Curves and Surfaces

Lecture Notes for Geometry 1

Henrik Schlichtkrull

Department of Mathematics
University of Copenhagen

Preface

The topic of these notes is *differential geometry*. Differential geometry is the study of geometrical objects using techniques of differential calculus, in particular differentiation of functions. The objects that will be studied here are curves and surfaces in two- and three-dimensional space, and they are primarily studied by means of parametrization. The main properties of these objects, which will be studied, are notions related to the shape. We will study tangents of curves and tangent spaces of surfaces, and the notion of curvature will be introduced. These notions are defined through differentiation of the parametrization, and they are related to first and second derivatives, respectively.

The notion of curvature is quite complicated for surfaces, and the study of this notion will take up a large part of the notes. The culmination is a famous theorem of Gauss, which shows that the so-called *Gauss curvature* of a surface can be calculated directly from quantities which can be measured on the surface itself, without any reference to the surrounding three dimensional space. This theorem has played a profound role in the development of more advanced differential geometry, which was initiated by Riemann.

The theory developed in these notes originates from mathematicians of the 18th and 19th centuries. Principal contributors were Euler (1707-1783), Monge (1746-1818) and Gauss (1777-1855), but the topic has much deeper roots, since it builds on the foundations laid by Euclid (325-265 BC).

In these notes a significant emphasis is placed on the interplay between intuitive geometry and exact mathematics. Ideas are explained by numerous illustrations, but they are also given rigorous proofs. It is my hope that the student of the text will perceive the importance of both viewpoints. The notes are adapted to an intensive course which runs over 7 weeks, so that each chapter corresponds approximately to one week of teaching.

The notes were written and used for the first time in 2005. The present version, intended for 2011, has been improved and corrected thanks to the suggestions of many students. Undoubtedly there are places where further revision would be desirable, and I will appreciate all comments and corrections. The drawings are made with Anders Thorup's program SPLINES, downloadable from http://www.math.ku.dk/∼thorup/splines/.

Henrik Schlichtkrull
March 2011

Contents

Chapter 1

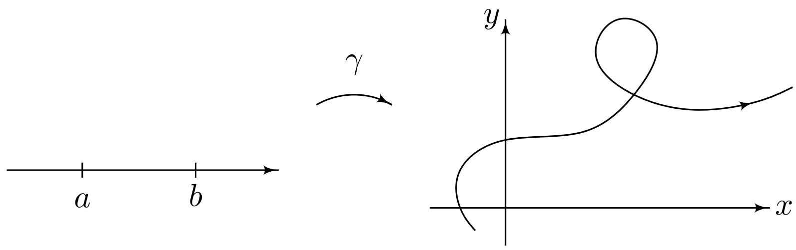
Parametrized curves and surfaces

In this chapter the basic concepts of curves and surfaces are introduced, and examples are given. These concepts will be described as subsets of $\mathbb{R}^2$ or $\mathbb{R}^3$ with a given parametrization, but also as subsets defined by equations. The connection from equations to parametrizations is drawn by means of the implicit function theorems (Theorems 1.5, 1.6 and 1.7).

1.1 Curves

It is well known from elementary geometry that a line in $\mathbb{R}^2$ or $\mathbb{R}^3$ can be described by means of a parametrization $t \mapsto p + tq$ where $q \neq 0$ and p are fixed vectors, and the parameter t runs over the real numbers. Likewise, a circle in $\mathbb{R}^2$ (say with center 0) can be parametrized by $t \mapsto (r\cos t, r\sin t)$ where $t \in \mathbb{R}$. The common nature of these examples is expressed in the following definition.

Definition 1.1. A *parametrized continuous curve* in $\mathbb{R}^n$ ($n = 2, 3, \dots$) is a continuous map $\gamma\colon I \to \mathbb{R}^n$, where $I \subset \mathbb{R}$ is an open interval (of end points $-\infty \leq a < b \leq \infty$).



The image set $\mathcal{C} = \gamma(I) \subset \mathbb{R}^n$ is called the *trace* of the curve. It is important to notice that we distinguish the curve and its trace. Physically, a curve describes the motion of a particle in n-space, and the trace is the trajectory of the particle. If the particle follows the same trajectory, but with different speed or direction, the curve is considered to be different.

For example, the positive x-axis is the trace of the parametrized curve $\gamma(t) = (t, 0)$ where $t \in I =]0, \infty[$, but it is also the trace of $\tilde{\gamma}(t) = (e^t, 0)$ with $t \in \mathbb{R}$.

Notice also that we do not require the parametrization to be injective. A point in the trace, which corresponds to more than one parameter value t, is called a *self-intersection* of the curve. For example, in the above parametrization of the circle, all points are self-intersections because values $t + 2\pi k$ correspond to the same point for all $k \in \mathbb{Z}$.

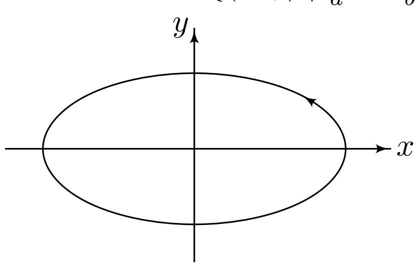
In these notes we will mainly be concerned with plane curves ($n = 2$) and space curves ($n = 3$), but in order to treat both cases simultaneously it is convenient not to specify n. We do not assume $n \leq 3$ for the time being, since it does not lead to any simplifications.

A parametrized continuous curve, for which the map $\gamma\colon I \to \mathbb{R}^n$ is differentiable up to all orders, is called a *parametrized smooth curve.* Recall that a map f into $\mathbb{R}^n$ is differentiable if each of its components $f_1, \dots, f_n$ is differentiable. The class of continuous curves is wide and the requirement of smoothness is a strong limitation. For example, the bizarre Peano curve, which is defined on $[0, 1]$ and has the entire unit square as trace, is continuous but not smooth. In these notes we will only study smooth curves, and we therefore adopt the convention that from now on *a parametrized curve is smooth, unless otherwise mentioned.*

We have already seen that lines and circles can be parametrized as smooth curves. Here are some further examples.

