the second fundamental form we use that

$$\begin{bmatrix} \frac{\partial \nabla F}{\partial x} & \frac{\partial \nabla F}{\partial y} & \frac{\partial \nabla F}{\partial z} \end{bmatrix} = \begin{bmatrix} \frac{\partial^2 F}{\partial x^2} & \frac{\partial^2 F}{\partial x \partial y} & \frac{\partial^2 F}{\partial x \partial z} \\ \frac{\partial^2 F}{\partial y \partial x} & \frac{\partial^2 F}{\partial y^2} & \frac{\partial^2 F}{\partial y \partial z} \\ \frac{\partial^2 F}{\partial z \partial x} & \frac{\partial^2 F}{\partial z \partial y} & \frac{\partial^2 F}{\partial z^2} \end{bmatrix}.$$

So

$$\mathrm{II}(X,X) = \frac{1}{|\nabla F|} \begin{bmatrix} -\frac{\partial F}{\partial y} & \frac{\partial F}{\partial x} & 0 \end{bmatrix} \begin{bmatrix} \frac{\partial^2 F}{\partial x^2} & \frac{\partial^2 F}{\partial x \partial y} & \frac{\partial^2 F}{\partial x \partial z} \\ \frac{\partial^2 F}{\partial y \partial x} & \frac{\partial^2 F}{\partial y^2} & \frac{\partial^2 F}{\partial y \partial z} \\ \frac{\partial^2 F}{\partial z \partial x} & \frac{\partial^2 F}{\partial z \partial y} & \frac{\partial^2 F}{\partial z^2} \end{bmatrix} \begin{bmatrix} -\frac{\partial F}{\partial y} \\ \frac{\partial F}{\partial x} \\ 0 \end{bmatrix},$$

$$\mathrm{II}(X,Y) = \frac{1}{|\nabla F|} \begin{bmatrix} -\frac{\partial F}{\partial y} & \frac{\partial F}{\partial x} & 0 \end{bmatrix} \begin{bmatrix} \frac{\partial^2 F}{\partial x^2} & \frac{\partial^2 F}{\partial x \partial y} & \frac{\partial^2 F}{\partial x \partial z} \\ \frac{\partial^2 F}{\partial y \partial x} & \frac{\partial^2 F}{\partial y^2} & \frac{\partial^2 F}{\partial y \partial z} \\ \frac{\partial^2 F}{\partial z \partial x} & \frac{\partial^2 F}{\partial z \partial y} & \frac{\partial^2 F}{\partial z^2} \end{bmatrix} \begin{bmatrix} \frac{\partial F}{\partial z}\frac{\partial F}{\partial x} \\ \frac{\partial F}{\partial z}\frac{\partial F}{\partial y} \\ -\left(\frac{\partial F}{\partial x}\right)^2 - \left(\frac{\partial F}{\partial y}\right)^2 \end{bmatrix},$$

$$\mathrm{II}(Y,Y) = \frac{1}{|\nabla F|} \begin{bmatrix} \frac{\partial F}{\partial z}\frac{\partial F}{\partial x} & \frac{\partial F}{\partial z}\frac{\partial F}{\partial y} & -\left(\frac{\partial F}{\partial x}\right)^2 - \left(\frac{\partial F}{\partial y}\right)^2 \end{bmatrix} \begin{bmatrix} \frac{\partial^2 F}{\partial x^2} & \frac{\partial^2 F}{\partial x \partial y} & \frac{\partial^2 F}{\partial x \partial z} \\ \frac{\partial^2 F}{\partial y \partial x} & \frac{\partial^2 F}{\partial y^2} & \frac{\partial^2 F}{\partial y \partial z} \\ \frac{\partial^2 F}{\partial z \partial x} & \frac{\partial^2 F}{\partial z \partial y} & \frac{\partial^2 F}{\partial z^2} \end{bmatrix} \begin{bmatrix} \frac{\partial F}{\partial z}\frac{\partial F}{\partial x} \\ \frac{\partial F}{\partial z}\frac{\partial F}{\partial y} \\ -\left(\frac{\partial F}{\partial x}\right)^2 - \left(\frac{\partial F}{\partial y}\right)^2 \end{bmatrix}.$$

Exercises.

