CLASSICAL DIFFERENTIAL GEOMETRY
Curves and Surfaces in Euclidean Space

Gert Heckman
Radboud University Nijmegen
G.Heckman@math.ru.nl

March 21, 2018

Contents

Preface

These are lectures on classicial differential geometry of curves and surfaces in Euclidean space $\mathbb{R}^3$, as it developped in the 18^{th} and 19^{th} century. Their principal investigators were Gaspard Monge (1746-1818), Carl Friedrich Gauss (1777-1855) and Bernhard Riemann (1826-1866).

In Chapter 1 we discuss smooth curves in the plane $\mathbb{R}^2$ and in space $\mathbb{R}^3$. The main results are the definition of curvature and torsion, the Frenet equations for the derivative of the moving frame, and the fundamental theorem for smooth curves being essentially charcterized by their curvature and torsion. The theory of smooth curves is also a preparation for the study of smooth surfaces in $\mathbb{R}^3$ via smooth curves on them.

The results of Chapter 2 on the first and second fundamental forms are essentially due to Monge and his contemporaries. At the end of this chapter we can give the definitions of mean curvature and Gaussian curvature.

In Chapter 3 we will discuss particular examples: surfaces of revolution, and a special attention for various charts on the sphere. We end this chapter with a discussion of ruled surfaces.

Chapter 4 on the Theorema Egregium deals with the main contributions by Gauss, as developped in his "Disquisitiones generalis circa superficies curvas" (General investigations on curved surfaces) from 1827. This chapter is a highlight of these lectures, and altogether we shall discuss four different proofs of the Theorema Egregium.

In Chapter 5 we discuss geodesics on a surface $\mathcal{S}$ in $\mathbb{R}^3$. They are defined as smooth curves on $\mathcal{S}$ whose acceleration vector in $\mathbb{R}^3$ is perpendicular to the surface along the curve. They turn out to be the curves that locally minimize the length between points.

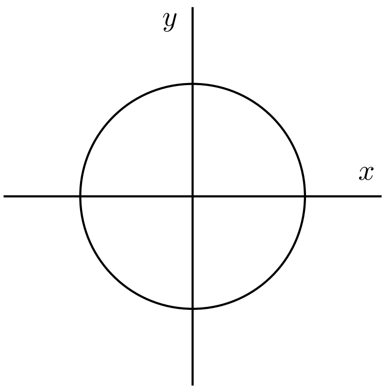
The last Chapter 6 deals with abstract surfaces in the spirit of Riemann, notably from his famous Habilitationsvortrag "Ueber die Hypothesen welche der Geometrie zu Grunde liegen" (On the hypotheses which lie at the basis of geometry) from 1854. The basic example of such an abstract Riemannian surface is the hyperbolic plane with its constant curvature equal to -1 Riemannian metric. We discuss the Riemann disc model and the Poincaré upper half plane model for hyperbolic geometry.

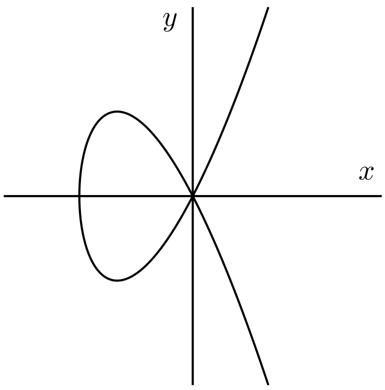
This course can be taken by bachelor students with a good knowledge of calculus, and some knowledge of differential equations. After taking this course they should be well prepared for a follow up course on modern Riemannian geometry. The text book "Elementary Differential Geometry" by Andrew Pressley from 2010 contains additional details and many exercises as well, and will be used for this course.

1 Smooth Curves

1.1 Plane Curves

A plane algebraic curve is given as the locus of points (x, y) in the plane $\mathbb{R}^2$ which satisfy a polynomial equation $F(x, y) = 0$. For example the unit circle with equation $F(x, y) = x^2 + y^2 - 1$ and the nodal cubic curve with equation $F(x, y) = y^2 - x^2(x + 1)$ are represented by the pictures





These curves can be parametrized. For the circle we can take the familiar parametrization

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

where the time t runs over $\mathbb{R}/2\pi\mathbb{Z}$. But parametrizations of a curve are highly nonunique. For example, the pencil of lines $y = t(x + 1)$ through the point $(-1, 0)$ on the circle intersects the circle in a unique other point, which gives the parametrization

$$t \mapsto \mathbf{r}(t) = (\frac{1 - t^2}{1 + t^2}, \frac{2t}{1 + t^2})$$

where time t runs over the real line $\mathbb{R}$ together with a point ∞ at infinity correspondig to the base point $(-1, 0)$ of the pencil. The first parametrization is the standard unit speed trigonometic parametrization, while the second parametrization is a rational parametrization of the circle.

The second method can also be used to find a parametrization of the nodal cubic curve. Consider the pencil of lines $y = tx$ with base point $(0, 0)$

the nodal singular point. Each line of the pencil intersects the nodal cubic curve in a unique other point, and we find the polynomial parametrization

$$t \mapsto \mathbf{r}(t) = (t^2 - 1, t^3 - t)$$

where time t runs over the real line $\mathbb{R}$.

The circle and the nodal cubic curve are so called rational curves, because they admit a rational parametization. However, it can be shown that the cubic curve with equation $F(x, y) = 4x^3 - ax - b - y^2$ is not a rational curve, as soon as the discriminant $\Delta = a^3 - 27b^2$ does not vanish.

The above parametrizations give in fact holomorphic parametrization of the complex points of the curves in question. The cubic curve has a holomorphic parametization of its complex points using the Weierstrass function $\wp(a, b; t)$. This holomorphic function of the complex variable t is doubly periodic, and as such is called an elliptic function. Likewise the trigonometric parametrization of the unit circle is simply periodic in the complex variable t with periods from $2\pi\mathbb{Z}$.

This leads us into the world of complex function theory and algebraic geometry. Although a highly interesting part of mathematics it is not the subject of these lectures. Instead we shall study real curves (and later real surfaces) given by smooth real equations through smooth real parametrizations. Here smooth means infinitely differentiable in the sense of calculus (or real analysis). From this perspective the implicit function theorem is a relevant general result.

Theorem 1.1. *Let $F : U \to \mathbb{R}$ be a smooth function on an open subset U in the plane $\mathbb{R}^2$. Let F_x and F_y denote the partial derivatives of F with respect to x and y respectively. If $F(x_0, y_0) = 0$ and $F_y(x_0, y_0) \neq 0$ for some point (x_0, y_0) in U then locally near (x_0, y_0) the curve with equation $F(x, y) = 0$ is the graph $y = f(x)$ of a smooth function f of the variable x near the point x_0 with $f(x_0) = y_0$.*

The curve with an implicit equation $F(x, y) = 0$ is locally an explicit graph $y = f(x)$, which explains the name implicit function theorem. However, our starting point for smooth curves will not be through their equations, but right from the definition through their parametrizations.

Definition 1.2. *A smooth plane curve $\mathcal{C}$ is a smooth injective map*

$$(\alpha, \beta) \ni t \mapsto \mathbf{r}(t) = (x(t), y(t)) \in \mathbb{R}^2$$

with $-\infty \leq \alpha < \beta \leq \infty$ and nowhere vanishing derivative $\mathbf{r}_t = (x_t, y_t)$.

For example, a curve given by an equation $F(x, y) = 0$ as in the implicit function theorem can be parametrized by

$$t \mapsto \mathbf{r}(t) = (t, f(t))$$

or more generally by

$$t \mapsto \mathbf{r}(t) = (x(t), f(x(t)))$$

with $x_t > 0$ or $x_t < 0$ throughout the interval domain of the parametrization.

Thinking in this way of a plane curve as a smooth path $t \mapsto \mathbf{r}(t)$ traced out in time t the first and second derivative

$$\mathbf{v} = \mathbf{r}_t \quad , \quad \mathbf{a} = \mathbf{r}_{tt}$$

are the velocity and the acceleration of the motion along the curve. What is their geometric meaning?

If we denote by $\mathbf{u} \cdot \mathbf{v}$ the inner product of the vectors $\mathbf{u}, \mathbf{v} \in \mathbb{R}^2$ then the first fundamental function

$$t \mapsto g = \mathbf{r}_t \cdot \mathbf{r}_t$$

is just the square of the speed at which the curve is traversed. The function

$$t \mapsto s(t) = \int_{t_0}^{t} \sqrt{g(u)}du$$

gives the length of the curve traced out between time t_0 and a later time t. The vector $\mathbf{t} = \mathbf{r}_t/\sqrt{g}$ is the unit tangent vector of the curve.

Let J denote the counterclockwise rotation of $\mathbb{R}^2$ over an angle $\pi/2$, so that $J\mathbf{e}_1 = \mathbf{e}_2$ and $J\mathbf{e}_2 = -\mathbf{e}_1$ with $\mathbf{e}_1 = (1, 0)$ and $\mathbf{e}_2 = (0, 1)$ the standard orthonormal basis of $\mathbb{R}^2$. The vector $\mathbf{n} = J\mathbf{t}$ is called the unit normal vector of the curve.

The second fundamental function

$$h = \mathbf{r}_{tt} \cdot \mathbf{n}$$

is the component of the acceleration in the normal direction.

Let $t \mapsto \tilde{t}$ from (α, β) to $(\tilde{\alpha}, \tilde{\beta})$ be a bijective and bismooth map. The smooth plane curve $\tilde{\mathcal{C}}$

$$\tilde{t} \mapsto \tilde{\mathbf{r}}(\tilde{t}) = \mathbf{r}(t(\tilde{t}))$$

is called a reparametrization of the original curve $\mathcal{C}$. The reparametrization is called proper (improper) if

$$\frac{d\tilde{t}}{dt} > 0 \quad \left(\frac{d\tilde{t}}{dt} < 0\right)$$

meaning that the curve is traced out in the same (opposite) direction.

Suppose $t \mapsto \tilde{t}$ is a reparametrization. Then we have

$$\begin{aligned} \frac{d\mathbf{r}}{dt} &= \frac{d\tilde{\mathbf{r}}}{d\tilde{t}}\frac{d\tilde{t}}{dt} \\ \frac{d^2\mathbf{r}}{dt^2} &= \frac{d^2\tilde{\mathbf{r}}}{d\tilde{t}^2}\left(\frac{d\tilde{t}}{dt}\right)^2 + \frac{d\tilde{\mathbf{r}}}{d\tilde{t}}\frac{d^2\tilde{t}}{dt^2} \end{aligned}$$

using the chain rule. Since the unit tangent vector and the unit normal vector remain unchanged under proper reparametizations we find

$$g(t) = \tilde{g}(\tilde{t})\left(\frac{d\tilde{t}}{dt}\right)^2 \ , \ h(t) = \tilde{h}(\tilde{t})\left(\frac{d\tilde{t}}{dt}\right)^2$$

as transformation rules for the first and second fundamental functions. So the quotient $h/g = \tilde{h}/\tilde{g}$ is invariant under proper reparametrizations of the curve. Since both $\mathbf{t}$ and $\mathbf{n}$ change sign under improper reparametrizations the quotient h/g changes sign as well under improper reparametrizations.

Definition 1.3. *The function $k = h/g$ is called the signed curvature of the plane curve $\mathcal{C}$.*

For example, the circle with radius $r > 0$ and parametrization

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

has unit tangent $\mathbf{t}$ and unit normal $\mathbf{n}$ given by

$$\mathbf{t} = (-\sin t, \cos t) \ , \ \mathbf{n} = J\mathbf{t} = (-\cos t, -\sin t)$$

and the fundamental functions become $g = r^2$ and $h = r$. Hence the signed curvature $k = h/g = 1/r$ is just the inverse of the radius of the circle.

If $k(t) = 0$ for some t then the corresponding point $\mathbf{r}(t)$ of the curve $\mathcal{C}$ is called a flex point. Equivalently, for a flexpoint the acceleration $\mathbf{r}_{tt}(t)$ is a scalar multiple of the velocity $\mathbf{r}_t(t)$.

The next result is called the fundamental theorem for plane curves.

Theorem 1.4. *If we have given two arbitrary smooth real valued functions $t \mapsto g(t) > 0$ and $t \mapsto h(t)$ defined on some open interval then there exists locally a smooth curve $t \mapsto \mathbf{r}(t)$ with the prescribed $g > 0$ and h as their first and second fundamental functions. The curve is unique up to a proper Euclidean motion.*

Proof. Since $\{\mathbf{t}, \mathbf{n}\}$ is an orthonormal basis we have the equation

$$\mathbf{r}_{tt} = (\mathbf{r}_{tt} \cdot \mathbf{t})\mathbf{t} + (\mathbf{r}_{tt} \cdot \mathbf{n})\mathbf{n}$$

and using $\mathbf{t} = \mathbf{r}_t/\sqrt{g}, \mathbf{n} = J\mathbf{r}_t/\sqrt{g}$ and $g = \mathbf{r}_t \cdot \mathbf{r}_t, g_t = 2\mathbf{r}_{tt} \cdot \mathbf{r}_t$ we arrive at the second order ordinary differential equation

$$\mathbf{r}_{tt} = \left(\frac{g_t}{2g} + \frac{h}{\sqrt{g}} J \right) \mathbf{r}_t$$

for the curve $t \mapsto \mathbf{r}(t)$ on the given interval. This second order differential equation has a unique solution $t \mapsto \mathbf{r}(t)$ on a sufficiently small open interval around an initial time t_0 with the initial position $\mathbf{r}(t_0)$ and initial velocity $\mathbf{r}_t(t_0)$ freely prescribed.

Taking the inner product of the second order differential equation with $\mathbf{r}_t$ we find the equation

$$(\mathbf{r}_t \cdot \mathbf{r}_t)_t / 2(\mathbf{r}_t \cdot \mathbf{r}_t) = g_t/2g$$

or equivalently

$$(\log(\mathbf{r}_t \cdot \mathbf{r}_t) - \log g)_t = 0$$

with general solution $(\mathbf{r}_t \cdot \mathbf{r}_t) = cg$ for some constant $c > 0$. Taking the solution curve $t \mapsto \mathbf{r}(t)$ with $\mathbf{r}_t(t_0) \cdot \mathbf{r}_t(t_0) = g(t_0)$ we get $c = 1$ and therefore $g = (\mathbf{r}_t \cdot \mathbf{r}_t)$ for all t on the given interval around t_0. Taking the inner product of the second order differential equation with the unit normal $\mathbf{n} = J\mathbf{r}_t/\sqrt{g}$ we also find

$$\mathbf{r}_{tt} \cdot \mathbf{n} = h(t)$$

and the solution curve $t \mapsto \mathbf{r}(t)$ of the second order differential equation has g and h as its first and second fundamental functions.

The curve $t \mapsto \mathbf{r}(t)$ is unique up to a translation in order to fix the initial position $\mathbf{r}(t_0)$ composed with a rotation to fix the direction of the initial velocity $\mathbf{r}_t(t_0)$. The composition of a translation with a rotation is called a proper Euclidean motion of $\mathbb{R}^2$. These transformations form a group under composition of maps, the proper Euclidean motion group of the Euclidean plane. □

An important reparametrization of a curve $t \mapsto \mathbf{r}(t)$ is the arclength $s = s(t)$ characterized by $ds/dt = \sqrt{g}$ with $g = \mathbf{r}_t \cdot \mathbf{r}_t$ the first fundamental function. If the curve $s \mapsto \mathbf{r}(s)$ is parametrized by arclength then $\mathbf{r}_s \cdot \mathbf{r}_s \equiv 1$. The second order differential equation in the above proof simplifies to

$$\mathbf{r}_{ss} = kJ\mathbf{r}_s$$

and the geometry of the locus of points traced out by the curve is essentially determined by the signed curvature $s \mapsto k(s)$ up to a proper Euclidean motion of the plane $\mathbb{R}^2$.

1.2 Space Curves

The definition of a space curve is just the obvious modification of the one of a plane curve as given in Definition 1.2.

Definition 1.5. *A smooth space curve $\mathcal{C}$ is a smooth injective map*

$$(\alpha, \beta) \ni t \mapsto \mathbf{r}(t) = (x(t), y(t), z(t)) \in \mathbb{R}^3$$

for some $-\infty \leq \alpha < \beta \leq \infty$ such that the derivative $\mathbf{r}_t = (x_t, y_t, z_t) \neq 0$ is nowhere vanishing.