Example 1.1.1. The *constant curve* given by $\gamma(t) = p$, $t \in I$, where $p \in \mathbb{R}^n$ is fixed and I some open interval, is a parametrized curve.

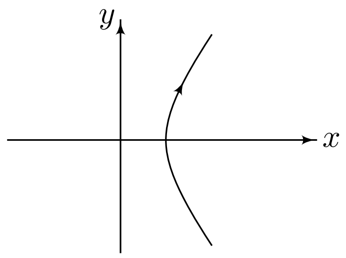
Example 1.1.2. The map $\gamma(t) = (a \cos t, b \sin t)$, where $a, b > 0$ are constants, parametrizes the *ellipse* $\mathcal{C} = \{(x, y) \mid \frac{x^2}{a^2} + \frac{y^2}{b^2} = 1\}$.



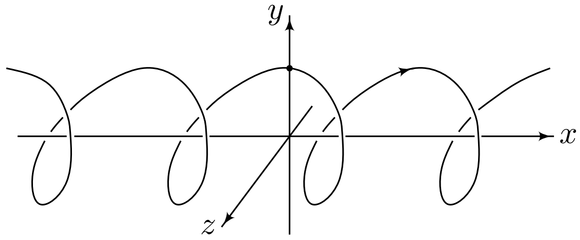
Example 1.1.3. Let $\gamma(t) = (a \cosh t, b \sinh t)$ where $a, b > 0$ and (see Appendix E)

$$\cosh t = \frac{e^t + e^{-t}}{2}, \quad \sinh t = \frac{e^t - e^{-t}}{2}.$$

Using the equation $\cosh^2 t - \sinh^2 t = 1$ we see that γ is a parametrization of the *hyperbola* (branch) $\mathcal{C} = \{(x, y) \mid \frac{x^2}{a^2} - \frac{y^2}{b^2} = 1, x > 0\}$.

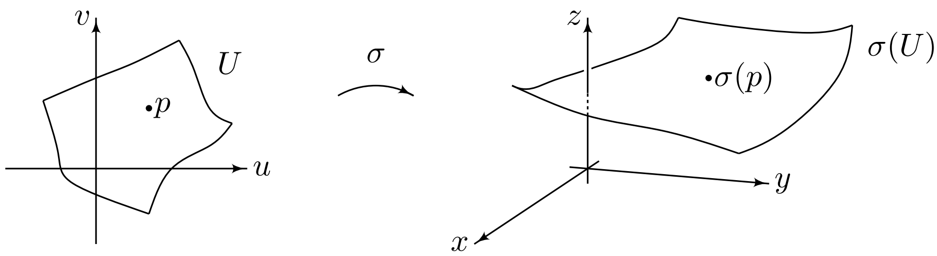


Example 1.1.4. The space curve $\gamma(t) = (\lambda t, r\cos(\omega t), r\sin(\omega t))$, where $r > 0$ and $\lambda, \omega \neq 0$ are constants, is called a *helix*. It is the spiraling motion of a point which moves along the x-axis with velocity λ while at the same time rotating around this axis with radius r and angular velocity ω.



1.2 Surfaces

Definition 1.2. A *parametrized continuous surface* in $\mathbb{R}^3$ is a continuous map $\sigma\colon U \to \mathbb{R}^3$, where $U \subset \mathbb{R}^2$ is an open, non-empty set.



It will often be convenient to consider the pair $(u, v) \in U$ as a set of *coordinates* of the point $\sigma(u, v)$ in the image $\mathcal{S} = \sigma(U)$. However, since σ is not assumed to be injective, the same point in $\mathcal{S}$ may have several pairs of coordinates.

We call a parametrized continuous surface *smooth* if the map $\sigma\colon U \to \mathbb{R}^3$ is smooth, that is, if the components σ_i, $i = 1, 2, 3$, of

$$\sigma(u, v) = (\sigma_1(u, v), \sigma_2(u, v), \sigma_3(u, v))$$

have continuous partial derivatives with respect to u and v, up to all orders. We adopt the convention that *a parametrized surface is smooth, unless otherwise mentioned.*

Example 1.2.1. Plane. Let $p, q_1, q_2 \in \mathbb{R}^3$ be fixed vectors and let

$$\sigma(u, v) = p + uq_1 + vq_2$$

for $(u, v) \in U = \mathbb{R}^2$. Then σ is a parametrized surface. If q_1, q_2 are linearly independent, the image $\sigma(U)$ is a plane. Otherwise it is a line or a point.

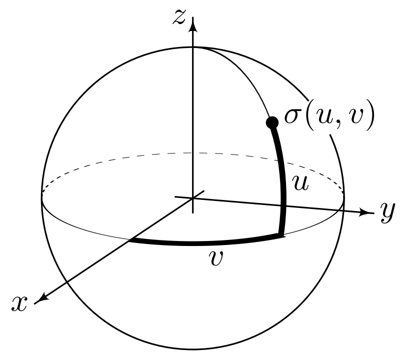
Example 1.2.2. Sphere. Let

$$\sigma(u,v) = (\cos u \cos v, \cos u \sin v, \sin u)$$

where $(u,v) \in \mathbb{R}^2$. This is a standard parametrization of the unit sphere

$$S^2 = \{(x,y,z) \in \mathbb{R}^3 \mid x^2 + y^2 + z^2 = 1\}.$$

The parameters u and v are called *latitude* and *longitude*, and together they are called *spherical coordinates.*

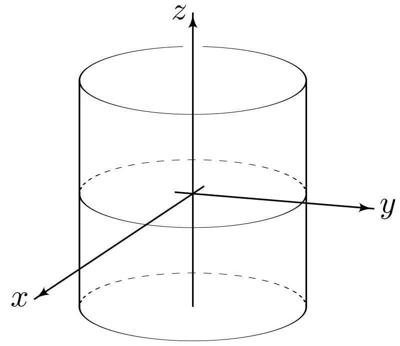


This parametrization covers the total sphere, but it is not injective. On the other hand, if we request, for example, that $u \in]-\frac{\pi}{2}, \frac{\pi}{2}[$ and $v \in]-\pi, \pi[$, then σ is injective, but it is not surjective, since a half-circle on the 'back' of the sphere will be outside the image of σ.