(1) If $\mathbf{q}$ is a curve, then it is a curve on $F = C$ if $\mathbf{q}(0)$ lies on the surface and $\dot{\mathbf{q}} \cdot \nabla F$ vanishes. If $\mathbf{q}$ is regular and a curve on $F = C$, then it can be reparametrized to be a geodesic if and only if the triple product $\det[\nabla F, \dot{\mathbf{q}}, \ddot{\mathbf{q}}] = 0$.

B.5. Geodesic Coordinates

This is a parametrization having a first fundamental form that looks like:

$$\mathrm{I} = \begin{bmatrix} 1 & 0 \\ 0 & g_{vv} \end{bmatrix}$$

This is as with surfaces of revolution, but now g_{vv} can depend on both u and v. Using a central v curve, we let the u curves be unit speed geodesics orthogonal to the fixed v curve. They are also often call Fermi coordinates after the famous physicist and seem to have been used in his thesis on general relativity. They were however also used by Gauss. These coordinates will be used time and again to simplify calculations in the proofs of several theorems. The v-curves are well defined as the curves that appear when u is constant. At $u = 0$ the u and v curves are perpendicular by construction, so by continuity they can't be tangent as long as u is sufficiently small.

Exercises.

(1) Consider a parametrization $\mathbf{q}(s,t)$ where the s-curves are unit speed geodesics and $\frac{\partial \mathbf{q}}{\partial s}(s,0) \perp \frac{\partial \mathbf{q}}{\partial t}(s,0)$. Show that

$$\frac{\partial \mathbf{q}}{\partial s}(s,t) \perp \frac{\partial \mathbf{q}}{\partial t}(s,t)$$

and conclude that such a parametrization defines geodesic coordinates.

(2) Show that for geodesic coordinates:

$$
\begin{aligned}
\Gamma_{uuu} &= 0, \\
\Gamma_{uvu} &= 0 = \Gamma_{vuu}, \\
\Gamma_{vvu} &= -\frac{1}{2}\frac{\partial g_{vv}}{\partial u}, \\
\Gamma_{vvv} &= \frac{1}{2}\frac{\partial g_{vv}}{\partial v}, \\
\Gamma_{uvv} &= \frac{1}{2}\frac{\partial g_{vv}}{\partial u} = \Gamma_{vuv}, \\
\Gamma_{uuv} &= 0, \\
\Gamma_{ij}^{u} &= \Gamma_{iju}, \\
\Gamma_{ij}^{v} &= \frac{1}{g_{vv}}\Gamma_{ijv},
\end{aligned}
$$

and

$$K = -\frac{\partial_u^2 \sqrt{g_{vv}}}{\sqrt{g_{vv}}} = -\frac{1}{2}\left(\frac{\partial_u^2 g_{vv}}{g_{vv}} - \left(\frac{\partial_u g_{vv}}{g_{vv}}\right)^2\right).$$

B.6. Chebyshev Nets

These correspond to a parametrization where the first fundamental form looks like:

$$
\begin{aligned}
\mathrm{I} &= \begin{bmatrix} 1 & c \\ c & 1 \end{bmatrix} \\
&= \begin{bmatrix} 1 & \cos\theta \\ \cos\theta & 1 \end{bmatrix},
\end{aligned}
$$

Real life interpretations that are generally brought up are fishnet stockings or non-stretchable cloth tailored to the contours of the body. The idea is to have a material where the fibers are not changed in length or stretched, but are allowed to change their mutual angles.

Note that such parametrizations are characterized as having unit speed parameter curves.

Exercises.

(1) Show that any surface locally admits Chebyshev nets. Hint: Fix a point $p = \mathbf{q}(u_0, v_0)$ for a given parametrization and define new parameters

$$
\begin{aligned}
s(u,v) &= \int_{u_0}^{u} \sqrt{g_{uu}(x,v)}dx \\
t(u,v) &= \int_{v_0}^{v} \sqrt{g_{vv}(u,y)}dy
\end{aligned}
$$

Show that $\frac{\partial s}{\partial v}(u_0, v_0) = 0 = \frac{\partial t}{\partial u}(u_0, v_0)$ and conclude that (s,t) defines a new parametrization that creates a Chebyshev net.