Likewise a bijective and bismooth map $(\alpha, \beta) \ni t \mapsto \tilde{t} \in (\tilde{\alpha}, \tilde{\beta})$ gives a reparametrization $\tilde{\mathcal{C}}$ of $\mathcal{C}$ by means of

$$(\tilde{\alpha}, \tilde{\beta}) \ni \tilde{t} \mapsto \tilde{\mathbf{r}}(\tilde{t}) = \mathbf{r}(t(\tilde{t})) \in \mathbb{R}^3$$

and we speak of a proper or improper reparametrization depending on whether $d\tilde{t}/dt$ is positive or negative respectively. We obtain

$$\begin{aligned}
\frac{d\mathbf{r}}{dt} &= \frac{d\tilde{\mathbf{r}}}{d\tilde{t}}\frac{d\tilde{t}}{dt} \\
\frac{d^2\mathbf{r}}{dt^2} &= \frac{d^2\tilde{\mathbf{r}}}{d\tilde{t}^2}\left(\frac{d\tilde{t}}{dt}\right)^2 + \frac{d\tilde{\mathbf{r}}}{d\tilde{t}}\frac{d^2\tilde{t}}{dt^2} \\
\frac{d^3\mathbf{r}}{dt^3} &= \frac{d^3\tilde{\mathbf{r}}}{d\tilde{t}^3}\left(\frac{d\tilde{t}}{dt}\right)^3 + 3\frac{d^2\tilde{\mathbf{r}}}{d\tilde{t}^2}\frac{d^2\tilde{t}}{dt^2}\frac{d\tilde{t}}{dt} + \frac{d\tilde{\mathbf{r}}}{d\tilde{t}}\frac{d^3\tilde{t}}{dt^3}
\end{aligned}$$

by the chain rule. Using these formulas one can check that (denoting for the derivative with respect to t a dot and for the derivative with respect to $\tilde{t}$ a prime)

$$\dot{\mathbf{r}} \times \ddot{\mathbf{r}} = \left(\frac{d\tilde{t}}{dt}\right)^3 \tilde{\mathbf{r}}' \times \tilde{\mathbf{r}}'' \ , \ \dddot{\mathbf{r}} \cdot (\dot{\mathbf{r}} \times \ddot{\mathbf{r}}) = \left(\frac{d\tilde{t}}{dt}\right)^6 \tilde{\mathbf{r}}''' \cdot (\tilde{\mathbf{r}}' \times \tilde{\mathbf{r}}'')$$

which in turn implies that the scalar expressions

$$\kappa = \frac{|\dot{\mathbf{r}} \times \ddot{\mathbf{r}}|}{|\dot{\mathbf{r}}|^3} \ , \ \tau = \frac{\dddot{\mathbf{r}} \cdot (\dot{\mathbf{r}} \times \ddot{\mathbf{r}})}{|\dot{\mathbf{r}} \times \ddot{\mathbf{r}}|^2}$$

are independent of the parametrization, and should have geometric meaning.

Definition 1.6. *The functions κ and τ are called the curvature and the torsion of the curve $\mathcal{C}$ respectively.*

Note that the torsion is only well defined if the curvature $\kappa > 0$ is positive. As soon as we deal with the torsion τ this will always be tacitly assumed. The name torsion is due to Valleé in his text book Traité de Geometrie Descriptive of 1825.

The length of the curve (taken from $t = t_0$ to $t = t_1$) is given by

$$\int_{t_0}^{t_1} |\dot{\mathbf{r}}| dt$$

and viewed as a new parameter $s = s(t) = \int |\dot{\mathbf{r}}| dt$ is called the arclength.

If we assume that the curve $\mathcal{C}$ is parametrized by arclength then the tangent vector $\mathbf{t} = \dot{\mathbf{r}}$ has unit length. Differentiation of the relation $\mathbf{t} \cdot \mathbf{t} = 1$ yields $\dot{\mathbf{t}} \cdot \mathbf{t} = 0$. The unit vector $\mathbf{n}$, defined by $\dot{\mathbf{t}} = \kappa \mathbf{n}$ for $\kappa > 0$, is therefore perpendicular to $\mathbf{t}$ and called the principal normal. The vector product $\mathbf{b} = \mathbf{t} \times \mathbf{n}$ is called the binormal. The above formulas for curvature and torsion in case of arclength parametrization take the form

$$\kappa = \dot{\mathbf{t}} \cdot \mathbf{n} \ , \ \tau = \dot{\mathbf{n}} \cdot \mathbf{b} \ .$$

Indeed, we have $\ddot{\mathbf{r}} = \dot{\mathbf{t}} = \kappa \mathbf{n}$ and $\tau = \ddot{\mathbf{t}} \cdot (\mathbf{t} \times \dot{\mathbf{t}})/\kappa^2 = \dot{\mathbf{n}} \cdot \mathbf{b}$.

Definition 1.7. *Let the space curve $s \mapsto \mathbf{r}(s)$ be parametrized by arclength. The orthonormal triple $\{\mathbf{t}, \mathbf{n}, \mathbf{b}\}$ of tangent vector $\mathbf{t}$, principal normal $\mathbf{n}$ and binormal $\mathbf{b}$ is called the moving frame, or in french the "répère mobile" of Élie Cartan.*

The name binormal was introduced by Barré de Saint Venant in 1845.

Theorem 1.8. *Let the space curve $s \mapsto \mathbf{r}(s)$ be parametrized by arclength. Suppose that the curvature $\kappa(s) > 0$ for all s, and so the moving frame $\mathbf{t}(s), \mathbf{n}(s), \mathbf{b}(s)$ and the torsion $\tau(s)$ are defined for all s. Then the moving frame of the curve satisfies the system of differential equations*

$$\begin{aligned} \dot{\mathbf{t}} &= 0\mathbf{t} + \kappa \mathbf{n} + 0\mathbf{b} \\ \dot{\mathbf{n}} &= -\kappa \mathbf{t} + 0\mathbf{n} + \tau \mathbf{b} \\ \dot{\mathbf{b}} &= 0\mathbf{t} - \tau \mathbf{n} + 0\mathbf{b} \end{aligned}$$

which are called the Frenet (or sometimes Frenet–Serret) equations.

The Frenet equations were found by Frenet in his dissertation of 1847, and were found independently by Serret and published in the Journal de Mathématique 16 of 1851.

Proof. Because $\{\mathbf{t}, \mathbf{n}, \mathbf{b}\}$ is an orthonormal basis we have

$$\begin{aligned} \dot{\mathbf{t}} &= (\dot{\mathbf{t}} \cdot \mathbf{t})\mathbf{t} + (\dot{\mathbf{t}} \cdot \mathbf{n})\mathbf{n} + (\dot{\mathbf{t}} \cdot \mathbf{b})\mathbf{b} \\ \dot{\mathbf{n}} &= (\dot{\mathbf{n}} \cdot \mathbf{t})\mathbf{t} + (\dot{\mathbf{n}} \cdot \mathbf{n})\mathbf{n} + (\dot{\mathbf{n}} \cdot \mathbf{b})\mathbf{b} \\ \dot{\mathbf{b}} &= (\dot{\mathbf{b}} \cdot \mathbf{t})\mathbf{t} + (\dot{\mathbf{b}} \cdot \mathbf{n})\mathbf{n} + (\dot{\mathbf{b}} \cdot \mathbf{b})\mathbf{b} \end{aligned}$$

Differentiation of $\mathbf{t} \cdot \mathbf{t} = \mathbf{n} \cdot \mathbf{n} = \mathbf{b} \cdot \mathbf{b} = 1$ yields $\dot{\mathbf{t}} \cdot \mathbf{t} = \dot{\mathbf{n}} \cdot \mathbf{n} = \dot{\mathbf{b}} \cdot \mathbf{b} = 0$. By the definition of curvature and torsion we have

$$\dot{\mathbf{t}} \cdot \mathbf{n} = \kappa \ , \ \dot{\mathbf{t}} \cdot \mathbf{b} = 0 \ , \ \dot{\mathbf{n}} \cdot \mathbf{b} = \tau$$

as explained above. Likewise, differentiation of $\mathbf{t} \cdot \mathbf{n} = \mathbf{t} \cdot \mathbf{b} = \mathbf{n} \cdot \mathbf{b} = 0$ yields $\dot{\mathbf{t}} \cdot \mathbf{n} = -\dot{\mathbf{n}} \cdot \mathbf{t}, \dot{\mathbf{t}} \cdot \mathbf{b} = -\dot{\mathbf{b}} \cdot \mathbf{t}, \dot{\mathbf{n}} \cdot \mathbf{b} = -\dot{\mathbf{b}} \cdot \mathbf{n}$. This proves the Frenet formulas. □

The next result is called the fundamental theorem for space curves.

Theorem 1.9. *If $s \mapsto \kappa(s) > 0$ and $s \mapsto \tau(s)$ are smooth functions on an open interval then there exists locally a space curve $s \mapsto \mathbf{r}(s)$ parametrized by arclength and with curvature $\kappa(s)$ and torsion $\tau(s)$. Moreover the curve is unique up to a proper Euclidean motion.*

Proof. Let $\{\mathbf{r}_1(t), \cdots, \mathbf{r}_n(t)\}$ be a set of n vectors in $\mathbb{R}^n$ depending in a smooth way on a parameter t in some interval (α, β) such that at some initial time t_0 the vectors form a positive orthonomal basis of $\mathbb{R}^n$. Then for all time t in (α, β) these vectors from a positive orthonormal basis of $\mathbb{R}^n$ if an only if the functions $a_{ij}(t)$ defined by the equations

$$\dot{\mathbf{r}}_i(t) = \sum_k a_{ik}(t)\mathbf{r}_k(t)$$

satisfy the skew symmetry relation

$$a_{ij}(t) + a_{ji}(t) = 0$$

for all $t \in (\alpha, \beta)$. Indeed observe that

$$\frac{d}{dt}(\mathbf{r}_i \cdot \mathbf{r}_j) = \sum_k (a_{ik}\mathbf{r}_k \cdot \mathbf{r}_j + a_{jk}\mathbf{r}_i \cdot \mathbf{r}_k)$$

by the Leibniz product rule. The left hand side is identically equal to 0 if and only if $\mathbf{r}_i \cdot \mathbf{r}_j = \delta_{ij}$ for all $t \in (\alpha, \beta)$. Hence the right hand side is identically equal to 0 if and only if $a_{ij}(t) + a_{ji}(t) = 0$ for all $t \in (\alpha, \beta)$.

The conclusion is that for given smooth functions $\kappa(s)$ and $\tau(s)$ on some interval there exists on a sufficiently small open interval around an initial parameter s_0 a positive orthonormal basis $\{\mathbf{t}(s), \mathbf{n}(s), \mathbf{b}(s)\}$ varying smoothly with s near s_0 as solution of the Frenet equations. Moreover it is unique up to a free (positive orthonormal) choice at the initial parameter s_0. Define the smooth curve $s \mapsto \mathbf{r}(s)$ by a direct integration

$$\mathbf{r}(s) = \int \mathbf{t}(s) ds$$

and again the initial position $\mathbf{r}(s_0)$ can be freely prescribed. Since $\dot{\mathbf{r}}(s) = \mathbf{t}(s)$ has unit length this curve is parametrized by arclength, and it has curvature $\kappa(s)$ (now use that $\kappa(s) > 0$) and torsion $\tau(s)$ at time s. Moreover the curve is unique up to a proper orthogonal linear transformation followed by a translation. $\square$

2 Surfaces in Euclidean Space

2.1 The First Fundamental Form

A smooth surface in $\mathbb{R}^3$ is often given by a smooth or even polynomial equation

$$F(x, y, z) = 0$$

and if both $F(x_0, y_0, z_0) = 0$ and $F_z(x_0, y_0, z_0) \neq 0$ then the implicit function theorem says that locally near (x_0, y_0, z_0) the surface is just the graph $z = f(x, y)$ of a smooth function f of the variables (x, y) near the point (x_0, y_0) with $f(x_0, y_0) = z_0$. In other words the surface $F(x, y, z) = 0$ has local coordinates coming from the orthogonal projection of the surface on the plane $z = 0$. However it is more convenient and more practical to give a more general definition, which is just the analogue of the similar situation for plane of space curves.

Definition 2.1. *A smooth surface $\mathcal{S}$ in Euclidean space is a smooth injective map*

$$U \ni (u, v) \mapsto \mathbf{r}(u, v) \in \mathbb{R}^3$$

with U an open subset of $\mathbb{R}^2$ and $\mathbf{r}_u \times \mathbf{r}_v \neq 0$ on all of U.

All our discussions of $\mathcal{S}$ are entirely local, and so we are allowed to shrink U if required. The condition

$$\mathbf{r}_u \times \mathbf{r}_v \neq 0$$

means that the pair of vectors $\{\mathbf{r}_u, \mathbf{r}_v\}$ is linearly independent. Their linear span at (u, v) is called the tangent space at $\mathbf{r}(u, v)$, while the vector

$$\mathbf{N} = \frac{\mathbf{r}_u \times \mathbf{r}_v}{|\mathbf{r}_u \times \mathbf{r}_v|}$$

is called the (unit) normal. Note that the triple

$$\{\mathbf{r}_u, \mathbf{r}_v, \mathbf{N}\}$$

is a basis of $\mathbb{R}^3$.

Definition 2.2. *If we denote*

$$E = \mathbf{r}_u \cdot \mathbf{r}_u \ , \ F = \mathbf{r}_u \cdot \mathbf{r}_v = \mathbf{r}_v \cdot \mathbf{r}_u \ , \ G = \mathbf{r}_v \cdot \mathbf{r}_v$$

as smooth functions on U then the expression

$$\mathrm{I} = Edu^2 + 2Fdudv + Gdv^2$$

is called the first fundamental form of $\mathcal{S}$ on U. Note that $E > 0$, $G > 0$ and $EG - F^2 > 0$.

The coordinates (u, v) on $\mathcal{S}$ are called conformal coordinates if $E = G$ and $F = 0$ for all $(u, v) \in U \subset \mathbb{R}^2$. This means that the angle between two intersecting curves in the coordinate patch U and the angle between the corresponding curves on $\mathcal{S}$ are equal.

If $t \mapsto (u(t), v(t))$ is a smooth curve in U then the arclength of the curve $t \mapsto \mathbf{r}(u(t), v(t))$ on $\mathcal{S}$ in $\mathbb{R}^3$ is given by

$$s = \int \sqrt{E\dot{u}^2 + 2F\dot{u}\dot{v} + G\dot{v}^2}\, dt$$

and therefore we also write

$$ds^2 = Edu^2 + 2Fdudv + Gdv^2$$

for the first fundamental form. The square root

$$ds = \sqrt{Edu^2 + 2Fdudv + gdv^2}$$

is called the length element on $\mathcal{S}$.

The area element dA on $\mathcal{S}$ takes in these coordinates the form

$$dA = \sqrt{EG - F^2}dudv$$

and integration over a compact region R inside U gives the area of the image of R under the map $(u, v) \mapsto \mathbf{r}(u, v)$ inside $\mathcal{S}$.

2.2 The Second Fundamental Form

As before let $U \ni (u, v) \mapsto \mathbf{r}(u, v) \in \mathbb{R}^3$ be a smooth surface $\mathcal{S}$ in Euclidean space $\mathbb{R}^3$, and let

$$\{\mathbf{r}_u, \mathbf{r}_v, \mathbf{N} = \frac{\mathbf{r}_u \times \mathbf{r}_v}{|\mathbf{r}_u \times \mathbf{r}_v|}\}$$

be the associated positive frame in $\mathbb{R}^3$ of tangent vectors and unit normal.

Definition 2.3. *If we denote*

$$L = \mathbf{r}_{uu} \cdot \mathbf{N} \ , \ M = \mathbf{r}_{uv} \cdot \mathbf{N} = \mathbf{r}_{vu} \cdot \mathbf{N} \ , \ N = \mathbf{r}_{vv} \cdot \mathbf{N}$$

as smooth functions on U then the expression

$$\mathrm{II} = L du^2 + 2M du dv + N dv^2$$

is called the second fundamental form of $\mathcal{S}$ on U.