Example 1.2.3. Cylinder. Let $r > 0$ and put

$$\sigma(u,v) = (r \cos v, r \sin v, u)$$

where $(u,v) \in \mathbb{R}^2$. The image $\mathcal{S}$ of σ is the cylinder $\{(x,y,z) \mid x^2 + y^2 = r^2\}$ of radius r.

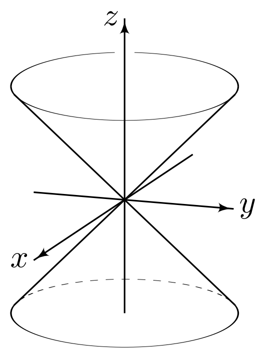


As before we have to restrict to a smaller set U if we want σ to be injective, for example by requiring v to belong in a fixed open interval of length 2π.

Example 1.2.4. Cone. Let $\lambda > 0$ and

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

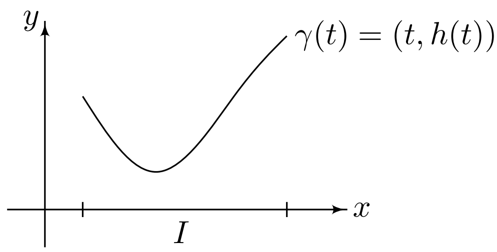
where $(u, v) \in \mathbb{R}^2$, then the image of σ is the cone $\{(x, y, z) \mid x^2+y^2 = \lambda^2 z^2\}$.



1.3 Graphs

By definition, the graph of a map $h: A \to B$, where A and B are arbitrary sets, is the set of all pairs $(x, h(x)) \in A \times B$, where $x \in A$.

Let $h: I \to \mathbb{R}$ be a smooth function, where $I \subset \mathbb{R}$ is an open interval. The map $t \mapsto (t, h(t))$ from I to $\mathbb{R}^2$ parametrizes the graph and makes it into a parametrized plane curve. We shall always regard the graph as being this parametrized curve.



Likewise we shall regard the graph of a smooth function $h: I \to \mathbb{R}^2$ as the parametrized curve $t \mapsto (t, h(t)) = (t, h_1(t), h_2(t))$ in $\mathbb{R}^3$.

Example 1.3.1. The graph of an affine linear function $h(t) = at+b$, $\mathbb{R} \to \mathbb{R}$ (where $a, b \in \mathbb{R}$), is the line in $\mathbb{R}^2$ parametrized by $(t, at + b)$. All lines which are not perpendicular to the x-axis can be parametrized in this fashion.

Similarly the graph of an affine linear function $h(t) = at + b$, $\mathbb{R} \to \mathbb{R}^2$ (where $a = (a_1, a_2), b = (b_1, b_2) \in \mathbb{R}^2$), is the line in $\mathbb{R}^3$ parametrized by $(t, a_1 t + b_1, a_2 t + b_2)$. All lines of direction not perpendicular to the x-axis can be parametrized in this fashion.

Example 1.3.2. The helix in Example 1.1.4 with $\lambda = 1$ is the 3-dimensional graph of the map $h: \mathbb{R} \to \mathbb{R}^2$ defined by $h(t) = (r \cos(\omega t), r \sin(\omega t))$.

We shall also consider surfaces, which are graphs. If $h\colon U \to \mathbb{R}$ is a smooth function defined on an open set $U \subset \mathbb{R}^2$, then the graph of h is the set

$$\{(u, v, h(u, v)) \mid (u, v) \in U\} \subset \mathbb{R}^3.$$

Equipped with the map

$$\sigma(u, v) = (u, v, h(u, v)), \quad (u, v) \in U,$$

the graph is clearly a parametrized smooth surface.

Example 1.3.3. The graph of an affine linear function $\mathbb{R}^2 \to \mathbb{R}$ is a plane in $\mathbb{R}^3$. Say $h(u, v) = au+bv+c$ where $a, b, c \in \mathbb{R}$, then $\sigma(u, v) = (u, v, au+bv+c)$. All planes, except those which are perpendicular to the xy-plane, can be parametrized in this fashion.

Example 1.3.6. The graph of the function $h(u, v) = \sqrt{1 - u^2 - v^2}$, defined on the unit disk $U = \{(u, v) \in \mathbb{R}^2 \mid u^2 + v^2 < 1\}$ is a half-sphere.

1.4 Level sets

Very often a plane curve is described, not by means of a parametrization, but by an equation. For example, a line is represented by an equation of the form $ax+by = c$ with $a, b, c \in \mathbb{R}$ and $(a, b) \neq (0, 0)$, and a circle is represented by an equation of the form $(x - x_0)^2 + (y - y_0)^2 = r^2$ with $r > 0$.

Similarly a surface can be described by an equation. For example, a plane in $\mathbb{R}^3$ is the set of solutions to an equation $ax + by + cz = d$, where $(a, b, c) \neq (0, 0, 0)$, and a sphere is represented by $(x - x_0)^2 + (y - y_0)^2 + (z - z_0)^2 = r^2$.

We shall now give a general definition which covers both situations.

Definition 1.4.1. Let $\Omega \subset \mathbb{R}^n$ be open and let $f\colon \Omega \to \mathbb{R}$ be a continuous function. The *level sets* for f are the sets

$$\mathcal{C} = \{x \in \Omega \mid f(x) = c\}$$

of solutions in Ω to the equation $f(x) = c$, where $c \in \mathbb{R}$ is a fixed constant.