(2) Show that Chebyshev nets $\mathbf{q}(u,v)$ satisfy the following properties

$$\frac{\partial^2 \mathbf{q}}{\partial u \partial v} \perp T_pM,$$

$$\begin{aligned}
\Gamma_{uvw} &= \Gamma_{uuu} = \Gamma_{vvv} = 0, \\
\Gamma_{uuv} &= -\frac{\partial \theta}{\partial u} \sin \theta, \\
\Gamma_{vvu} &= -\frac{\partial \theta}{\partial v} \sin \theta, \\
\frac{\partial^2 \theta}{\partial u \partial v} &= -K \sin \theta.
\end{aligned}$$

(3) Show that the geodesic curvature κ_g of the u-coordinate curves in a Chebyshev net satisfy

$$\kappa_g = -\frac{\partial \theta}{\partial u}.$$

(4) (Hazzidakis) Show that $\sqrt{\det [\mathrm{I}]} = \sin \theta$, and that integrating the Gauss curvature over a coordinate rectangle yields:

$$-\int_{[a,b]\times[c,d]} K \sin \theta du dv = 2\pi - \alpha_1 - \alpha_2 - \alpha_3 - \alpha_4$$

where the angles α_i are the interior angles.

B.7. Isothermal Coordinates

These are also more generally known as *conformally flat coordinates* and have a first fundamental form that looks like:

$$\mathrm{I} = \begin{bmatrix} \lambda^2 & 0 \\ 0 & \lambda^2 \end{bmatrix}$$

The proof that these always exist is called the *local uniformization theorem.* It is not a simple result, but the importance of these types of coordinates in the development of both classical and modern surface theory cannot be understated. There is also a global result which we will mention at a later point. Gauss was the first to work with such coordinates, and Riemann also heavily depended on their use. They have the properties that

$$\begin{aligned}
\Gamma_{uuu} &= \frac{\partial \ln \lambda}{\partial u} \\
\Gamma_{uvu} &= \frac{\partial \ln \lambda}{\partial v} = \Gamma_{vuu} \\
\Gamma_{vvv} &= \frac{\partial \ln \lambda}{\partial v} \\
\Gamma_{uvv} &= \frac{\partial \ln \lambda}{\partial u} = \Gamma_{vuv} \\
\Gamma_{uuv} &= -\frac{\partial \ln \lambda}{\partial v} \\
\Gamma_{vvu} &= -\frac{\partial \ln \lambda}{\partial u}, \\
\Gamma_{w_1 w_2}^{w_3} &= \frac{1}{\lambda^2} \Gamma_{w_1 w_2 w_3},
\end{aligned}$$

$$K = -\frac{1}{\lambda^2} \left(\frac{\partial^2 \ln \lambda}{\partial u^2} + \frac{\partial^2 \ln \lambda}{\partial v^2} \right).$$

Exercises.

(1) A particularly nice special case occurs when

$$\lambda^2(u,v) = U^2(u) + V^2(v)$$

These types of metrics are called *Liouville metrics.* Compute their Christoffel symbols, Gauss curvature, and show that when geodesics are written as $v(u)$ or $u(v)$ they they solve a separable differential equation. Show also that the geodesics have the property that

$$U^2 \sin^2 \omega - V^2 \cos^2 \omega$$

is constant, where ω is the angle the geodesic forms with the u curves.

(2) Show that when

$$\lambda = \frac{1}{a(u^2+v^2) + b_u u + b_v v + c}$$

we obtain a metric with constant Gauss curvature

$$K = 4ac - b_u^2 - b_v^2$$

It can be shown that no other choices for λ will yield constant curvature.

Bibliography

[1] D. Hilbert and S. Cohn-Vossen, *Geometry and the Imagination*, New York: Chelsea, 1990.
[2] D.J. Struik, *Lectures on Classical Differential Geometry*, New York: Dover, 1988.