If $t \mapsto (u(t), v(t))$ is a smooth curve in the coordinate region U of the surface $\mathcal{S}$ given by $(u, v) \mapsto \mathbf{r}(u, v)$ then

$$t \mapsto \mathbf{r}(u(t), v(t))$$

is a curve on $\mathcal{S}$ with

$$\mathbf{v} = \mathbf{r}_u \dot{u} + \mathbf{r}_v \dot{v} \ , \ \mathbf{a} = \mathbf{r}_{uu} \dot{u}^2 + 2\mathbf{r}_{uv} \dot{u}\dot{v} + \mathbf{r}_{vv} \dot{v}^2 + \mathbf{r}_u \ddot{u} + \mathbf{r}_v \ddot{v}$$

as velocity and acceleration. Suppose that time t is the arclength parameter s for this curve, so that the velocity $\mathbf{v}$ is the unit tangent $\mathbf{t}$ for all s. Then $\mathbf{a} = \dot{\mathbf{t}}$ implies that $\mathbf{a} \cdot \mathbf{t} = 0$ for all t, and $\mathbf{a} = \kappa \mathbf{n}$ with κ the curvature of the curve, and $\mathbf{n}$ the principal normal (defined as long as $\kappa > 0$). Since the triple $\{\mathbf{t}, \mathbf{N}, \mathbf{t} \times \mathbf{N}\}$ is a positive orthonormal frame we can decompose

$$\dot{\mathbf{t}} = \kappa \mathbf{n} = \kappa_n \mathbf{N} + \kappa_g \mathbf{t} \times \mathbf{N}$$

into a normal and a tangential component. The functions κ_n and κ_g are called the normal curvature and the geodesic curvature of the given curve on $\mathcal{S}$ respectively. Note that $\kappa^2 = \kappa_n^2 + \kappa_g^2$ since the above decomposition of $\dot{\mathbf{t}}$ is orthogonal. It is also clear that the normal curvature is given by

$$\kappa_n = \dot{\mathbf{t}} \cdot \mathbf{N} = \frac{\mathrm{II}}{\mathrm{I}}$$

as quotient of second and first fundamental form in the direction of the vector $(du/ds, dv/ds)$. Indeed the denominator I is just equal to 1 in the direction of $(du/ds, dv/ds)$ by the definition of arclength.

Definition 2.4. *If the curve $s \mapsto \mathbf{r}(u(s), v(s))$ is parametrized by arclength then it is called a geodesic on $\mathcal{S}$ if its geodesic curvature vanishes identically.*

The geodesic curvature is a measure how much the curve $t \mapsto \mathbf{r}(u(t), v(t))$ on $\mathcal{S}$ curves inside $\mathcal{S}$. Geodesics are therefore those curves on $\mathcal{S}$ that are as straight as possible.

The formula for the normal curvature

$$\kappa_n = \frac{\mathrm{II}}{\mathrm{I}} = \frac{L\dot{u}^2 + 2M\dot{u}\dot{v} + N\dot{v}^2}{E\dot{u}^2 + 2F\dot{u}\dot{v} + G\dot{v}^2}$$

as the quotient of the second and first fundamental form in the direction of the vector $(\dot{u}, \dot{v})$ remains valid for any time parameter t for the curve on $\mathcal{S}$. This formula means that two curves on $\mathcal{S}$ that touch each other at a point $\mathbf{r}$ on $\mathcal{S}$ have the same normal curvature κ_n at $\mathbf{r}$.

Let us fix a point $\mathbf{r}$ on $\mathcal{S}$ and let $\mathbf{N}$ be the normal to $\mathcal{S}$ at $\mathbf{r}$. The pencil of planes through the normal line $\mathbf{r} + \mathbb{R}\mathbf{N}$ intersects $\mathcal{S}$ in a family of plane curves, for which at $\mathbf{r}$ the normal curvature κ_n, if traversed in the right direction, is just the signed curvature k. In turn we arrive at the equation

$$k = \frac{L\dot{u}^2 + 2M\dot{u}\dot{v} + N\dot{v}^2}{E\dot{u}^2 + 2F\dot{u}\dot{v} + G\dot{v}^2}$$

with t a time parameter for this family of plane curves on $\mathcal{S}$ through $\mathbf{r}$. Having the point $\mathbf{r}$ on $\mathcal{S}$ fixed and the coefficients E, F, G and L, M, N also fixed numbers we shall view $k = k(\dot{u}, \dot{v})$ as function of the direction $(\dot{u}, \dot{v})$. The extrema of k when both $\dot{u}$ and $\dot{v}$ vary (but not both equal to zero) are given by differentiation of the above formula with respect to $\dot{u}$ and $\dot{v}$. The two equations $\partial k/\partial \dot{u} = 0$ and $\partial k/\partial \dot{v} = 0$ amount to

$$\begin{aligned}
(E\dot{u}^2 + 2F\dot{u}\dot{v} + G\dot{v}^2)(L\dot{u} + M\dot{v}) - (L\dot{u}^2 + 2M\dot{u}\dot{v} + N\dot{v}^2)(E\dot{u} + F\dot{v}) &= 0 \\
(E\dot{u}^2 + 2F\dot{u}\dot{v} + G\dot{v}^2)(M\dot{u} + N\dot{v}) - (L\dot{u}^2 + 2M\dot{u}\dot{v} + N\dot{v}^2)(F\dot{u} + G\dot{v}) &= 0
\end{aligned}$$

and can be simplified to

$$\begin{gathered}
(L - kE)\dot{u} + (M - kF)\dot{v} = 0 \\
(M - kF)\dot{u} + (N - kG)\dot{v} = 0
\end{gathered}$$

which means that k is a root of the characteristic equation

$$\begin{vmatrix} L - kE & M - kF \\ M - kF & N - kG \end{vmatrix} = 0 \, .$$

This equation becomes

$$(EG - F^2)k^2 - (EN - 2MF + GL)k + (LN - M^2) = 0 \, .$$

The two roots of this quadratic equation are called the principal curvatures of $\mathcal{S}$, and will be denoted k_1 and k_2. According to Dirk Struik the principal curvatures were introduced by Monge in 1784 [11].

If $k_1 = k_2$ at some point $\mathbf{r}$ of $\mathcal{S}$ then the first and second fundamental forms are proportional at that point, and $\mathbf{r}$ is called an umbilic of $\mathcal{S}$. If $k_1 \neq k_2$ then the two corresponding solutions $(\dot{u}, \dot{v})$ are well defined up to a nonzero multiple, and called the principal directions. These directions are orthogonal on $\mathcal{S}$, which can be proved using the familiar result from linear algebra that eigenvectors of a symmetric matrix corresponding to different eigenvalues are orthogonal.

Indeed, if we write

$$\mathrm{I} = \begin{pmatrix} E & F \\ F & G \end{pmatrix} \quad , \quad \mathrm{II} = \begin{pmatrix} L & M \\ M & N \end{pmatrix}$$

then $k = k_{1,2}$ are the eigenvalues of the symmetric matrix $\mathrm{I}^{-1/2}\ \mathrm{II}\ \mathrm{I}^{-1/2}$, whose corresponding eigenvectors $(\dot{u}_1, \dot{v}_1)\mathrm{I}^{1/2}$ and $(\dot{u}_2, \dot{v}_2)\mathrm{I}^{1/2}$ are orthogonal in $\mathbb{R}^2$. This means that $(\dot{u}_1, \dot{v}_1)$ and $(\dot{u}_2, \dot{v}_2)$ are orthogonal with respect to I, that is

$$E\dot{u}_1\dot{u}_2 + F(\dot{u}_1\dot{v}_2 + \dot{v}_1\dot{u}_2) + G\dot{v}_1\dot{v}_2 = 0,$$

which means that these two directions are orthogonal on the surface $\mathcal{S}$. Equivalently, the two corresponding planes of the pencil of planes through the line $\mathbf{r} + \mathbb{R}\mathbf{N}$ are perpendicular.

The fixed point $\mathbf{r}$ on $\mathcal{S}$ is called an elliptic point if $LN - M^2 > 0$, and a hyperbolic point if $LN - M^2 < 0$. So for an elliptic point the principal curvatures have the same sign, while for a hyperbolic point the principal curvatures have opposite sign.

The important quantities

$$\begin{aligned} H = \frac{EN - 2FM + GL}{2(EG - F^2)} &= \tfrac{1}{2}(k_1 + k_2) \\ K = \frac{LN - M^2}{EG - F^2} &= k_1 k_2 \end{aligned}$$

are called the mean curvature H and the Gaussian curvature K at the given point of $\mathcal{S}$. A surface whose mean curvature H vanishes identically is called a minimal surface. These are the surfaces one encounters in the Plateau problem: if we have given a closed curve in $\mathbb{R}^3$ then find a (preferably the) surface with boundary this curve and minimal area. Such surfaces are also called soap bubble surfaces.

Since the principal curvatures k_1 and k_2 have geometric meaning they are invariants with respect to proper coordinate transformations. Therefore H and K also remain invariant under proper coordinate transformations. Under improper coordinate transformation both k_1 and k_2 change sign, and therefore H also changes sign while K still remains invariant.

The Gaussian curvature K is the most important notion of curvature for surfaces, and will be further investigated in later chapters.

3 Examples of Surfaces

3.1 Surfaces of Revolution

Consider a surface of revolution $\mathcal{S}$ in standard form

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

as the revolution of the plane curve $\mathcal{C} : u \mapsto \mathbf{r}(u) = (0, f(u), g(u))$ in the plane $x = 0$ around the z axis. The curves on $\mathcal{S}$ with parameter u while v is constant are called meridians, and those with u is constant and parameter v are called circles of latitude.

The coefficients of the first and second fundamental forms are given by

$$E = (f')^2 + (g')^2 \ , \ F = 0 \ , \ G = f^2$$
$$L = \frac{f'g'' - f''g'}{\sqrt{(f')^2 + (g')^2}} \ , \ M = 0 \ , \ N = \frac{fg'}{\sqrt{(f')^2 + (g')^2}}$$

and so the various curvatures become

$$K = \frac{g'(f'g'' - f''g')}{f((f')^2 + (g')^2)^2} \ , \ H = \frac{g'((f')^2 + (g')^2) + f(f'g'' - f''g')}{2f((f')^2 + (g')^2)\sqrt{(f')^2 + (g')^2}}$$
$$k_1 = \frac{g'}{f\sqrt{(f')^2 + (g')^2}} \ , \ k_2 = \frac{f'g'' - f''g'}{((f')^2 + (g')^2)\sqrt{(f')^2 + (g')^2}}$$

by direct calculation. For the unit sphere we take $f(u) = \cos u, g(u) = \sin u$ and as a result get $K = H = k_1 = k_2 = 1$. Hence the unit sphere has constant Gaussian curvature equal to 1.

3.2 Charts for the Sphere

Consider the unit sphere $\mathbb{S}$ with equation $x^2 + y^2 + z^2 = 1$ as a surface of revolution

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

by rotation of the unit circle $\mathcal{C} : u \mapsto \mathbf{r}(u) = (0, f(u), g(u))$ with $f^2 + g^2 \equiv 1$ in the plane $x = 0$ around the z axis.

The Archimedes projection is given by

$$f(u) = \sqrt{1 - u^2} \ , \ g(u) = u \ , \ -1 < u < 1$$

which is just the horizontal projection from the cylinder

$$x^2 + y^2 = 1 \ , \ -1 < z < 1$$

onto the sphere minus the north pole $\mathbf{n}$ and the south pole $\mathbf{s}$. Since

$$f' = \frac{-u}{\sqrt{1-u^2}} \ , \ g' = 1$$

we get

$$E = \frac{1}{1-u^2} \ , \ F = 0 \ , \ G = (1-u^2)$$

which implies $EG - F^2 = 1$. Hence the Archimedes projection yields equiareal coordinates. In this way Archimedes (287-212 BC) found the area of the unit sphere to be 4π. In cartography the Archimedes projection is usually called the Gall–Peters projection named after the cartographers James Gall (1808-1895) and Arno Peters (1916-2002). The Gall–Peters projection gave some controversy in the late 20th century by its claim of politically correct map design.

The equirectangular projection introduced by the Greek mathematician Marinus of Tyre (70-130 AD) is given by

$$f(u) = \cos u \ , \ g(u) = \sin u \ , \ -\pi/2 < u < \pi/2 \ .$$

It is a suitable projection from the cylinder

$$x^2 + y^2 = 1 \ , \ -\pi/2 < z < \pi/2$$

onto the sphere minus the north and south pole. Since

$$f' = -\sin u \ , \ g' = \cos u$$

we get

$$E = 1 \ , \ F = 0 \ , \ G = \cos^2 u$$

and so the equirectangular projection is equidistant along meridians.

The Mercator projection is given by

$$f(u) = \frac{1}{\cosh u} \ , \ g(u) = \frac{\sinh u}{\cosh u} \ , \ -\infty < u < \infty$$

which is a suitable nonlinear projection from the full cylinder

$$x^2 + y^2 = 1, -\infty < z < \infty$$

onto the sphere minus the north and south pole. Since

$$f' = \frac{-\sinh u}{\cosh^2 u} \ , \ g' = \frac{1}{\cosh^2 u}$$

we get

$$E = \frac{1}{\cosh^2 u} \ , \ F = 0 \ , \ G = \frac{1}{\cosh^2 u}$$

and so the Mercator projection yields conformal coordinates on $\mathbb{S} - \{\mathbf{n}, \mathbf{s}\}$. The straight lines $u = av + b$ with constant a in the plane map onto curves on the sphere which intersect the meridians under a constant angle. These curves are called loxodromes, and were used in the old days for navigation. This was the reason Mercator (1512-1594) invented his projection.

The stereographic projection is the linear projection from the plane $z = 0$ with center the north pole $\mathbf{n} = (0, 0, 1)$. A point $\mathbf{p} = (u, v, 0)$ in the plane is projected onto the point $\mathbf{q} = \lambda\mathbf{p} + (1 - \lambda)\mathbf{n}$ on the sphere. Since $|\mathbf{q}| = 1$ we get $\lambda = 2/(u^2 + v^2 + 1)$. So the stereographic projection is given by

$$\mathbf{r}(u, v) = \left(\frac{2u}{u^2 + v^2 + 1}, \frac{2v}{u^2 + v^2 + 1}, \frac{u^2 + v^2 - 1}{u^2 + v^2 + 1}\right)$$

and projects onto the sphere minus the north pole $\mathbf{n}$. A direct computation yields

$$\begin{aligned} \mathbf{r}_u &= (-2u^2 + 2v^2 + 2, -4uv, -4u)/(u^2 + v^2 + 1)^2 \\ \mathbf{r}_v &= (-4uv, +2u^2 - 2v^2 + 2, -4u)/(u^2 + v^2 + 1)^2 \end{aligned}$$

which in turn implies

$$E = \frac{4}{(u^2 + v^2 + 1)^2} \ , \ F = 0 \ , \ G = \frac{4}{(u^2 + v^2 + 1)^2} \ .$$

The conclusion is that stereographic projection yields conformal coordinates on $\mathbb{S} - \{\mathbf{n}\}$. Stereographic projection was already known in ancient Greek civilization to Hipparchus and Ptolemy. The first known world map based upon stereographic projection, mapping each hemisphere onto a circular disc, was made in 1507 by Gualterious Lud.

It is easy to verify that the inverse stereographic projection is given by

$$(x, y, z) \mapsto \left(\frac{x}{1 - z}, \frac{y}{1 - z}\right) \ .$$

The composition of inverse stereographic projection and Mercator projection becomes

$$(u, v) \mapsto \left(\frac{\cos v}{\cosh u}, \frac{\sin v}{\cosh u}, \frac{\sinh u}{\cosh u}\right) \mapsto \left(\frac{\cos v}{\cosh u - \sinh u}, \frac{\sin v}{\cosh u - \sinh u}\right)$$

or in complex notation

$$w = u + iv \mapsto \frac{\cos v + i \sin v}{\cosh u - \sinh u} = \frac{\exp(iv)}{\exp(-u)} = \exp(u + iv) = e^w$$

which is a holomorphic function as should. Indeed, it is conformal as composition of two conformal transformations. On the complex plane it becomes a holomorphic transformation with nowhere vanishing derivative.