In this course the function f will be assumed to be smooth. However, the smoothness alone does not ensure that the level sets for f can be parametrized as smooth curves or surfaces (in case $n = 2$ or 3). For example, a level set for the trivial function $f = 0$ on $\mathbb{R}^2$, that is, the set of solutions to an equation $0 = c$, is either the empty set or the full set $\mathbb{R}^2$. Some extra condition will be needed on f in order that the set is a curve.

Definition 1.4.2. Let $f\colon \Omega \to \mathbb{R}$ be smooth, where $\Omega \subset \mathbb{R}^n$ is open. A point $p \in \Omega$ is called *critical* if

$$\frac{\partial f}{\partial x_1}(p) = \cdots = \frac{\partial f}{\partial x_n}(p) = 0.$$

Let us consider some examples in the plane case $n = 2$. It will be seen in all the examples that if we exclude critical points, the level sets can be parametrized as curves. A precise statement to this effect is given in the corollary in Section 1.5.

Example 1.4.1. Consider the linear equation $ax + by = c$ whose solutions comprise a level set for $f(x, y) = ax + by$. If $(a, b) \neq (0, 0)$ then there are no critical points. In this case the set of solutions form a line, hence can be parametrized as a curve. On the other hand, if $(a, b) = (0, 0)$ then $f(x, y) = ax + by$ is the trivial function and all points are critical.

Example 1.4.2. Let $f(x, y) = x^2 + y^2$, then $\frac{\partial f}{\partial x} = 2x$ and $\frac{\partial f}{\partial y} = 2y$, so $(0, 0)$ is the only critical point. The level sets for $c > 0$ contain no critical points. They are circles, hence can be parametrized as smooth curves. The level set for $c = 0$ consists only of the critical point $(0, 0)$ and it is exactly in this case the circles degenerate to a point.

Example 1.4.3. Consider the equation $f(x, y) = xy = 0$. Here $\partial f/\partial x = y$ and $\partial f/\partial y = x$, and hence the origin $(0, 0)$ is the only critical point. In fact, the level set given by $f(x, y) = 0$ is the union of the two axes, which exactly fails to be a 'reasonable' curve at the origin.

1.5 The implicit function theorem, two variables

The implicit function theorem describes conditions under which a given equation in two variables can be solved to obtain one of the variables as a function of the other variable. For some simple equations, for example $y^2 - 2xy + 1 = 0$, explicit solutions are easily obtained by simple algebra, here $y = x + \sqrt{x^2 - 1}$ and $y = x - \sqrt{x^2 - 1}$, but for other equations such explicit solutions cannot be derived. The reason for this need not be lack of algebraic skill on our side, since a solution may not exist at all. The implicit function theorem expresses a simple condition which guarantees the existence of a function h, such that the solution is $y = h(x)$.

Theorem 1.5. *Let $f\colon \Omega \to \mathbb{R}$ be a smooth function, where $\Omega \subset \mathbb{R}^2$ is open. Let*

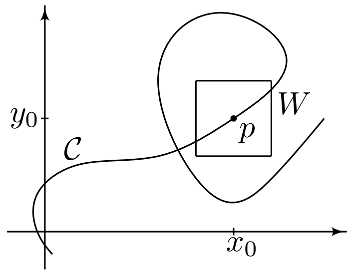
$$\mathcal{C} = \{(x, y) \in \Omega \mid f(x, y) = c\}$$

be the set of solutions to the equation $f(x, y) = c$. Let $p = (x_0, y_0) \in \mathcal{C}$ be given, and assume that $\frac{\partial f}{\partial y} \neq 0$ at p.

Then there exist open intervals I and J around x_0 and y_0, respectively, such that the rectangle $W = I \times J$ is contained in Ω, and a smooth map $h\colon I \to J$ such that

$$\mathcal{C} \cap W = \{(x, h(x)) \mid x \in I\}, \tag{1}$$

that is, in the neighborhood W of p, $\mathcal{C}$ is the graph of h.

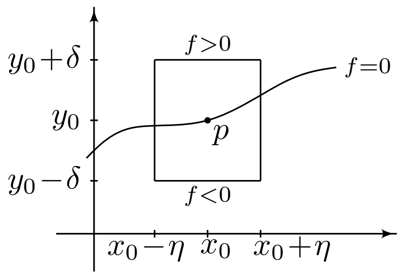


Proof. Assume for simplicity that $c = 0$, and that the value of $\frac{\partial f}{\partial y}$ at p is positive. These properties can be arranged by a simple replacement of f which does not affect the set $\mathcal{C}$. Choose $\delta > 0$ such that the neighborhood $\{(x, y) \mid |x - x_0| \leq \delta, |y - y_0| \leq \delta\}$ of p lies inside Ω, and such that $\frac{\partial f}{\partial y} \geq a$ on this neighborhood, for some constant $a > 0$ (continuity of $\frac{\partial f}{\partial y}$ is used). Then $y \mapsto f(x, y)$ is strictly increasing on the interval $[y_0 - \delta, y_0 + \delta]$, for each fixed x with $|x - x_0| < \delta$.

In particular, since $p \in \mathcal{C}$ we have $f(p) = f(x_0, y_0) = 0$, and hence

$$f(x_0, y_0 - \delta) < 0 \quad \text{and} \quad f(x_0, y_0 + \delta) > 0.$$

By continuity in x_0 of each of the maps $x \mapsto f(x, y_0 \pm \delta)$, there exists a positive number $\eta \leq \delta$ such that $f(x, y_0 - \delta) < 0$ and $f(x, y_0 + \delta) > 0$ for all x with $|x - x_0| < \eta$.



Let $I = \{x \mid |x - x_0| < \eta\}$, and let $x \in I$. Since $y \mapsto f(x, y)$ is strictly increasing and continuous, and since $f(x, y_0 - \delta) < 0$ and $f(x, y_0 + \delta) > 0$, there exists a unique y between $y_0 - \delta$ and $y_0 + \delta$ with $f(x, y) = 0$. This value of y is denoted $h(x)$. Then h maps I into $J =]y_0 - \delta, y_0 + \delta[$ and satisfies $f(x, h(x)) = 0$. The identity of the sets in (1) follows from the uniqueness of y. We will complete the proof of the theorem by showing that h is smooth.

We first prove that h is continuous. Fix $x \in I$ and let $y = h(x)$, then $f(x, y) = 0$. Let Δx be sufficiently small so that $x + \Delta x \in I$. Associated to Δx we define Δy such that $y + \Delta y = h(x + \Delta x)$, then also $f(x + \Delta x, y + \Delta y) = 0$.

The asserted continuity amounts to the statement that $\Delta y \to 0$ when $\Delta x \to 0$. The function

$$t \mapsto \varphi(t) = f(x + t\Delta x, y + t\Delta y)$$

is zero both for $t = 0$ and $t = 1$. By the mean value theorem (Rolle's theorem) there exists a number $\theta \in (0, 1)$ (depending on Δx) such that

$$\varphi'(\theta) = 0.$$

Differentiating φ by means of the chain rule we thus obtain

$$\frac{\partial f}{\partial x}(x + \theta\Delta x, y + \theta\Delta y)\Delta x + \frac{\partial f}{\partial y}(x + \theta\Delta x, y + \theta\Delta y)\Delta y = 0.$$

Hence

$$\Delta y = -\frac{\frac{\partial f}{\partial x}(x + \theta\Delta x, y + \theta\Delta y)}{\frac{\partial f}{\partial y}(x + \theta\Delta x, y + \theta\Delta y)}\Delta x,$$

and since $|\frac{\partial f}{\partial x}|$ is bounded, and $\frac{\partial f}{\partial y} \geq a > 0$, it follows that $\Delta y \to 0$ when $\Delta x \to 0$, as claimed.

Next we prove that h is differentiable, which with the notation from above amounts to the convergence of $\Delta y/\Delta x$ as $\Delta x \to 0$. In fact, since $\frac{\partial f}{\partial x}$ and $\frac{\partial f}{\partial y}$ are continuous, this follows immediately from the equation above. Moreover, the limit is given by

$$\lim_{\Delta x \to 0} \frac{\Delta y}{\Delta x} = -\frac{\frac{\partial f}{\partial x}(x, y)}{\frac{\partial f}{\partial y}(x, y)}.$$

Hence h is differentiable and satisfies

$$h'(x) = -\frac{\frac{\partial f}{\partial x}(x, h(x))}{\frac{\partial f}{\partial y}(x, h(x))}. \tag{2}$$

Finally, we prove by induction that h is smooth. Assuming that h is r times differentiable for some natural number r, we see from equation (2) that so is h'. Hence h is $r + 1$ times differentiable. □

Corollary 1.5. *Let $f: \Omega \to \mathbb{R}$ be a smooth function, where $\Omega \subset \mathbb{R}^2$ is open. Let*

$$\mathcal{C} = \{(x, y) \in \Omega \mid f(x, y) = c\}$$

and let $p = (x_0, y_0) \in \mathcal{C}$. Assume that p is not a critical point.

Then there exists an open rectangle $W \subset \Omega$ around p, such that $\mathcal{C} \cap W$ is the graph of a smooth function h, considered either as $y = h(x)$ or as $x = h(y)$.

In particular, it follows that the level set can be parametrized as a smooth curve in a neighborhood of each non-critical point.

Proof. By assumption $\frac{\partial f}{\partial x}$ and $\frac{\partial f}{\partial y}$ are not both zero at p. If $\frac{\partial f}{\partial y}(p) \neq 0$ the conclusion is already in the previous theorem. Otherwise, we interchange x and y. □

Example 1.5.1. Let $f:\mathbb{R}^2 \to \mathbb{R}$ be given by $f(x,y) = y^2 - 2xy + 1$, and consider the level set $\mathcal{C} = \{(x,y) \mid f(x,y) = 0\}$. Then $\frac{\partial f}{\partial y} = 2y - 2x$, which is zero if and only if $x = y$. Inserting $y = x$ in the equation $y^2 - 2xy + 1 = 0$, we see that the only points in $\mathcal{C}$ where $\frac{\partial f}{\partial y} = 0$ are $p = (1,1)$ and $q = (-1,-1)$.

We can then conclude from the theorem that the level set $\mathcal{C}$ can be attained as a graph of the form $y = h(x)$ in a neighborhood of each of its points, except possibly p and q (the theorem gives no information in case $\frac{\partial f}{\partial y} = 0$).

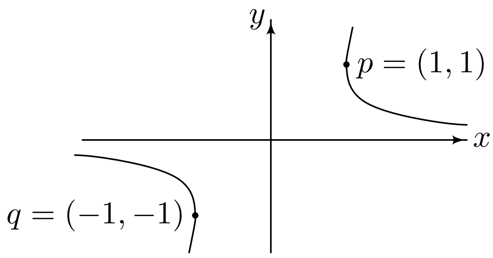
On the other hand, the partial derivative $\frac{\partial f}{\partial x} = -2y$ is never zero on $\mathcal{C}$ (since $y = 0$ in $y^2 - 2xy + 1 = 0$ leads to a contradiction), and thus $\mathcal{C}$ has the form of a graph $x = h(y)$ in a neighborhood of all its points.

In fact, the equation can be easily solved with respect to both x and y:

$$y = x \pm \sqrt{x^2 - 1}, \qquad x = \frac{1}{2}(y + \frac{1}{y}).$$

The formula on the left gives two expressions, each with y a function of x. Only one of these is relevant in a neighborhood of a given point $(x,y) \in \mathcal{C}$, provided $|x| > 1$. However, at the points where $x = \pm 1$, these two expressions for y collapse, and neither of them gives a well-defined function in a neighborhood, because $|x| < 1$ is not allowed in the square root. Notice that these are exactly the two points where $\frac{\partial f}{\partial y} = 0$.