It is convenient to compose the stereographic projection from the south pole $\mathbf{s} = (0, 0, -1)$ with complex conjugation of the (u, v) plane, in order that the transition from one stereographic projection to the other preserves the orientation. It is given by the equation $\mathbf{q} = \lambda\mathbf{p} + (1 - \lambda)\mathbf{s}$ with $|\mathbf{q}| = 1$ and $\mathbf{p} = (u, -v) \in \mathbb{R}^2$ with $\lambda = 2/(u^2 + v^2 + 1)$, and boils down to the formula

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

Because $1 - z = 2(u^2 + v^2)/(u^2 + v^2 + 1)$ the composition of inverse stereographic projection from the north pole with stereographic projection from the south pole is given in complex notation by

$$w = u + iv \mapsto \frac{u - iv}{u^2 + v^2} = \frac{1}{w}$$

which is again holomorphic on the punctured complex plane with nowhere vanishing deriviative.

3.3 Ruled Surfaces

A ruled surface $\mathcal{S}$ is a smooth surface given by a parametrization

$$(u, v) \mapsto \mathbf{r}(u, v) = \boldsymbol{\gamma}(u) + v\boldsymbol{\delta}(u)$$

with $t \mapsto \boldsymbol{\gamma}(t)$ and $t \mapsto \boldsymbol{\delta}(t)$ two smooth space curves. We shall always have in mind that v is taken from an open interval around 0 that might depend on u. The line segments with u constant and v as parameter are called the rulings of the ruled surface. In order that $\mathcal{S}$ is smooth we need to require that

$$\mathbf{r}_u \times \mathbf{r}_v = (\dot{\boldsymbol{\gamma}} + v\dot{\boldsymbol{\delta}}) \times \boldsymbol{\delta} \neq 0$$

for all (u, v) in the domain of definition. This amounts to

$$\dot{\boldsymbol{\gamma}} \times \boldsymbol{\delta} \neq 0$$

by possibly shrinking the interval of definition for the parameter v. Since $\mathbf{r}_{vv} = 0$ the third coefficient

$$N = \mathbf{r}_{vv} \cdot \mathbf{N}$$

of the second fundamental form vanishes everywhere, which in turn implies that the Gaussian curvature

$$K = \frac{-M^2}{EG - F^2} \leq 0$$

is nonpositive everywhere on $\mathcal{S}$. Examples of ruled surfaces are the one sheeted hyperboloid with equation

$$x^2 + y^2 - z^2 = 1$$

(check this by taking $\gamma(u) = (\cos u, \sin u, 0)$ and looking for $\delta(u)$) and the Möbius band with parametrization

$$\mathbf{r}(u, v) = (\cos u, \sin u, 0) + v(-\sin(u/2)\cos u, -\sin(u/2)\sin u, \cos(u/2))$$

with $u \in \mathbb{R}/2\pi\mathbb{Z}$ and say $|v| < 1/2$, because

$$\boldsymbol{\gamma}(u) = \begin{pmatrix} \cos u & -\sin u & 0 \\ \sin u & \cos u & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$$

while

$$\boldsymbol{\delta}(u) = \begin{pmatrix} \cos u & -\sin u & 0 \\ \sin u & \cos u & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \cos(u/2) & 0 & -\sin(u/2) \\ 0 & 1 & 0 \\ \sin(u/2) & 0 & \cos(u/2) \end{pmatrix} \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$$

This Möbius band has everywhere $K < 0$ and so is metrically different from the flat Möbius band with $K \equiv 0$ obtained by joining the ends of a strip of paper after first performing a halft twist! This remark should be read again after the reader has absorbed the Theorema Egregium in the next chapter.

4 Theorema Egregium of Gauss

4.1 The Gauss Relations

Suppose we have given a smooth surface $\mathcal{S}$ in Euclidean space $\mathbb{R}^3$ given by a smooth coordinate map

$$U \ni (u,v) \mapsto \mathbf{r}(u,v) \in \mathbb{R}^3$$

for some open subset $U$of the Euclidean plane $\mathbb{R}^2$. The coefficients of the first fundamental form

$$\mathrm{I} = Edu^2 + 2Fdudv + Gdv^2$$

are by definition the coefficients of the Gram matrix of the basis $\{\mathbf{r}_u, \mathbf{r}_v\}$ of the tangent space to $\mathcal{S}$. If we denote the unit normal of $\mathcal{S}$ by

$$\mathbf{N} = \frac{\mathbf{r}_u \times \mathbf{r}_v}{|\mathbf{r}_u \times \mathbf{r}_v|}$$

then $\{\mathbf{r}_u, \mathbf{r}_v, \mathbf{N}\}$ is a basis of $\mathbb{R}^3$, and we obtain the Gauss relations

$$\begin{aligned}
\mathbf{r}_{uu} &= \Gamma^1_{11}\mathbf{r}_u + \Gamma^2_{11}\mathbf{r}_v + \ L\mathbf{N} \\
\mathbf{r}_{uv} &= \Gamma^1_{12}\mathbf{r}_u + \Gamma^2_{12}\mathbf{r}_v + M\mathbf{N} \\
\mathbf{r}_{vv} &= \Gamma^1_{22}\mathbf{r}_u + \Gamma^2_{22}\mathbf{r}_v + \ N\mathbf{N}
\end{aligned}$$

with suitable coefficients Γ^k_{ij} and L, M, N being smooth functions on U. Taking the inner product with $\mathbf{N}$ shows that L, M, N are the coefficients of the second fundamental form

$$\mathrm{II} = Ldu^2 + 2Mdudv + Ndv^2$$

of $\mathcal{S}$ in accordance with Definition 2.3. The coefficients Γ^k_{ij} are called the Christoffel symbols. The next theorem is due to Gauss.

Theorem 4.1. *The Christoffel symbols are given by the Gauus relations*

$$\begin{aligned}
\Gamma^1_{11} &= \frac{GE_u - 2FF_u + FE_v}{2(EG-F^2)} \quad , \quad \Gamma^2_{11} = \frac{2EF_u - EE_v - FE_u}{2(EG-F^2)} \\
\Gamma^1_{12} &= \frac{GE_v - FG_u}{2(EG-F^2)} \quad , \quad \Gamma^2_{12} = \frac{EG_u - FE_v}{2(EG-F^2)} \\
\Gamma^1_{22} &= \frac{2GF_v - GG_u - FG_v}{2(EG-F^2)} \quad , \quad \Gamma^2_{22} = \frac{EG_v - 2FF_v + FG_u}{2(EG-F^2)}
\end{aligned}$$

as functions of the coefficients E, F, G and their first order derivatives.

Proof. Taking the inner product of the Gauss relations with the tangent vectors $\mathbf{r}_u$ and $\mathbf{r}_v$ yields the linear equations

$$\begin{aligned}
\mathbf{r}_{uu}\cdot\mathbf{r}_u &= \Gamma^1_{11}E+\Gamma^2_{11}F \quad , \quad \mathbf{r}_{uu}\cdot\mathbf{r}_v = \Gamma^1_{11}F+\Gamma^2_{11}G \\
\mathbf{r}_{uv}\cdot\mathbf{r}_u &= \Gamma^1_{12}E+\Gamma^2_{12}F \quad , \quad \mathbf{r}_{uv}\cdot\mathbf{r}_v = \Gamma^1_{12}F+\Gamma^2_{12}G \\
\mathbf{r}_{vv}\cdot\mathbf{r}_u &= \Gamma^1_{22}E+\Gamma^2_{22}F \quad , \quad \mathbf{r}_{vv}\cdot\mathbf{r}_v = \Gamma^1_{22}F+\Gamma^2_{22}G
\end{aligned}$$

with solutions

$$\begin{aligned}
\Gamma^1_{11} &= \frac{G\mathbf{r}_{uu}\cdot\mathbf{r}_u - F\mathbf{r}_{uu}\cdot\mathbf{r}_v}{EG-F^2} \quad , \quad \Gamma^2_{11} = \frac{E\mathbf{r}_{uu}\cdot\mathbf{r}_v - F\mathbf{r}_{uu}\cdot\mathbf{r}_u}{EG-F^2} \\
\Gamma^1_{12} &= \frac{G\mathbf{r}_{uv}\cdot\mathbf{r}_u - F\mathbf{r}_{uv}\cdot\mathbf{r}_v}{EG-F^2} \quad , \quad \Gamma^2_{12} = \frac{E\mathbf{r}_{uv}\cdot\mathbf{r}_v - F\mathbf{r}_{uv}\cdot\mathbf{r}_u}{EG-F^2} \\
\Gamma^1_{22} &= \frac{G\mathbf{r}_{vv}\cdot\mathbf{r}_u - F\mathbf{r}_{vv}\cdot\mathbf{r}_v}{EG-F^2} \quad , \quad \Gamma^2_{22} = \frac{E\mathbf{r}_{vv}\cdot\mathbf{r}_v - F\mathbf{r}_{vv}\cdot\mathbf{r}_u}{EG-F^2}
\end{aligned}$$

On the other hand

$$\begin{aligned}
E_u = 2\mathbf{r}_{uu}\cdot\mathbf{r}_u \ , \ E_v = 2\mathbf{r}_{uv}\cdot\mathbf{r}_u \ , \ F_u = \mathbf{r}_{uu}\cdot\mathbf{r}_v + \mathbf{r}_{uv}\cdot\mathbf{r}_u \\
G_u = 2\mathbf{r}_{uv}\cdot\mathbf{r}_v \ , \ G_v = 2\mathbf{r}_{vv}\cdot\mathbf{r}_v \ , \ F_v = \mathbf{r}_{uv}\cdot\mathbf{r}_v + \mathbf{r}_{vv}\cdot\mathbf{r}_u
\end{aligned}$$

which in turn implies that

$$\begin{aligned}
\mathbf{r}_{uu}\cdot\mathbf{r}_u = E_u/2 \ , \ \mathbf{r}_{uv}\cdot\mathbf{r}_u = E_v/2 \ , \ \mathbf{r}_{uu}\cdot\mathbf{r}_v = F_u - E_v/2 \\
\mathbf{r}_{uv}\cdot\mathbf{r}_v = G_u/2 \ , \ \mathbf{r}_{vv}\cdot\mathbf{r}_v = G_v/2 \ , \ \mathbf{r}_{vv}\cdot\mathbf{r}_u = F_v - G_u/2
\end{aligned}$$

and the given expressions for the Christoffel symbols follow by substitution. □

This theorem enables one to obtain a partial analogue of Theorem 1.4.

Corollary 4.2. *A smooth surface in Euclidean space $\mathbb{R}^3$ is determined up to a proper Euclidean motion by its first and second fundamental forms.*

Proof. Given the coefficents of the first and second fundamental forms the Gauss relations become a system of second order partial differential equations, and as such have a unique solution

$$(u,v)\mapsto \mathbf{r}(u,v)$$

up to a proper Euclidean motion of $\mathbb{R}^3$. Indeed the freedom is the prescription of $\mathbf{r}$ and of $\{\mathbf{r}_u,\mathbf{r}_v\}$ for some initial point $(u_0,v_0)\in U$ with the restriction that the Gram matrix of $\{\mathbf{r}_u,\mathbf{r}_v\}$ at the initial point is given by the first fundamental form. □

However it is not true that both the coefficients E, F, G and L, M, N can be independently prescribed. The Gauss relations have to satisfy the compatibility conditions

$$(\mathbf{r}_{uu})_v = (\mathbf{r}_{uv})_u \ , \ (\mathbf{r}_{vv})_u = (\mathbf{r}_{uv})_v$$

and then can be solved according to the Frobenius integrability theorem. These compatibility conditions boil down to three independent differential relations among the 6 coefficients E, F, G and L, M, N. This is what one should expect, because we look for the 3 components of the solution vector $\mathbf{r}(u, v)$ and so in the coefficients of the Gauss relations there should also be 3 independent coefficients. In the next section we shall work out the compatibility conditions for the Gauss relations explicitly.

4.2 The Codazzi–Mainardi and Gauss Equations

Let $U \ni (u, v) \mapsto \mathbf{r}(u, v) \in \mathbb{R}^3$ be a smooth surface. The frame $\{\mathbf{r}_u, \mathbf{r}_v, \mathbf{N}\}$ is a basis of $\mathbb{R}^3$. Since $\mathbf{N} \cdot \mathbf{N} \equiv 1$ the partial derivatives $\mathbf{N}_u$ and $\mathbf{N}_v$ can be written as a linear combination of $\{\mathbf{r}_u, \mathbf{r}_v\}$.

Theorem 4.3. *We have the Weingarten equations*

$$\begin{aligned} \mathbf{N}_u &= \frac{MF - LG}{EG - F^2}\mathbf{r}_u + \frac{LF - ME}{EG - F^2}\mathbf{r}_v \\ \mathbf{N}_v &= \frac{NF - MG}{EG - F^2}\mathbf{r}_u + \frac{MF - NE}{EG - F^2}\mathbf{r}_v \end{aligned}$$

Proof. If we write

$$\mathbf{N}_u = a\mathbf{r}_u + b\mathbf{r}_v \ , \ \mathbf{N}_v = c\mathbf{r}_u + d\mathbf{r}_v$$

for suitable $a, b, c, d \in \mathbb{R}$ then

$$\begin{aligned} Ea + Fb = \mathbf{N}_u \cdot \mathbf{r}_u = -L \quad &, \quad Ec + Fd = \mathbf{N}_v \cdot \mathbf{r}_u = -M \\ Fa + Gb = \mathbf{N}_u \cdot \mathbf{r}_v = -M \quad &, \quad Fc + Gd = \mathbf{N}_v \cdot \mathbf{r}_v = -N \end{aligned}$$

by taking inner products with $\mathbf{r}_u$ and $\mathbf{r}_v$. Solving these linear equations proves the result. $\square$

These equations were found by Weingarten in 1861. After this preparation we shall now work out the compatibility conditions

$$(\mathbf{r}_{uu})_v = (\mathbf{r}_{uv})_u \ , \ (\mathbf{r}_{vv})_u = (\mathbf{r}_{uv})_v$$

for the Gauss relations of the previous section. Their components in the normal direction lead to the Codazzi–Mainardi equations. These formulas were found independently by Codazzi in 1860 and Mainardi in 1856. Their components in the tangential directions lead to the Gauss equations.

Theorem 4.4. *The coefficients of the first and second fundamental forms satisfy the Codazzi–Mainardi equations*

$$\begin{aligned} L_v - M_u &= L\Gamma_{12}^1 + M(\Gamma_{12}^2 - \Gamma_{11}^1) - N\Gamma_{11}^2 \\ M_v - N_u &= L\Gamma_{22}^1 + M(\Gamma_{22}^2 - \Gamma_{12}^1) - N\Gamma_{12}^2 \end{aligned}$$

and the Gauss equations

$$\begin{aligned} EK &= (\Gamma_{11}^2)_v - (\Gamma_{12}^2)_u + \Gamma_{11}^1\Gamma_{12}^2 + \Gamma_{11}^2\Gamma_{22}^2 - \Gamma_{12}^1\Gamma_{11}^2 - \Gamma_{12}^2\Gamma_{12}^2 \\ FK &= (\Gamma_{12}^1)_u - (\Gamma_{11}^1)_v + \Gamma_{12}^2\Gamma_{12}^1 - \Gamma_{11}^2\Gamma_{22}^1 \\ FK &= (\Gamma_{12}^2)_v - (\Gamma_{22}^2)_u + \Gamma_{12}^1\Gamma_{12}^2 - \Gamma_{22}^1\Gamma_{11}^2 \\ GK &= (\Gamma_{22}^1)_u - (\Gamma_{12}^1)_v + \Gamma_{22}^1\Gamma_{11}^1 + \Gamma_{22}^2\Gamma_{12}^1 - \Gamma_{12}^1\Gamma_{12}^1 - \Gamma_{12}^2\Gamma_{22}^1 \end{aligned}$$

as normal and tangential components of the compatibility conditions.