The expression for x, on the other hand, is defined and smooth for all $y \neq 0$ (and $y = 0$ never occurs in $\mathcal{C}$).



Example 1.5.2. Let $f(x,y) = x^4 - x^2 + y^2$, and $\mathcal{C} = \{(x,y) \mid f(x,y) = 0\}$. The derivatives

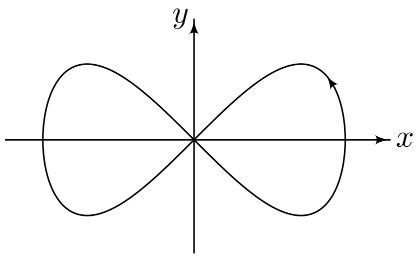
$$\frac{\partial f}{\partial x} = 4x^3 - 2x \quad \text{and} \quad \frac{\partial f}{\partial y} = 2y$$

are both zero if and only if (x,y) is one of the three points

$$(0,0), \quad (\sqrt{\tfrac{1}{2}}, 0), \quad (-\sqrt{\tfrac{1}{2}}, 0).$$

Only the first one of these belongs to $\mathcal{C}$, and this point is therefore the only critical point in $\mathcal{C}$.

The set $\mathcal{C}$ is shown in the following figure. It can be shown that $\mathcal{C}$ is the trace of the parametrized curve $\gamma(t) = (\cos t, \cos t \sin t)$, which has a self-intersection exactly in the critical point $(0, 0)$.



1.6 The implicit function theorem, more variables

We will now consider the analogue for surfaces of the theory of the preceding section. Where the solution set for an equation in two variables was described as a parametrized curve, the analogous theorem describes the solution set for an equation in three variables as a parametrized surface.

In fact, it is convenient to state the theorem as a theorem treating an equation in n variables, with arbitrary $n \geq 2$. In this fashion the theorem becomes a generalization rather than an analogue.

In order to compare easily with the previous theorem we denote our variables in $\mathbb{R}^n$ by $(x, y) = (x_1, \dots, x_{n-1}, y)$ where $x \in \mathbb{R}^{n-1}$ and $y \in \mathbb{R}$. By definition, an interval in $\mathbb{R}^n$ is a product $I_1 \times \cdots \times I_n$ of intervals in $\mathbb{R}$.

Theorem 1.6. *Let $f\colon \Omega \to \mathbb{R}$ be a smooth function, where $\Omega \subset \mathbb{R}^n$ is open. Let*

$$\mathcal{S} = \{(x, y) \in \Omega \mid f(x, y) = c\}$$

and let $p = (x^0, y^0) \in \mathcal{S}$. Assume that $\frac{\partial f}{\partial y} \neq 0$ at p.

Then there exist open intervals $I \subset \mathbb{R}^{n-1}$ and $J \subset \mathbb{R}$ around x^0 and y^0, respectively, such that the interval $W = I \times J$ is contained in Ω, and a smooth map $h\colon I \to J$ such that

$$\mathcal{S} \cap W = \{(x, h(x)) \mid x \in I\},$$

that is, in the neighborhood W of p, $\mathcal{S}$ is the graph of h.

Proof. Notice the similarity with Theorem 1.5, the only difference being that $x \in \mathbb{R}$ is replaced by $(x_1, \dots, x_{n-1}) \in \mathbb{R}^{n-1}$. In fact, the proof is a rather straightforward generalization of the proof of Theorem 1.5, and it is therefore omitted. $\square$

Let us take $n = 3$, and replace the notation (x_1, x_2, y) by (x, y, z). We obtain the following result which is analogous to the corollary in Section 1.5.

Corollary 1.6. *Let $f\colon \Omega \to \mathbb{R}$ be a smooth function, where $\Omega \subset \mathbb{R}^3$ is open. Let*

$$\mathcal{S} = \{(x, y, z) \in \Omega \mid f(x, y, z) = c\}$$

and let $p = (x_0, y_0, z_0) \in \mathcal{S}$. Assume that p is not a critical point.

Then there exist an open interval $W \subset \Omega$ around p, such that $\mathcal{S} \cap W$ is the graph of a smooth function h, considered either as $z = h(x, y)$, as $y = h(x, z)$ or as $x = h(y, z)$.

In particular, it follows that the level set can be parametrized as a smooth surface in a neighborhood of each non-critical point.

Proof. By assumption at least one of the partial derivatives $\frac{\partial f}{\partial x}$, $\frac{\partial f}{\partial y}$ and $\frac{\partial f}{\partial z}$ is not zero at p. Interchanging z with x or y if necessary, we may assume that it is $\frac{\partial f}{\partial z}$. The conclusion then follows from the previous theorem. $\square$

Example 1.6.1. The equation for a plane in $\mathbb{R}^3$, $ax + by + cz = d$, where $(a, b, c) \neq (0, 0, 0)$, satisfies the assumption of the preceding corollary. If $c \neq 0$, the plane is the graph $z = h(x, y)$ of the function $h(x, y) = (d - ax - by)/c$. If $c = 0$ the plane is vertical, and we cannot exhibit it as a graph of the form $z = h(x, y)$, but then we can exhibit it as a graph over one of the other coordinate planes.

Example 1.6.2. The surfaces introduced in Examples 1.2.2-1.2.4, sphere, cylinder and cone, are level sets for, respectively, $f(x, y, z) = x^2 + y^2 + z^2$, $f(x, y, z) = x^2 + y^2$ and $f(x, y, z) = x^2 + y^2 - z^2$. These functions all satisfy the assumption of the corollary above, except for the vertex $(0, 0, 0)$ of the cone.

Example 1.6.3. Let $f(x, y, z) = ze^x + yz$ and consider the equation $f(x, y, z) = 1$ in a neighborhood of the point $p = (0, 0, 1)$ (which solves the equation). The partial derivative $\frac{\partial f}{\partial z} = e^x + y$ is 1 at this point, so by the implicit function theorem the solution z exists as a function of (x, y) in a neighborhood of $(0, 0, 1)$. In fact, the equation has the solution $z = 1/(e^x + y)$. However, if the equation is replaced by for example $f(x, y, z) = ze^x + \sin(yz) = 1$, then it is impossible to write down a solution to the equation in terms of known functions, but the conclusion from the implicit function theorem remains the same since we still have $\frac{\partial f}{\partial z} = 1$ at p.