Proof. Substitution of the Gauss relations in the compatibility conditions

$$(\mathbf{r}_{uu})_v = (\mathbf{r}_{uv})_u \ , \ (\mathbf{r}_{vv})_u = (\mathbf{r}_{uv})_v$$

gives the equations

$$\begin{aligned} \frac{\partial}{\partial v}(\Gamma_{11}^1\mathbf{r}_u + \Gamma_{11}^2\mathbf{r}_v + L\mathbf{N}) &= \frac{\partial}{\partial u}(\Gamma_{12}^1\mathbf{r}_u + \Gamma_{12}^2\mathbf{r}_v + M\mathbf{N}) \\ \frac{\partial}{\partial u}(\Gamma_{22}^1\mathbf{r}_u + \Gamma_{22}^2\mathbf{r}_v + N\mathbf{N}) &= \frac{\partial}{\partial v}(\Gamma_{12}^1\mathbf{r}_u + \Gamma_{12}^2\mathbf{r}_v + M\mathbf{N}) \end{aligned}$$

for any surface $\mathcal{S}$ in $\mathbb{R}^3$. Writing out these equations in terms of the basis $\{\mathbf{r}_u, \mathbf{r}_v, \mathbf{N}\}$ gives (using the Weingarten equations)

$$\begin{gathered} (\Gamma_{11}^1)_v + \Gamma_{11}^1\Gamma_{12}^1 + \Gamma_{11}^2\Gamma_{22}^1 + L\frac{NF - MG}{EG - F^2} = \\ (\Gamma_{12}^1)_u + \Gamma_{12}^1\Gamma_{11}^1 + \Gamma_{12}^2\Gamma_{12}^1 + M\frac{MF - LG}{EG - F^2} \ , \\ (\Gamma_{22}^1)_u + \Gamma_{22}^1\Gamma_{11}^1 + \Gamma_{22}^2\Gamma_{12}^1 + N\frac{MF - LG}{EG - F^2} = \\ (\Gamma_{12}^1)_v + \Gamma_{12}^1\Gamma_{12}^1 + \Gamma_{12}^2\Gamma_{22}^1 + M\frac{NF - MG}{EG - F^2} \end{gathered}$$

as the coefficients of $\mathbf{r}_u$, and

$$(\Gamma_{11}^2)_v + \Gamma_{11}^1\Gamma_{12}^2 + \Gamma_{11}^2\Gamma_{22}^2 + L\frac{MF-NE}{EG-F^2} =$$
$$(\Gamma_{12}^2)_u + \Gamma_{12}^1\Gamma_{11}^2 + \Gamma_{12}^2\Gamma_{12}^2 + M\frac{LF-ME}{EG-F^2}\ ,$$
$$(\Gamma_{22}^2)_u + \Gamma_{22}^1\Gamma_{11}^2 + \Gamma_{22}^2\Gamma_{12}^2 + N\frac{LF-ME}{EG-F^2} =$$
$$(\Gamma_{12}^2)_v + \Gamma_{12}^1\Gamma_{12}^2 + \Gamma_{12}^2\Gamma_{22}^2 + M\frac{MF-NE}{EG-F^2}$$

as the coefficients of $\mathbf{r}_v$, and

$$M\Gamma_{11}^1 + N\Gamma_{11}^2 + L_v = L\Gamma_{12}^1 + M\Gamma_{12}^2 + M_u$$
$$L\Gamma_{22}^1 + M\Gamma_{22}^2 + N_u = M\Gamma_{12}^1 + N\Gamma_{12}^2 + M_v$$

as the coefficients of $\mathbf{N}$ respectively. The tangential components of the compatibility conditions reduce to the Gauss equations, while the normal components give the Codazzi–Mainardi equations. $\square$

Given smooth functions E, F, G and L, M, N on a planar region U with $E, G > 0$ and $EG - F^2 > 0$ that satisfy the Codazzi-Mainardi equations and the Gauss equations of Theorem 4.4 (with the Christoffel symbols given by Theorem 4.1) it follows that there exists a smooth surface in Euclidean space $\mathbb{R}^3$ with these functions as coefficients of the first and second fundamental form. This result is called the fundamental theorem for surfaces in $\mathbb{R}^3$. It goes back to Bonnet (1867) and follows from the Frobenius integrability theorem.

4.3 The Remarkable Theorem of Gauss

The results of the previous two sections are rather computational. But the catch is that the Christoffel symbols as given by the Gauss equations in Theorem 4.1 and the expressions for the Gaussian curvature K as given through the Gauss equations of Theorem 4.4 lead to a remarkable theorem.

Theorem 4.5. *The Gauss curvature K of a surface $\mathcal{S}$ in $\mathbb{R}^3$ depends only of the coefficients E, F, G of the first fundamental form (through their partial derivatives of order at most 2).*

Proof. Indeed the Gauss equations of Theorem 4.4 give a formula for K in terms of the partial derivatives of at most first order of the Christoffel symbols, while in turn the Christoffel symbols depend on E, F, G through partial derivatives of at most first order by Theorem 4.1. $\square$

Gauss called this theorem the "Theorema Egregium", and it is indeed a remarkable result. Let us call a differential geometric quantity for a surface $\mathcal{S}$ in $\mathbb{R}^3$ an inner quantity if it relies only on the length element ds^2 on $\mathcal{S}$, so if in local coordinates $(u,v) \mapsto \mathbf{r}(u,v)$ on $\mathcal{S}$ it relies only on the coefficients E, F, G of the first fundamental form. Inner differential geometry is the differential geometry that is meaningful to a flatlander living on $\mathcal{S}$.

The coefficients L, M, N of the second fundamental form are not inner quantities. They are given as the components of the second order derivatives of the coordinates $(u,v) \mapsto \mathbf{r}(u,v)$ in the direction of the normal $\mathbf{N}$. So their calculation requires the ambient Euclidean space $\mathbb{R}^3$ containing $\mathcal{S}$. Therefore it is remarkable that the Gaussian curvature $K = (LN - M^2)/(EG - F^2)$ is an inner quantity, expressible in terms of the coefficients E, F, G of the first fundamental form.

Corollary 4.6. *It is impossible to choose local coordinates $(u,v) \mapsto \mathbf{r}(u,v)$ for the unit sphere in $\mathbb{R}^3$ such that the length element on the sphere becomes the planar Euclidean length element $ds^2 = du^2 + dv^2$.*

Proof. The Gaussian curvature of the unit sphere is constant equal to 1 while the Gaussian curvature of a planar Euclidean region is constant equal to 0. Hence the Theorema Egregium prevents such flat coordinates for the sphere. $\square$

This fact must have been conjectured by many cartographers before Gauss, but the Theorema Egregium provides the first rigorous proof. There is another application of the Theorema Egregium that was already alluded to at the end of Section 3.3.

Corollary 4.7. *It is impossible to choose new local coordinates (x,y) on the Möbius band $\mathcal{S}$ given by*

$$\mathbf{r}(u,v) = (\cos u, \sin u, 0) + v(-\sin(u/2)\cos u, -\sin(u/2)\sin u, \cos(u/2))$$

with $u \in \mathbb{R}/2\pi\mathbb{Z}$ and say $|v| < 1/2$ such that the length element in the new coordinates becomes the planar Euclidean length element $ds^2 = dx^2 + dy^2$.

Proof. Indeed, we computed in Section 3.3 that the Gauss curvature K of the Möbius band $\mathcal{S}$ is strictly negative on all of $\mathcal{S}$. If these new coordinates would exist then K vanishes identically, which is a contradiction with the Theorema Egregium. $\square$

The Gauss equations in Theorem 4.4 give in fact 4 expressions for the Gaussian curvature K as function of E, F, G and their partial derivatives up to order 2. It turns out that all 4 expressions lead to the same formula, whose explicit form was worked out by Brioschi (1852) and Baltzer (1866).

Theorem 4.8. *The explicit expression for the Gaussian curvature K as a function of E, F, G in the Theorema Egregium takes the form*

$$K(EG - F^2)^2 = (-E_{vv}/2 + F_{uv} - G_{uu}/2)(EG - F^2) +$$
$$\begin{vmatrix} 0 & E_u/2 & F_u - E_v/2 \\ F_v - G_u/2 & E & F \\ G_v/2 & F & G \end{vmatrix} - \begin{vmatrix} 0 & E_v/2 & G_u/2 \\ E_v/2 & E & F \\ G_u/2 & F & G \end{vmatrix}$$

Proof. We start with the formula

$$K = \frac{LN - M^2}{EG - F^2}$$

that was used as definition of the Gaussian curvature. From

$$L = \mathbf{r}_{uu} \cdot \mathbf{N} \ , \ M = \mathbf{r}_{uv} \cdot \mathbf{N} \ , \ N = \mathbf{r}_{vv} \cdot \mathbf{N}$$

together with

$$\mathbf{N} = \frac{\mathbf{r}_u \times \mathbf{r}_v}{\sqrt{EG - F^2}}$$

we obtain

$$K(EG - F^2)^2 = (\mathbf{r}_{uu} \cdot (\mathbf{r}_u \times \mathbf{r}_v))(\mathbf{r}_{vv} \cdot (\mathbf{r}_u \times \mathbf{r}_v)) - (\mathbf{r}_{uv} \cdot (\mathbf{r}_u \times \mathbf{r}_v))^2 \ .$$

Using $\mathbf{a} \cdot (\mathbf{b} \times \mathbf{c}) = \det(\mathbf{a}\ \mathbf{b}\ \mathbf{c})$ we get

$$(\mathbf{a} \cdot (\mathbf{b} \times \mathbf{c}))(\mathbf{d} \cdot (\mathbf{e} \times \mathbf{f})) = \begin{vmatrix} \mathbf{a} \cdot \mathbf{d} & \mathbf{a} \cdot \mathbf{e} & \mathbf{a} \cdot \mathbf{f} \\ \mathbf{b} \cdot \mathbf{d} & \mathbf{b} \cdot \mathbf{e} & \mathbf{b} \cdot \mathbf{f} \\ \mathbf{c} \cdot \mathbf{d} & \mathbf{c} \cdot \mathbf{e} & \mathbf{c} \cdot \mathbf{f} \end{vmatrix}$$

for any six vectors $\mathbf{a}, \mathbf{b}, \mathbf{c}, \mathbf{d}, \mathbf{e}, \mathbf{f}$ in $\mathbb{R}^3$. Hence

$$K(EG - F^2)^2 = \begin{vmatrix} \mathbf{r}_{uu} \cdot \mathbf{r}_{vv} & \mathbf{r}_{uu} \cdot \mathbf{r}_u & \mathbf{r}_{uu} \cdot \mathbf{r}_v \\ \mathbf{r}_u \cdot \mathbf{r}_{vv} & E & F \\ \mathbf{r}_v \cdot \mathbf{r}_{vv} & F & G \end{vmatrix} - \begin{vmatrix} \mathbf{r}_{uv} \cdot \mathbf{r}_{uv} & \mathbf{r}_{uv} \cdot \mathbf{r}_u & \mathbf{r}_{uv} \cdot \mathbf{r}_v \\ \mathbf{r}_u \cdot \mathbf{r}_{uv} & E & F \\ \mathbf{r}_v \cdot \mathbf{r}_{uv} & F & G \end{vmatrix}$$

and so

$$K(EG - F^2)^2 = (\mathbf{r}_{uu} \cdot \mathbf{r}_{vv} - \mathbf{r}_{uv} \cdot \mathbf{r}_{uv})(EG - F^2) +$$
$$\begin{vmatrix} 0 & \mathbf{r}_{uu} \cdot \mathbf{r}_u & \mathbf{r}_{uu} \cdot \mathbf{r}_v \\ \mathbf{r}_u \cdot \mathbf{r}_{vv} & E & F \\ \mathbf{r}_v \cdot \mathbf{r}_{vv} & F & G \end{vmatrix} - \begin{vmatrix} 0 & \mathbf{r}_{uv} \cdot \mathbf{r}_u & \mathbf{r}_{uv} \cdot \mathbf{r}_v \\ \mathbf{r}_u \cdot \mathbf{r}_{uv} & E & F \\ \mathbf{r}_v \cdot \mathbf{r}_{uv} & F & G \end{vmatrix}$$

and the result follows if we can express the right hand side in terms of the first fundamental form.

By definition the coefficients E, F, G of the first fundamental form are given by

$$E = \mathbf{r}_u \cdot \mathbf{r}_u \ , \ F = \mathbf{r}_u \cdot \mathbf{r}_v \ , \ G = \mathbf{r}_v \cdot \mathbf{r}_v$$

and differentiation results in the identities (as in the proof of Theorem 4.1)

$$\begin{aligned} \mathbf{r}_{uu} \cdot \mathbf{r}_u &= E_u/2 \ , \ \mathbf{r}_{uv} \cdot \mathbf{r}_u = E_v/2 \ , \ \mathbf{r}_{uu} \cdot \mathbf{r}_v = F_u - E_v/2 \\ \mathbf{r}_{uv} \cdot \mathbf{r}_v &= G_u/2 \ , \ \mathbf{r}_{vv} \cdot \mathbf{r}_v = G_v/2 \ , \ \mathbf{r}_{vv} \cdot \mathbf{r}_u = F_v - G_u/2 \end{aligned}$$

Differentiation of the third expression with respect to v and of the fourth with respect to u, followed by a subtraction, yields the identity

$$\mathbf{r}_{uu} \cdot \mathbf{r}_{vv} - \mathbf{r}_{uv} \cdot \mathbf{r}_{uv} = -E_{vv}/2 + F_{uv} - G_{uu}/2$$

and the desired formula for the Gaussian curvature K follows by a direct substitution. □

These clever calculations give a second proof of the Theorema Egregium, but it is quite clear that the theorem was not discovered this way. Both proofs of the Theorema Egregium work with general coordinates (u, v) which might be part of the reason why the calculations tend to be cumbersome.

We end this section with a geometric but intuitive argument of Hilbert why the Theorema Egregium ought to be true. The Gauss–Rodrigues map for a surface $\mathcal{S}$ in $\mathbb{R}^3$ as given in local coordinates

$$(u, v) \mapsto \mathbf{r}(u, v)$$

is the map from $\mathcal{S}$ to the two dimensional unit sphere $\mathcal{S}^2$ by sending the point $\mathbf{r}(u, v)$ to the unit normal $\mathbf{N}(u, v)$. The important formula

$$\mathbf{N}_u \times \mathbf{N}_v = K \mathbf{r}_u \times \mathbf{r}_v$$

says that the Jacobian of the Gauss–Rodrigues map equals the Gaussian curvature. The proof of this formula follows directly from the Weingarten equations. In most text books the Gauss–Rodrigues map is just called the Gauss map, but it was already introduced before Gauss by Rodrigues in 1815, who used it to prove that the total Gaussian curvature of an ellipsoid is equal to 4π.

The point is now that we shall think of the smooth surface $\mathcal{S}$ as build from a triangulation of Euclidean triangles. The total Gaussian curvature

$$K(u, v) dA = K(u, v) \sqrt{EG - F^2} dudv$$

then becomes a weighted sum of delta functions located at the vertices of the triangulation. The weight at a given vertex is the area of the region R of unit normal vectors, directed in outward direction for the given orientation.

Say at the given vertex n triangles come together with angles $\alpha_1, \cdots, \alpha_n$. Then it is not difficult to see that the region R of outward unit normal vectors is a spherical polygon with n vertices and angles equal to $\pi - \alpha_1, \cdots, \pi - \alpha_n$. This is clear by replacing the n triangles by n quadrangles, each with one pair of opposite orthogonal angles and the other opposite pair with angles α_i and $\pi - \alpha_i$. Moreover take the pairwise glued edges of the same length. It is instructive to make a paper model in case $n = 4$.

The total Gaussian curvature at the vertex becomes the area of the polygon R, and equals

$$\sum_1^n (\pi - \alpha_i) - (n-2)\pi = 2\pi - \sum_1^n \alpha_i$$

by the Girard formula. This explains that the Gaussian curvature is an inner quantity. Note that if more than three triangles come together at the given vertex then the triangulated surface can be locally bent without distortation, but the local contribution to the Gaussian curvature from the given vertex remains the same. Did Gauss discover the Theorema Egregium along these lines, by approximation of a smooth surface by a triangulated surface?

Thinking along these lines we can also explain a truely remarkable result, called the Gauss–Bonnet theorem.

Theorem 4.9. *Let $\mathcal{S}$ be a surface build out of Euclidean triangles, which is compact and has no boundary. Let v be the number of vertices, e the number of edges and f be the number of faces of the triangulation. Then the total Gaussian curvature of $\mathcal{S}$ is given by*

$$\int_{\mathcal{S}} K dA = 2\pi\chi(\mathcal{S})$$

with $\chi(\mathcal{S}) = (v - e + f) \in \mathbb{Z}$ the so called Euler characteristic of $\mathcal{S}$.

Proof. The expression $\int_{\mathcal{S}} K dA$ is called the total Gaussian curvature, and becomes in the triangulated approximation equal to

$$2\pi v - \sum_\alpha \alpha = 2\pi v - \pi f = 2\pi(v - e + f)$$

since $3f = 2e$ as each triangle is bounded by three edges, while each edge bounds exactly two triangles. ∎

The remarkable aspect of the Gauss–Bonnet theorem is that the integral Gaussian curvature on $\mathcal{S}$, which a priori is a real number depending on the chosen length element ds^2 of the surface, remains invariant if the surface is smoothly deformed. Indeed an integral multiple of 2π remains constant under smooth deformations. For example the total Gaussian curvature of the ellipsoids

$$x^2/a^2 + y^2/b^2 + z^2/c^2 = 1$$

is equal to 4π for any $a, b, c > 0$, since the area of the unit sphere with constant Gaussian curvature $K = 1$ is equal to 4π. We can also approximate the unit sphere by a tetrahedron, octahedron or icosahedron, and in each case $\chi = v - e + f = 2$ as known to Euler.

4.4 The upper and lower index notation

So far we have used the classical notation for surfaces in $\mathbb{R}^3$ with local coordinates u, v and coefficients E, F, G and L, M, N of the first and second fundamental forms respectively. This notation has been completely standard in the nineteenth century literature, when this theory was developed. It is still used in the more recent text books of the past century like Vorlesungen über Differentialgeometrie von Wilhelm Blaschke from 1945, Lectures on Classical Differential Geometry by Dirk Struik from 1950, Differential Geometry by James Stoker from 1969, and the more recent Elementary Differential Geometry by Andrew Pressley from 2010, which we also use as side book for this course.

In this section we adopt the following change of notation

$$u \mapsto u^1, v \mapsto u^2 \qquad \partial/\partial u \mapsto \partial_1, \partial/\partial v \mapsto \partial_2$$

with upper indices for the coordinates and lower indices for the partial derivatives. The coefficients of the first and second fundamental forms are changed into g_{ij} and h_{ij} respectively, with indices $i, j, k, \cdots = 1, 2$. So the surface in $\mathbb{R}^3$ is given locally by a smooth map

$$(u^1, u^2) \mapsto \mathbf{r} = \mathbf{r}(u^1, u^2)$$

with $\mathbf{N} = \partial_1\mathbf{r} \times \partial_2\mathbf{r}/|\partial_1\mathbf{r} \times \partial_2\mathbf{r}|$ the everywhere defined unit normal. So in the new notation

$$g_{ij} = \partial_i\mathbf{r} \cdot \partial_j\mathbf{r} \qquad h_{ij} = \partial_i\partial_j\mathbf{r} \cdot \mathbf{N}$$

for $i, j = 1, 2$. The first fundamental form

$$ds^2 = \sum g_{ij} du^i du^j$$

is also called the metric tensor. Whenever we write a sum symbol $\sum$ it is understood that the summation runs over those indices in the summand which appear both as upper and as lower index. In Einstein summation convention even the summation sign is left out, and whenever the same index appears as upper and lower index in some expression it is tacitly assumed that it is summed over. For example, we have

$$\boxed{\partial_i \partial_j \mathbf{r} = \sum \Gamma^k_{ij} \partial_k \mathbf{r} + h_{ij} \mathbf{N}}$$

for the definition of the Christoffel symbols Γ^k_{ij} and the second fundamental form tensor $\sum h_{ij} du^i du^j$. We write g^{ij} for the coefficients of the inverse matrix of g_{ij}, so for example $g^{12} = -g_{21}/(g_{11}g_{22} - g_{21}g_{12})$.

Theorem 4.10. *We have* $\Gamma^k_{ij} = \sum(\partial_i g_{jl} + \partial_j g_{il} - \partial_l g_{ij})g^{lk}/2$ *and so the Christoffel symbols are algebraic expressions in the coefficients of the metric tensor and their first order derivatives.*

Proof. Indeed, taking the scalar product of the boxed equation with $\partial_l \mathbf{r}$ gives

$$\partial_i \partial_j \mathbf{r} \cdot \partial_l \mathbf{r} = \sum \Gamma^k_{ij} g_{kl}$$

and since

$$(\partial_i g_{jl} + \partial_j g_{il} - \partial_l g_{ij})/2 = \partial_i \partial_j \mathbf{r} \cdot \partial_l \mathbf{r}$$

the formula for the Christoffel symbols follows from linear algebra. □

A comparison with the old notation used in Section 4.1 illustrates that the new notation is both compact and efficient. We can view the boxed equation as a second order system of *partial* differential equations for $\mathbf{r}$ as function of (u^1, u^2), and the theory of such equations gives that a smooth surface in $\mathbb{R}^3$ is uniquely determined by its first and second fundamental form coefficients, up to a proper Euclidean motion coming from the free choice of an initial point $\mathbf{r}(u^1_0, u^2_0)$ and initial tangent vectors $\partial_i \mathbf{r}(u^1_0, u^2_0)$ under the constraint $\partial_i \mathbf{r}(u^1_0, u^2_0) \cdot \partial_j \mathbf{r}(u^1_0, u^2_0) = g_{ij}(u^1_0, u^2_0)$. This is the fundamental theorem for smooth surfaces in $\mathbb{R}^3$.

The situation is analogous to that of smooth curves being determined by curvature $\kappa > 0$ and torsion τ. However in the former case we could prescribe *arbitrarily* smooth functions $\kappa > 0$ and τ of a parameter s and the local existence and uniqueness of a smooth arclength parametrized smooth curve with these curvature and torsion followed by local existence and uniqueness of the *ordinary* differential equation in question, the Frenet equation. However in the situation of surfaces we can not prescribe the coefficients g_{ij} (under the obvious constraints $g_{11} > 0, g_{11}g_{22} - g_{21}g_{12} > 0$) and h_{ij} arbitrarily.

Indeed for a solution of the boxed equation to exist we must necessarily have

$$\partial_i(\partial_j\partial_k\mathbf{r}) = \partial_j(\partial_i\partial_k\mathbf{r})$$

and so the expression

$$\partial_i\{\sum\Gamma^l_{jk}\partial_l\mathbf{r} + h_{jk}\mathbf{N}\}$$

is symmetric under $i \leftrightarrow j$. The theory of partial differential equations of boxed type gives that these so called *integrability conditions* also are sufficient for local existence and uniqueness. Hence the expression

$$\sum\partial_i\Gamma^l_{jk}\partial_l\mathbf{r} + \sum\Gamma^l_{jk}\Gamma^m_{il}\partial_m\mathbf{r} + \sum\Gamma^l_{jk}h_{il}\mathbf{N} + \partial_i h_{jk}\mathbf{N} + h_{jk}\partial_i\mathbf{N}$$

should be symmetric under $i \leftrightarrow j$. The normal component of these equations leads to the Codazzi–Mainardi equations, and the tangential components to the Gauss equations.

Since $\mathbf{N}\cdot\mathbf{N} = 1$ we get $\partial_i\mathbf{N}\cdot\mathbf{N} = 0$ and so the Codazzi–Mainardi equations simply become

$$\partial_i h_{jk} - \partial_j h_{ik} + \sum\{\Gamma^l_{jk}h_{il} - \Gamma^l_{ik}h_{jl}\} = 0$$

for all i, j, k. Likewise $\mathbf{N}\cdot\partial_n\mathbf{r} = 0$ implies that

$$\partial_i\mathbf{N}\cdot\partial_n\mathbf{r} = -\mathbf{N}\cdot\partial_i\partial_n\mathbf{r} = -h_{in}$$

and so $\partial_i\mathbf{N} = \sum n^l_i\partial_l\mathbf{r}$ with coefficients n^l_i given by $n^l_i = \sum -h_{in}g^{nl}$. These are the so called *Weingarten equations.* If we denote

$$R^l_{ijk} := \partial_i\Gamma^l_{jk} - \partial_j\Gamma^l_{ik} + \sum\{\Gamma^m_{jk}\Gamma^l_{im} - \Gamma^m_{ik}\Gamma^l_{jm}\}$$

for the coefficients of the *Riemann curvature tensor* then the Gauss equations take the form

$$R^l_{ijk} = \{h_{jk}h_{in} - h_{ik}h_{jn}\}g^{nl}$$

or equivalently

$$\{h_{jk}h_{il} - h_{ik}h_{jl}\} = \sum R^n_{ijk}g_{nl} =: R_{ijkl}$$

for all i, j, k, l. Since the Riemann curvature tensor coefficients R^l_{ijk} or R_{ijkl} are (admittedly rather complicated) algebraic expressions in the coefficients of the metric tensor and their first and second order partial derivatives we arrive at the Theorema Egregium

$$K := \frac{h_{11}h_{22} - h_{21}h_{12}}{g_{11}g_{22} - g_{21}g_{12}} = \frac{-R_{1212}}{g_{11}g_{22} - g_{21}g_{12}}$$

for $i = 1, j = 2, k = 1, l = 2$.

This ends our discussion of the Theorem Egregium as consequence of the integrability conditions for the fundamental theorem. Compared to the same explanations in Sections 4.1 and 4.2 the index notation used here is compact and more efficient. Another bonus is that in fact the story told here is valid for hypersurfaces of dimension n in Euclidean space $\mathbb{R}^{n+1}$.

5 Geodesics

5.1 The Geodesic Equations

Suppose $\mathcal{S}$ is a smooth surface given by the coordinates

$$(u, v) \mapsto \mathbf{r}(u, v)$$

with first fundamental form

$$ds^2 = Edu^2 + 2Fdudv + Gdv^2$$

given by the usual formulas

$$E = \mathbf{r}_u \cdot \mathbf{r}_u \ , \ F = \mathbf{r}_u \cdot \mathbf{r}_v \ , \ G = \mathbf{r}_v \cdot \mathbf{r}_v \ .$$

We recall the definition of a geodesic on $\mathcal{S}$ from Section 2.2.

Definition 5.1. *A curve $\boldsymbol{\gamma}(t) = \mathbf{r}(u(t), v(t))$ on $\mathcal{S}$ traversed in time t is called a geodesic if the acceleration $\ddot{\boldsymbol{\gamma}}$ is a multiple of the unit normal $\mathbf{N}$ for all time t.*

It is clear that geodesics are traversed with constant speed since

$$\frac{d}{dt}(\dot{\boldsymbol{\gamma}} \cdot \dot{\boldsymbol{\gamma}}) = 2\ddot{\boldsymbol{\gamma}} \cdot \dot{\boldsymbol{\gamma}} = 2\ddot{\boldsymbol{\gamma}} \cdot (\dot{u}\mathbf{r}_u + \dot{v}\mathbf{r}_v) = 0$$

because $\ddot{\boldsymbol{\gamma}}$ is a multiple of $\mathbf{N}$.

Theorem 5.2. *The curve $\boldsymbol{\gamma}(t) = \mathbf{r}(u(t), v(t))$ is a geodesic on $\mathcal{S}$ if and only if the so called geodesic equations*

$$\begin{aligned} \frac{d}{dt}(E\dot{u} + F\dot{v}) &= (E_u\dot{u}^2 + 2F_u\dot{u}\dot{v} + G_u\dot{v}^2)/2 \\ \frac{d}{dt}(F\dot{u} + G\dot{v}) &= (E_v\dot{u}^2 + 2F_v\dot{u}\dot{v} + G_v\dot{v}^2)/2 \end{aligned}$$

do hold.

Proof. By definition $\boldsymbol{\gamma}(t) = \mathbf{r}(u(t), v(t))$ is a geodesic on $\mathcal{S}$ if

$$\{\frac{d}{dt}(\dot{u}\mathbf{r}_u + \dot{v}\mathbf{r}_v)\} \cdot \mathbf{r}_u = 0 \ , \ \{\frac{d}{dt}(\dot{u}\mathbf{r}_u + \dot{v}\mathbf{r}_v)\} \cdot \mathbf{r}_v = 0 \ .$$

These equations can be rewritten as

$$\begin{aligned} \frac{d}{dt}\{(\dot{u}\mathbf{r}_u + \dot{v}\mathbf{r}_v) \cdot \mathbf{r}_u\} &= (\dot{u}\mathbf{r}_u + \dot{v}\mathbf{r}_v) \cdot (\dot{u}\mathbf{r}_{uu} + \dot{v}\mathbf{r}_{uv}) \\ \frac{d}{dt}\{(\dot{u}\mathbf{r}_u + \dot{v}\mathbf{r}_v) \cdot \mathbf{r}_v\} &= (\dot{u}\mathbf{r}_u + \dot{v}\mathbf{r}_v) \cdot (\dot{u}\mathbf{r}_{uv} + \dot{v}\mathbf{r}_{vv}) \end{aligned}$$

and using the relations (as seen before in the proof of Theorem 4.1)

$$\mathbf{r}_{uu} \cdot \mathbf{r}_u = E_u/2 \ , \ \mathbf{r}_{uv} \cdot \mathbf{r}_u = E_v/2 \ , \ \mathbf{r}_{uu} \cdot \mathbf{r}_v = F_u - E_v/2$$
$$\mathbf{r}_{uv} \cdot \mathbf{r}_v = G_u/2 \ , \ \mathbf{r}_{vv} \cdot \mathbf{r}_v = G_v/2 \ , \ \mathbf{r}_{vv} \cdot \mathbf{r}_u = F_v - G_u/2$$

the geodesic equations follow. □

We shall derive another form for the geodesic equations in the next theorem. Observe that the above geodesic equations can be rewritten as

$$\begin{aligned} E\ddot{u} + F\ddot{v} &= (-E_u\dot{u}^2 - 2E_v\dot{u}\dot{v} + (-2F_v + G_u)\dot{v}^2)/2 \\ F\ddot{u} + G\ddot{v} &= ((E_v - 2F_u)\dot{u}^2 - 2G_u\dot{u}\dot{v} - G_v\dot{v}^2)/2 \end{aligned}$$

and since $(EG - F^2) > 0$ one can solve $\ddot{u}$ and $\ddot{v}$ from these equations by direct linear algebra.

Theorem 5.3. *A curve $t \mapsto \mathbf{r}(u(t), v(t))$ is a geodesic on $\mathcal{S}$ if and only if the geodesic equations*

$$\ddot{u} + \Gamma^1_{11}\dot{u}^2 + 2\Gamma^1_{12}\dot{u}\dot{v} + \Gamma^1_{22}\dot{v}^2 = 0$$
$$\ddot{v} + \Gamma^2_{11}\dot{u}^2 + 2\Gamma^2_{12}\dot{u}\dot{v} + \Gamma^2_{22}\dot{v}^2 = 0$$

hold with Γ^k_{ij} the Christoffel symbols of Theorem 4.1.

Remark 5.4. *Let us write as before*

$$\mathbf{v} = \mathbf{r}_u\dot{u} + \mathbf{r}_v\dot{v} \ , \ \mathbf{a} = \mathbf{r}_u\ddot{u} + \mathbf{r}_v\ddot{v} + \mathbf{r}_{uu}\dot{u}^2 + 2\mathbf{r}_{uv}\dot{u}\dot{v} + \mathbf{r}_{vv}\dot{v}^2$$

for the velocity and acceleration of the curve $t \mapsto \mathbf{r}(u(t), v(t))$ on the surface $\mathcal{S}$ in $\mathbb{R}^3$. The orthogonal projection along $\mathbf{N}$ of the accelaration $\mathbf{a}$ on the tangent space spanned by $\mathbf{r}_u, \mathbf{r}_v$ is given by $\mathbf{a} - (\mathbf{a} \cdot \mathbf{N})\mathbf{N}$ and becomes

$$(\ddot{u} + \Gamma^1_{11}\dot{u}^2 + 2\Gamma^1_{12}\dot{u}\dot{v} + \Gamma^1_{22}\dot{v}^2)\mathbf{r}_u + (\ddot{v} + \Gamma^2_{11}\dot{u}^2 + 2\Gamma^2_{12}\dot{u}\dot{v} + \Gamma^2_{22}\dot{v}^2)\mathbf{r}_v$$

by the very definition of the Christoffel symbols. Hence the above theorem is clear indeed, and the proof given above was just a repetition of calculations in the proof of Theorem 4.1.