1.7 The implicit function theorem, more equations

We have seen that an equation $f(x, y) = c$ in $\mathbb{R}^2$ defines a plane curve, and that an equation $f(x, y, z) = c$ in $\mathbb{R}^3$ defines a surface (under suitable circumstances). In order to define a curve in $\mathbb{R}^3$ we need two equations. For example, the x-axis is the set defined by equations $y = 0$ and $z = 0$. Consider two equations of the form

$$f_1(x, y, z) = c_1,$$
$$f_2(x, y, z) = c_2,$$

where f_1, f_2 maps an open set $\Omega \subset \mathbb{R}^3$ into $\mathbb{R}$. We say that we have two equations in three variables. It is actually more convenient to write the equations in the form

$$f(x, y, z) = c$$

where $f = (f_1, f_2)$ maps Ω into $\mathbb{R}^2$, and where $c = (c_1, c_2)$. We want to generalize Theorem 1.6 in order to deal with this situation.

In fact, we will generalize even further, to functions $f\colon \Omega \subset \mathbb{R}^n \to \mathbb{R}^m$, that is, to the case of m equations in n variables, the only condition being that $n > m$. For the application to space curves, only $n = 3$ and $m = 2$ is needed, and the reader is encouraged to specialize to this case at first reading.

It is convenient for the comparison with Theorem 1.6 to write elements in $\mathbb{R}^n$ as (x, y) where $x = (x_1, \dots, x_{n-m}) \in \mathbb{R}^{n-m}$ and $y = (y_1, \dots, y_m) \in \mathbb{R}^m$. The goal is to obtain y as a function of x.

Theorem 1.7. *Let $f\colon \Omega \to \mathbb{R}^m$ be a smooth function, where $\Omega \subset \mathbb{R}^n$ is open. Let $c \in \mathbb{R}^m$ be fixed. Let*

$$\mathcal{C} = \{(x, y) \in \Omega \mid f(x, y) = c\}$$

and let $p = (x^0, y^0) \in \mathcal{C}$. Assume that the determinant of the $m \times m$ matrix

$$A = \frac{\partial f_i}{\partial y_j}(p),$$

consisting of the last m columns of the Jacobian $Df(p)$, is non-zero.

Then there exist open intervals $I \subset \mathbb{R}^{n-m}$ and $J \subset \mathbb{R}^m$ around x^0 and y^0, respectively, such that $W = I \times J \subset \Omega$, and a smooth map $h\colon I \to J$ such that

$$\mathcal{C} \cap W = \{(x, h(x)) \mid x \in I\},$$

that is, in the neighborhood W of p, $\mathcal{C}$ is the graph of $h\colon \mathbb{R}^{n-m} \to \mathbb{R}^m$.

Proof. The most common technique for solving several equations in several variables is elimination of variables. That is, we use one of the equations to express a particular variable in terms of the others, and insert this expression in the remaining equations. The chosen variable has then been eliminated, and the number of equations is reduced by one. This will also be our strategy in the present proof.

We prove the theorem by induction on m. The case $m = 1$ was already treated in Theorem 1.6. Thus, we assume that $m \geq 2$ and that the conclusion of the theorem is valid for functions into $\mathbb{R}^{m-1}$. We can safely assume that $c = 0$, since this can be arranged by subtraction of the constant from f.

Since $\det A$ is non-zero, A is invertible. We want to replace $f\colon \mathbb{R}^n \to \mathbb{R}^m$ by the function $A^{-1} \circ f\colon \mathbb{R}^n \to \mathbb{R}^m$, obtained by multiplying all image vectors $f(x, y) \in \mathbb{R}^m$ with the constant matrix A^{-1}. Since multiplication by A^{-1}

is a bijection, the equations $f(x,y) = 0$ and $A^{-1}f(x,y) = 0$ are equivalent. The Jacobian of the linear map, multiplication by A^{-1}, is the matrix A^{-1} itself (see Example B.1), and hence it follows from the chain rule that

$$D(A^{-1} \circ f)(p) = A^{-1} \cdot Df(p).$$

We see that the last m columns of $D(A^{-1} \circ f)(p)$ comprise a unit matrix δ_{kj}. By the replacement of f with $A^{-1} \circ f$ we thus obtain a function whose Jacobian matrix at p has a unit matrix in its last m columns, and for which the solution set $\mathcal{C}$ is unaltered. From now on we assume this replacement has been carried out, that is, we assume $\partial f_k/\partial y_j = \delta_{kj}$.

In particular, for the function f_m whose derivatives are in the last row of Df, we have that $\partial f_m/\partial y_m(p) = 1$. We will apply Theorem 1.6 to the equation $f_m(x,y) = 0$. The effect of the theorem is that the last variable, y_m, can be written as a smooth function of the remaining variables. We write the remaining variables as $(x, y') \in \mathbb{R}^{n-1}$ where $y' = (y_1, \dots, y_{m-1})$. More precisely, it then follows that there exists an interval neighborhood $W_1 = I_1 \times J_1$ around p, where $I_1 \subset \mathbb{R}^{n-1}$ and $J_1 \subset \mathbb{R}$, and a smooth function $h\colon I_1 \to J_1$ such that for $(x,y) = (x, y', y_m) \in W_1$ we have $f_m(x,y) = 0$ if and only if

$$y_m = h(x, y').$$

Let the function $F\colon I_1 \subset \mathbb{R}^{n-1} \to \mathbb{R}^{m-1}$ be defined by

$$F_k(x, y') = f_k(x, y', h(x, y')) \tag{3}$$

for $k = 1, \dots, m-1$, where as before $y' = (y_1, \dots, y_{m-1})$. The partial derivatives of F_k are obtained by applying the chain rule to (3):