Meridians on a surface of revolution are the intersection of the surface with a plane through the axis of rotation. Therefore it is clear on geometric grounds that the meridians on a surface of revolution are always geodesics. For example for the unit sphere $\mathcal{S}^2$ all great circles are geodesics. But for general surfaces it is rare that geodesics can be computed explicitly. However two important conlusions about geodesics can be drawn at this point.

The first conclusion is that for each point on the surface and for each tangent vector at that point there exists locally a unique geodesic through that point with the given tangent vector as velocity. Indeed, the geodesic equations are a system of ordinary second order differential equations. Hence this follows from the general existence and uniqueness theorem for such differential equations.

The second conclusion is similar in spirit to the Theorema Egregium in the sense that geodesics on a surface only depend on the coefficients E, F, G of the first fundamental form and their first partial derivatives. Apparently the notion of geodesic is an inner quantity of the surface, despite the fact that the definition of geodesic requires the ambient Euclidean space $\mathbb{R}^3$ containing $\mathcal{S}$. In the next section we shall see that geodesics are those curves on the surface $\mathcal{S}$ for which the length along the curve between nearby points is minimal. This geometric characterization of geodesics is clearly inner, and therefore it should this time not come as a surprise that geodesics are an inner concept.

In terms of classical mechanics one can think of a geodesic as the orbit of a free particle on $\mathcal{S}$. Indeed Newton's law of motion $\mathbf{F} = m\ddot{\mathbf{r}}$ together with the geodesic equation $\ddot{\boldsymbol{\gamma}} \propto \mathbf{N}$ means that the tangential component of the force vanishes everywhere. From this point of view geodesics have constant speed as a consequence of conservation of Hamiltonian energy $H = (\dot{\boldsymbol{\gamma}} \cdot \dot{\boldsymbol{\gamma}})/2$.

5.2 Geodesic Parallel Coordinates

Let $U \subset \mathbb{R}^2$ be an open rectangle containing the origin $(0, 0)$ and let

$$U \ni (u, v) \mapsto \mathbf{r}(u, v) \in \mathbb{R}^3$$

be a smooth surface $\mathcal{S}$ with first fundamental form coefficients

$$E = \mathbf{r}_u \cdot \mathbf{r}_u \ , \ F = \mathbf{r}_u \cdot \mathbf{r}_v \ , \ G = \mathbf{r}_v \cdot \mathbf{r}_v$$

defining the length element $ds^2 = Edu^2 + 2Fdudv + Gdv^2$ on $\mathcal{S}$.

Theorem 5.5. *All curves $u \mapsto \mathbf{r}(u, v)$ with v being constant are unit speed geodesics which intersect the axis $\{(0, v); v \in \mathbb{R}\}$ everywhere in U in a perpendicular way if and only if the first fundamental form becomes*

$$ds^2 = du^2 + G(u, v)dv^2$$

on all of U.

Proof. The curves $u \mapsto \mathbf{r}(u,v)$ with v constant are all unit speed curves if and only if $E = 1$ on all of U, and they intersect the axis $\{(0,v); v \in \mathbb{R}\}$ at each point in U orthogonally if and only if $F(0,v) = 0$ for all v. If the curves $u \mapsto \mathbf{r}(u,v)$ with v constant are in addition geodesics then the second geodesic equation in Theorem 5.2 implies that $F_u = 0$ on all of U, and hence $F = 0$ on all of U. This shows that the first fundamental form becomes $ds^2 = du^2 + Gdv^2$ on all of U as required.

Conversely if $ds^2 = du^2 + Gdv^2$ then the coordinate curves intersect at each point orthogonally. The curves $u \mapsto \mathbf{r}(u,v)$ with v constant are unit speed curves, and by direct inspection solutions of both geodesic equations in Theorem 5.2. $\square$

The coordinates of the theorem are called geodesic parallel cordinates. Geodesic parallel coordinates exist nearby a freely prescribed smooth curve $t \mapsto \mathbf{r}(0,t)$ on $\mathcal{S}$. Because

$$\int_{t_1}^{t_2} \sqrt{1 + G(t, v(t))\dot{v}^2}\, dt \geq \int_{t_1}^{t_2} dt = t_2 - t_1$$

for all $t_2 > t_1$ the length between nearby points on a geodesic is minimal under small deformations of the curve keeping begin and end point fixed throughout the deformation. Therefore geodesics have a geometric meaning as locally length minimizing curves, and it should not be a surprise that their characterizing geodesic equations are inner, as mentioned in the previous section.

A familiar example of geodesic parallel coordinates are polar coordinates in the Euclidean plane, since the coordinate transformation

$$x = r\cos\theta \ , \ y = r\sin\theta$$

implies that $ds^2 = dx^2 + dy^2 = dr^2 + r^2 d\theta^2$ is of the desired form.

Using the geodesic equations in Theorem 5.2 it is easy to check that in geodesic parallel coordinates with first fundamental form

$$ds^2 = du^2 + G(u,v)dv^2$$

the curve $v \mapsto \mathbf{r}(0,v)$ is a unit speed geodesic as well if and only if

$$G(0,v) = 1 \ , \ G_u(0,v) = 0$$

for all $(0,v) \in U$. Hence the first fundamental form becomes

$$ds^2 = du^2 + dv^2 + O(u^2)$$

and so is Euclidean up to first order along the curve $u = 0$.

5.3 Geodesic Normal Coordinates

Let $\mathcal{S}$ be a smooth surface in $\mathbb{R}^3$ given by local coordinates $U \ni (u,v) \mapsto \mathbf{r}(u,v)$ with U an open disc around the origin $(0,0)$.

Definition 5.6. *These coordinates are called geodesic normal coordinates around the point* $\mathbf{r}(0,0)$ *of* $\mathcal{S}$ *if in polar coordiates*

$$u = r\cos\theta \ , \ v = r\sin\theta$$

the lines θ *is constant through* $(0,0)$ *become geodesics on* $\mathcal{S}$ *with arclength parameter* r*.*

It follows from the existence and uniqueness of geodesics through a given point with a given tangent vector at that point that around each point of $\mathcal{S}$ geodesic normal coordinates exist and are in fact unique up to the action of the orthogonal group $\mathrm{O}(2,\mathbb{R})$ in the coordinates (u,v). The next result is called Gauss' Lemma.

Lemma 5.7. *In geodesic normal coordinates the geodesic circles* r *equal to a positive constant intersect the central geodesics* θ *equal to a constant in a perpendicular way.*

Proof. We have $|\mathbf{r}_r(r,\theta)| = 1$ for $0 < r < \epsilon$ for some $\epsilon > 0$, and therefore

$$\int_0^\rho (\mathbf{r}_r \cdot \mathbf{r}_r)\, dr = \rho$$

for all $0 < \rho < \epsilon$. Differentiation of both sides with respect to θ gives

$$0 = \int_0^\rho (\mathbf{r}_r \cdot \mathbf{r}_{r\theta})\, dr = \int_0^\rho (\mathbf{r}_r \cdot \mathbf{r}_\theta)_r\, dr - \int_0^\rho (\mathbf{r}_{rr} \cdot \mathbf{r}_\theta)\, dr$$

for all $0 < \rho < \epsilon$. Since $r \mapsto \mathbf{r}(r,\theta)$ with θ constant is a geodesic with arclength r we conclude that $\mathbf{r}_{rr} \propto \mathbf{N}$ and so $\mathbf{r}_{rr} \cdot \mathbf{r}_\theta = 0$ for all $0 < r < \epsilon$. Because $\mathbf{r}(0,\theta)$ is constant we get

$$\mathbf{r}_r(\rho,\theta) \cdot \mathbf{r}_\theta(\rho,\theta) = \mathbf{r}_r(0,\theta) \cdot \mathbf{r}_\theta(0,\theta) = 0$$

for all $\rho > 0$ and $\theta \in \mathbb{R}/2\pi\mathbb{Z}$. $\square$

Remark 5.8. *In order to understand the line of thought of Riemann, as we shall discuss in the next chapter, it will be necessary to have an inner geometric proof of Gauss' Lemma. Such a proof goes as follows. If*

$$u = r\cos\theta \ , \ v = r\sin\theta$$

are geodesic normal coordinates then the line element becomes

$$ds^2 = Edu^2 + 2Fdudv + Gdv^2 = edr^2 + 2fdrd\theta + gd\theta^2$$

with

$$\begin{aligned} e &= E\cos^2\theta + 2F\cos\theta\sin\theta + G\sin^2\theta \\ f &= (-E\cos\theta\sin\theta + F(\cos^2\theta - \sin^2\theta) + G\sin\theta\cos\theta)r \\ g &= (E\sin^2\theta - F\sin\theta\cos\theta + G\cos^2\theta)r^2 \end{aligned}$$

as follows from $du = \cos\theta dr - r\sin\theta d\theta, dv = \sin\theta dr + r\cos\theta d\theta$. *Therefore* $f = O(r)$ *and* $g = O(r^2)$ *for* $r \downarrow 0$, *and so*

$$f(+0,\theta) = 0 \quad , \quad g(+0,\theta) = g_r(+0,\theta) = 0$$

for all θ. *Clearly* $e = 1$ *because the curves* θ *is constant are unit speed parametized for time* r. *In turn the second geodesic equation in Theorem 5.2 for the geodesics* θ *is constant with parameter* r *gives* $f_r = 0$ *for all small* $r > 0$, *and together with* $f(+0,\theta) = 0$ *we deduce* $f(r,\theta) = 0$ *for all* θ *and all small* $r > 0$. *Hence in geodesic normal coordinates*

$$ds^2 = dr^2 + g(r,\theta)d\theta^2$$

and Gauss' lemma follows.

The next result gives a normal form of the length element ds^2 in geodesic normal coordinates.

Theorem 5.9 (Riemann's formula)**.** *In geodesic normal coordinates* (u,v) *the length element* ds^2 *on* $\mathcal{S}$ *takes the form*

$$ds^2 = dr^2 + g(r,\theta)d\theta^2 = du^2 + dv^2 + H(u,v)(udv - vdu)^2$$

with $H(u,v) = (g(r,\theta) - u^2 - v^2)/(u^2+v^2)^2$ *a smooth function around the origin. The Gaussian curvature* K_0 *at the origin is given by* $K_0 = -3H(0,0)$ *and therefore* $K_0 = 0$ *if and only if the length element* ds^2 *is Euclidean up to second order.*

Proof. The first equality sign in the formula for ds^2 is a direct consequence of Theorem 5.5 together with Gauss' Lemma. Since $u = r\cos\theta$ and $v = r\sin\theta$ we find

$$r = \sqrt{u^2+v^2}\ ,\ \theta = \arctan\frac{v}{u}$$

which in turn implies

$$dr = \frac{udu + vdv}{\sqrt{u^2+v^2}} \ , \ d\theta = \frac{d(v/u)}{1+(v/u)^2} = \frac{udv - vdu}{u^2+v^2}$$

and since

$$dr^2 = \frac{u^2du^2 + 2uvdudv + v^2dv^2}{u^2+v^2} = du^2 + dv^2 - \frac{(udv - vdu)^2}{u^2+v^2}$$

the second equality in the formula for ds^2 follows by direct substitution.

In the geodesic normal coordinates (u, v) we have

$$E = 1 + H(u,v)v^2 \ , \ F = -H(u,v)uv \ , \ G = 1 + H(u,v)u^2$$

and

$$EG - F^2 = 1 + H(u,v)(u^2+v^2)$$

and so the Brioschi–Baltzer formula of Theorem 4.8 gives

$$K(u,v) = \frac{-E_{vv} + 2F_{uv} - G_{uu}}{2(EG - F^2)} + O(r^2) = -3H(0,0) + O(r)$$

for $r^2 = u^2 + v^2 \downarrow 0$. Hence $K_0 = -3H(0,0)$ and this completes the proof of the theorem. □

Theorem 5.10. *If $L(r)$ is the length of the circle $u^2 + v^2 = r^2$ and $A(r)$ the area of the disc $u^2 + v^2 \le r^2$ for $0 < r < \epsilon$ then*

$$K_0 = 6 \lim_{r\downarrow 0} \frac{2\pi r - L(r)}{2\pi r \cdot r^2} \ , \ K_0 = 12 \lim_{r\downarrow 0} \frac{\pi r^2 - A(r)}{\pi r^2 \cdot r^2}$$

with $K_0 = K(0,0)$ the Gaussian curvature at the origin.

Proof. We have $ds^2 = dr^2 + g(r,\theta)d\theta^2$ with

$$g(r,\theta) = r^2 + H(0,0)r^4 + O(r^5)$$

for $r \downarrow 0$ by Theorem 5.9. Hence the length $L(r)$ equals

$$\begin{aligned}\int_0^{2\pi} \sqrt{g(r,\theta)}\ d\theta &= \int_0^{2\pi} r[1 + H(0,0)r^2/2 + O(r^3)]\ d\theta \\ &= 2\pi r(1 + H(0,0)r^2/2) + O(r^4)\end{aligned}$$

while the area $A(r)$ becomes

$$\int_0^r \int_0^{2\pi} \sqrt{g(\rho,\theta)}\, d\rho d\theta = \int_0^r \int_0^{2\pi} \rho[1 + H(0,0)\rho^2/2 + O(\rho^3)]\, d\rho d\theta$$
$$= \pi r^2(1 + H(0,0)r^2/4) + O(r^5)$$

for $r \downarrow 0$. Since $H(0,0) = -K_0/3$ we conclude

$$L(r) = 2\pi r(1 - K_0 r^2/6) + O(r^4)\ ,\ A(r) = \pi r^2(1 - K_0 r^2/12) + O(r^5)$$

and the theorem follows. □

The first formula of this theorem was obtained by Bertrand and Puiseux, and the second formula is due to Diquet, both in the year 1848. Therefore a correct inner geometric approach to the Gaussian curvature can be achieved as follows. Define the Gaussian curvature by the Brioschi–Baltzer formula of Theorem 4.8. This is a rather complicated algebraic formula, and it is a priori highly unclear that the Gaussian curvature is defined independent of a choice of local coordinates. A direct proof of that fact will be a very tricky algebraic computation. If someone would be courageous to do these computations then the immediate next question would be the meaning of Gaussian curvature.

In geodesic normal coordinates (which have inner geometric meaning) the Brioschi–Baltzer formula simplifies, and this is what we used in the the proof of Theorem 5.9. The formulae of Bertrand–Puiseux and Diquet are derived from this theorem, and result in two inner geometric meanings of the Gaussian curvature. As a consequence the Brioschi–Baltzer formula for the Gaussian curvature is indeed independent of the choice of local coordinates. The Bertrand–Puiseux and Diquet formulae give two more proofs of the Theorema Egregium that the Gaussian curvature is of inner geometric origin. Altogether we have given six different arguments for the validity of the Theorema Egregium, four in Section 4.3 and another two in this section.

6 Surfaces of Constant Curvature

6.1 Riemannian surfaces

Locally a Riemannian surface is given by an open set U in the Euclidean plane $\mathbb{R}^2$ with coordinates u, v and three smooth functions $E, F, G : U \to \mathbb{R}$ such that $E > 0$ and $EG - F^2 > 0$ on all of U. But apart from these restrictions the functions E, F, G are *arbitrary* functions. The expression

$$ds^2 = Edu^2 + 2Fdudv + Gdv^2$$

is called the first fundamental form (old terminology of Monge and Gauss) or the Riemannian metric (modern terminology in honour of Riemann). The Riemannian metric enables one to measure the length of piecewise smooth curves in U. Indeed, if $[\alpha, \beta] \ni t \mapsto (u(t), v(t)) \in U$ is a piecewise smooth curve in U then the length of such a curve is defined by

$$\int_\alpha^\beta ds = \int_\alpha^\beta \sqrt{E\dot{u}^2 + 2F\dot{u}\dot{v} + G\dot{v}^2}\, dt$$

just as defined in Section 2.1. Of course, the motivating example remains that of a smooth surface $\mathcal{S}$ embedded in $\mathbb{R}^3$, with the Riemannian metric induced by the embedding.

In his famous Habilitation lecture, held on June 10 in 1854 in Göttingen and entitled "Über die Hypothesen, welche der Geometrie zu Grunde liegen" (On the Hypotheses, which underlie Geometry), Riemann took the above intrinsic approach as point of departure. The Riemannian metric enables one to measure length of curves, and thereby also to define the concept of geodesics via the geodesic equations, as we did in Section 5.1. Subsequently, one shows that geodesics are curves which minimize the lengths between nearby points on the geodesic, as we did in Section 5.2. Finally, on proves the existence of geodesic normal coordinates, as we did in Section 5.3. In geodesic normal coordinates around a given point Riemann showed that the Riemannian metric has the form

$$ds^2 = du^2 + dv^2 + H(u, v)(udv - vdu)^2$$

for some smooth function $H(u, v)$, say defined around the origin corresponding to the given point. Riemann then defines the Gauss curvature at the given point as $K_0 = -H(0, 0)/3$ in accordance with Theorem 5.9. The Riemannian metric in geodesic normal coordinates is Euclidean up to first order, and the second order deviation of the flat Euclidean metric $du^2 + dv^2$

is given by the Gauss curvature at the given point. Can it be explained in simpler geometric terms? Curvature is what measures the deviation from flatness!

6.2 The Riemann Disc

Let $\mathbb{S}_r$ be the sphere in $\mathbb{R}^3$ with equation $x^2 + y^2 + z^2 = r^2$ for $r > 0$. The stereographic projection from the north pole $(0, 0, r)$ assigns to a point (x, y, z) on $\mathbb{S}_r$ the intersection of the line through $(0, 0, r)$ and (x, y, z) with the plane $z = 0$. This point of intersection has coordinates $(u, v, 0)$ with

$$u = \frac{rx}{r - z} \ , \ v = \frac{ry}{r - z}$$

because $(x, y, z) = (1 - \lambda)(0, 0, r) + \lambda(u, v, 0)$ with $\lambda = (r - z)/r$. The stereographic projection is a smooth bijection from $\mathbb{S}_r - \{(0, 0, r)\}$ onto the plane $z = 0$ with coordinates (u, v) with inverse given by

$$x = \frac{2r^2u}{u^2 + v^2 + r^2} \ , \ y = \frac{2r^2v}{u^2 + v^2 + r^2} \ , \ z = \frac{r(u^2 + v^2 - r^2)}{u^2 + v^2 + r^2}$$

for $(u, v) \in \mathbb{R}^2$. A straightforward calculation gives

$$ds^2 = \frac{4(du^2 + dv^2)}{(1 + (u^2 + v^2)/r^2)^2}$$

for the length element of the sphere $\mathbb{S}_r$ in these coordinates.

Theorem 6.1. *For certain constants $m, n \in \mathbb{R}$ the intersection of $\mathbb{S}_r$ with the plane $z = mx + n$ is mapped onto the quadric curve with equation*

$$r(u^2 - 2rmu + v^2 - r^2) - n(u^2 + v^2 + r^2) = 0$$

which in turn implies that circles on $\mathbb{S}_r$ are stereographically projected onto circles and (in case $n = r$ if the plane goes through the north pole) lines in $\mathbb{R}^2$. For $n = 0$ the plane $z = mx$ intersects $\mathbb{S}_r$ in a great circle whose projection

$$(u - rm)^2 + v^2 = r^2(m^2 + 1)$$

for $m \neq 0$ intersects the equator $u^2+v^2 = r^2$ in two antipodal points $(0, \pm r)$.

The proof of the theorem is by direct substitution. By rotational symmetry around the third axis we see that great circles on $\mathbb{S}_r$ are stereographically projected onto circles and lines intersecting the equator $u^2+v^2 = r^2$ in a pair of antipodal points. The conclusion is that the plane $\mathbb{R}^2$ with coordinates (u, v) and length element

$$ds^2 = \frac{4(du^2 + dv^2)}{(1 + K(u^2 + v^2))^2}$$

is a geometric model for a surface with constant positive Gaussian curvature $K = 1/r^2 > 0$.

Motivated by these calculations Riemann suggested as geometric model for a surface with constant negative Gaussian curvature $K = -1/r^2 < 0$ the disc $\mathbb{D}_r = \{u^2 + v^2 < r^2\}$ with length element

$$ds^2 = \frac{4(du^2 + dv^2)}{(1 - (u^2 + v^2)/r^2)^2}$$

by the exact same formula as above. Just replace r^2 by $-r^2$ in the formula for the length element on the Riemann sphere $\mathbb{S}_r$. The restriction to the disc $\mathbb{D}_r$ is necessary because the length element blows up near the boundary of the disc. The disc $\mathbb{D}_r$ with the above length element is called the Riemann disc. By construction the Riemann disc is an abstract surface, not given as surface in $\mathbb{R}^3$, but just in coordinates with a prescribed length element.

What curves in the Riemann disc $\mathbb{D}_r$ do we expect as geodesics? The transition from $K = 1/r^2 > 0$ to $K = -1/r^2 < 0$ is made by the formal algebraic substitutions

$$r \mapsto \sqrt{-1}r\ ,\ z \mapsto \sqrt{-1}z\ , m \mapsto \sqrt{-1}m\ , n \mapsto \sqrt{-1}n$$

while x, y, u, v remain the same. The geodesic

$$(u - rm)^2 + v^2 = r^2(m^2 + 1)$$

in the spherical geometry with $K = 1/r^2$ goes via these substitutions over in the geodesic

$$(u + rm)^2 + v^2 = r^2(m^2 - 1)$$

in the hyperbolic geometry with $K = -1/r^2$. Apparently one should take $m^2 > 1$ and the geodesic with this equation becomes a circle intersecting the boundary $u^2 + v^2 = r^2$ of $\mathbb{D}_r$ in two points

$$(-r/m, \pm r\sqrt{1 - 1/m^2})$$

in a perpendicular way. This last claim follows from

$$r^2/m^2 + r^2(1 - 1/m^2) + r^2(m - 1/m)^2 + r^2(1 - 1/m^2) = r^2m^2$$

and the Pythagoras theorem. So we expect as geodesics for the Riemann disc $\mathbb{D}_r$ circular arcs perpendicular to the boundary of $\mathbb{D}_r$.

6.3 The Poincaré Upper Half Plane

In the previous section we have seen that the unit disc $\mathbb{D} = \{u^2 + v^2 < 1\}$ with length element

$$ds^2 = \frac{4(du^2 + dv^2)}{(1 - u^2 - v^2)^2}$$

is a model for geometry with constant Gauss curvature $K = -1$. In complex notation $w = u + iv$ and $z = x + iy$ the fractional linear transformations

$$z = -i\frac{w+1}{w-1}\ , \ w = \frac{z-i}{z+i}$$

interchange $\mathbb{D}$ with the upper half plane $\mathbb{H} = \{y > 0\}$. The derivative is given by

$$\frac{dw}{dz} = \frac{2i}{(z+i)^2}$$

which in turn implies that

$$\frac{4dwd\overline{w}}{(1 - w\overline{w})^2} = \frac{16dzd\overline{z}}{((z+i)(\overline{z+i}) - (z-i)(\overline{z-i}))^2}$$

and hence

$$ds^2 = \frac{4(du^2 + dv^2)}{(1 - u^2 - v^2)^2} = \frac{dx^2 + dy^2}{y^2}$$

for the length element on $\mathbb{H}$. The upper half plane $\mathbb{H}$ with this length element is called the Poincaré upper half plane.

Theorem 6.2. *The group* $\mathrm{PSL}_2(\mathbb{R})$ *acts on the upper half plane* $\mathbb{H}$ *by fractional linear transformations*

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} z = \frac{az+b}{cz+d}$$

and this action preserves the length element $ds^2 = (dx^2 + dy^2)/y^2$.

Proof. The derivative of the mapping $z \mapsto \zeta = (az+b)/(cz+d)$ is equal to $d\zeta/dz = 1/(cz+d)^2$ and therefore

$$\frac{d\zeta d\overline{\zeta}}{(\Im\zeta)^2} = \frac{dzd\overline{z}}{(\Im z)^2}$$

by a direct computation. □

In fact $\mathrm{PSL}_2(\mathbb{R})$ is the full group of orientation preserving isometries of the Poincaré upper half plane. The positive imaginary axis is a geodesic, since

$$\int \frac{\sqrt{\dot{x}^2+\dot{y}^2}}{y}dt \geq \int \frac{dy}{y}$$

and by the distance transitive action of $\mathrm{PSL}_2(\mathbb{R})$ it follows that all lines and circles perpendicular to the real axis as boundary of $\mathbb{H}$ are geodesics. Note that $t \mapsto ie^t$ is a unit speed geodesic, and so the length from i to the boundary at 0 or at $i\infty$ in infinite. This description of geodesics in the Poincaré upper half plane is a confirmation of our formal description of geodesics for the Riemann unit disc in the previous section.

6.4 The Beltrami Trumpet

Consider a surface of revolution in $\mathbb{R}^3$ given by

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

with f, g smooth functions of a real variable u and $v \in \mathbb{R}/2\pi\mathbb{Z}$. The formula for the Gaussian curvature

$$K = \frac{g'(f'g''-f''g')}{f((f')^2+(g')^2)}$$

was derived in Section 3.1. In case the profile curve is parametrized by arclength we have $(f')^2+(g')^2 = 1$ and also $f'f''+g'g'' = 0$. Hence the Gaussian curvature becomes $K = -f''/f$ by direct verification.

The surface of revolution has constant Gaussian curvature $K = 1$ if and only if $f''+f = 0$. Hence $f(u) = \cos u$ and $g(u) = \int \sqrt{1-\sin^2 u du} = \sin u$ and the surface of revolution is the unit sphere $x^2+y^2+z^2 = 1$ as expected.

Likewise a surface of revolution has constant Gaussian curvature $K = -1$ if and only if $f''-f = 0$. Hence we find

$$f(u) = e^u \ , \ g(u) = \int \sqrt{1-e^{2u}}du \ .$$

The integral can be evaluated by the substitution $e^u = \cos\theta$, and

$$g(u) = -\int \frac{\sin^2\theta}{\cos\theta}d\theta = \int \{\cos\theta - \frac{1}{\cos\theta}\}d\theta = \sin\theta - \tfrac{1}{2}\log(\frac{1+\sin\theta}{1-\sin\theta})$$

which leads to

$$g(u) = \sqrt{1-e^{2u}} - \log(e^{-u}+\sqrt{e^{-2u}-1})$$

with $u < 0$.

The profile curve for this surface of revolution has graph

$$z = \sqrt{1-x^2} - \log \frac{1+\sqrt{1-x^2}}{x}$$

for $0 < x < 1$ and is called the tractrix. This name refers to a remarkable geometric property of this curve. At each point $\mathbf{r} = (x_0, z_0)$ of the tractrix the tangent line $\mathcal{L}$ intersects the vertical axis in a point $\mathbf{q} = (0, z_1)$ with $|\mathbf{r} - \mathbf{q}| = 1$. Indeed by a direct calculation one finds

$$\frac{dz}{dx} = \frac{\sqrt{1-x^2}}{x}$$

for $0 < x < 1$. Hence the tangent line $\mathcal{L}$ at $\mathbf{r}$ has equation

$$z - z_0 = \frac{\sqrt{1-x_0^2}}{x_0}(x - x_0)$$

and meets the vertical axis in the point $\mathbf{q} = (0, z_1)$ with

$$z_1 - z_0 = -\sqrt{1-x_0^2}\,.$$

The square of the distance from $\mathbf{r}$ to $\mathbf{q}$ is equal to

$$x_0^2 + (z_1 - z_0)^2 = x_0^2 + 1 - x_0^2 = 1$$

and so the distance from $\mathbf{r}$ to $\mathbf{q}$ remains constant and equal to 1 as $\mathbf{r}$ moves along the tractrix. The surface obtained by revolution of the tractrix around the vertical axis was called the pseudosphere by Beltrami, who introduced this surface in 1869. The pseudosphere also goes under the name of the Beltrami trumpet.

The first fundamental form of the pseudosphere becomes

$$ds^2 = du^2 + e^{2u} dv^2$$

and with the substition $x = v, y = e^{-u}$ this becomes the familiar length element

$$ds^2 = \frac{dx^2 + dy^2}{y^2}$$

of $\mathbb{H}$ with $x \in \mathbb{R}/2\pi\mathbb{Z}$ and $y > 1$. The pseudosphere is therefore also the quotient space

$$\{z = x + iy; y > 1\}/2\pi\mathbb{Z}$$

for the isometric action of $2\pi\mathbb{Z}$ on $\mathbb{H}$ by horizontal translations. The length of the pseudosphere is infinite, while its area relative to the area element

$$dA = \frac{dxdy}{y^2}$$

is finite and in fact equal to 2π.

The question whether the full Poincaré upper half plane $\mathbb{H}$ can be realized as a closed surface in $\mathbb{R}^3$ remained open until Hilbert in 1909 proved the impossibility of such an isometric embedding in $\mathbb{R}^3$. For a proof of Hilbert's theorem we refer to the text book by Stoker.

It should be mentioned that the formal description by Riemann of the hyperbolic disc leads to an isometic embedding as the lower sheet of the hyperboloid with two sheets

$$x^2 + y^2 + 1 = z^2 \ , \ z < 0$$

inside $\mathbb{R}^{2,1}$ with coordinates via stereographic projection from $(0, 0, 1)$ onto the unit disc in the plane $z = 0$. Here $\mathbb{R}^{2,1}$ is the Lorentzian space with pseudolength element

$$ds^2 = dx^2 + dy^2 - dz^2$$

but such expressions only became the subject of study in the 20th century with the theory of special and general relativety.

It was shown by the French mathematicians Janet in 1926 and Cartan in 1927 that an arbitrary Riemannian space (U, ds^2) of dimension m can always be *locally* isometrically embedded in $\mathbb{R}^n$ with $n = m(m + 1)/2$ under the restriction that the coefficients of ds^2 and the embedding have convergent power series expansions. This number n is clearly the minimal possible as ds^2 encodes $m(m + 1)/2$ unknown functions. But it is unknown whether such an isometric embedding does exist in the smooth context. More than 150 years after the seminal lecture by Riemann it is still an open problem in the case of surfaces of dimension $m = 2$ whether a length element

$$ds^2 = Edu^2 + 2Fdudv + Gdv^2$$

with arbitrary *smooth* coefficients with $E, G > 0$ and $EG - F^2 > 0$ *locally* near a point of *vanishing* Gaussian curvature *of sufficiently high order* can always be isometrically realized as a smooth surface inside the Euclidean space $\mathbb{R}^3$ of dimension $n = 3$.

References

[1] Marcel Berger, A Panoramic View of Riemannian Geometry, Springer Verlag, Berlin, 2003.

[2] Manfredo do Carmo, Differential Geometry of Curves and Surfaces, Prentice-Hall, 1976.

[3] Manfredo do Carmo, Riemannian Geometry, Birkhäuser, Boston, 1992.

[4] Carl Friedrich Gauss, General Investigations on Curved Surfaces, 1827.

[5] David Hibert and Stephan Cohn-Vossen, Geometry and the Imagination, Chelsea, New York, 1952.

[6] Felix Klein, Vorlesungen über die Entwicklung der Mathematik im 19. Jahrhundert, Springer-Verlag, Berlin, 1926.

[7] Andrew Pressley, Elementary Differential Geometry, Springer Undergraduate Mathematics Series, Springer Verlag, London, 2010.

[8] Bernhard Riemann, Ueber die Hypothesen, welche der Geometrie zu Grunde liegen, Habilitationsvortrag Göttingen, 10 Juni 1854.

[9] James J. Stoker, Differential Geometry, Wiley Interscience, New York, 1969.

[10] Dirk J. Struik, Lectures on Classical Differential Geometry, Addison Wesley, 1950.

[11] Dirk J. Struik, A Source Book in Mathematics, 1200-1800, Harvard University Press, Cambridge, Massachusetts, 1969.

Index