$$\frac{\partial F_k}{\partial y_j} = \frac{\partial f_k}{\partial y_j} + \frac{\partial f_k}{\partial y_m}\frac{\partial h}{\partial y_j} \qquad (j = 1, \dots, m-1),$$

and at p we thus have $\frac{\partial F_k}{\partial y_j} = \frac{\partial f_k}{\partial y_j} = \delta_{kj}$ (because $\frac{\partial f_k}{\partial y_m} = 0$). The determinant of this matrix being non-zero, we can apply our induction hypothesis to F, and we obtain the existence of an interval neighborhood $W_2 = I_2 \times J_2$ around $(x^0, y^{0\prime})$, where $I_2 \subset \mathbb{R}^{n-m}$ and $J_2 \subset \mathbb{R}^{m-1}$, and a smooth function $g\colon I_2 \to J_2$ such that the solution set for the equation $F(x, y') = 0$ in W_2 is the graph of g, that is, $F(x, y') = 0$ if and only if $y' = g(x)$.

Let the interval $W \subset \mathbb{R}^n$ be defined by

$$W = W_1 \cap \{(x,y) \mid (x, y') \in W_2\}.$$

We now see that for (x,y) in this set we have

$$(x,y) \in \mathcal{C}$$

if and only if

$$f_k(x,y) = 0, \quad k = 1,\dots,m$$

if and only if

$$f_k(x,y) = 0, \quad k = 1,\dots,m-1 \quad \text{and} \quad y_m = h(x,y')$$

if and only if

$$F(x,y') = 0 \quad \text{and} \quad y_m = h(x,y')$$

if and only if

$$y' = g(x) \quad \text{and} \quad y_m = h(x,y')$$

if and only if

$$y = (g(x), h(x, g(x))).$$

The function $x \mapsto (g(x), h(x, g(x)))$ is thus seen to be the desired function whose graph is $\mathcal{C}$ in a neighborhood of p. $\square$

Corollary 1.7. *Let $f\colon \Omega \to \mathbb{R}^2$ be smooth, where $\Omega \subset \mathbb{R}^3$ is open. Let $c \in \mathbb{R}^2$ and*

$$p \in \mathcal{C} = \{(x,y,z) \in \Omega \mid f(x,y,z) = c\}.$$

Assume the rows of $Df(p)$ (a 2×3 matrix) are linearly independent.

Then there exist an open interval $W \subset \Omega$ around p, such that $\mathcal{C} \cap W$ can be parametrized as a smooth curve in the form of a graph, considered either as $(y,z) = h(x)$, as $(x,z) = h(y)$ or as $(x,y) = h(z)$.

Proof. At least one of the three 2×2 submatrices of $Df(p)$ has non-zero determinant. With suitable reorganization of variables the theorem can be applied. $\square$

Example 1.7. The set of equations

$$x + xy + z^2 = 3 \quad \wedge \quad x^3 + 2xz - y^2z^2 = 2 \tag{4}$$

has the form $f(x,y,z) = c$ with

$$f(x,y,z) = \begin{pmatrix} x + xy + z^2 \\ x^3 + 2xz - y^2z^2 \end{pmatrix}, \qquad c = \begin{pmatrix} 3 \\ 2 \end{pmatrix}.$$

The Jacobian matrix is

$$Df(x,y,z) = \begin{pmatrix} 1+y & x & 2z \\ 3x^2 + 2z & -2yz^2 & 2x - 2y^2z \end{pmatrix}.$$

In the point $(1,1,1)$ the equations (4) are satisfied and the Jacobian is

$$\begin{pmatrix} 2 & 1 & 2 \\ 5 & -2 & 0 \end{pmatrix}.$$

The determinant of the last two columns is $\det A = 4$. Since this determinant is not zero, the implicit function theorem assures that the equations can be solved for (y, z) as function of x, in a neighborhood of (1,1,1). In this neighborhood the set of solutions can thus be parametrized as a smooth curve in the form of a 3-dimensional graph $(t, h(t))$ where $h(t) = (y(t), z(t)) \in \mathbb{R}^2$.

This example demonstrates the theoretical power of the implicit function theorem, since the explicit solving of (4) for y and z as functions of x is clearly a difficult task.

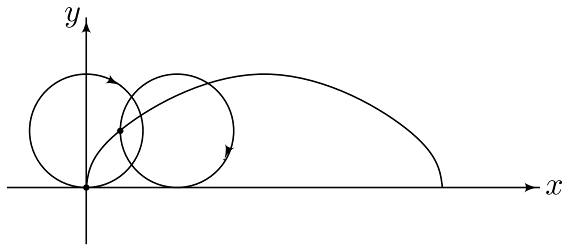
1.8 Exercises

1 Let $\gamma: I \to \mathbb{R}^n$ be a parametrized curve with $\gamma''(t) = 0$ for all t. Show that it is a line or a constant curve.

2 The following parametrized curve is called the *cycloid*

$$\gamma(t) = (t - \sin t, 1 - \cos t), \qquad (t \in \mathbb{R}).$$

It is constructed by a circle of radius 1 rolling without slipping along the positive x-axis. The curve is the path of a point on the circumference of the circle. Explain the formula above from this construction.



3 Determine a parametrized surface with image $\{(x, y, z) \mid x + 2y - 2z = 1\}$.

4 Consider the equation $x^3 + xy^2 - 2ay^2 = 0$ in $\mathbb{R}^2$, where $a > 0$ is a constant. Show that the parametrized curve

$$\gamma(t) = \left(\frac{2at^2}{1+t^2}, \frac{2at^3}{1+t^2}\right),$$

where $t \in \mathbb{R}$, is bijective onto the set of solutions. This curve is called the *cissoid of Diocles*. Draw a sketch of it (say for $a = 1$).

5 Let $\mathcal{S} \subset \mathbb{R}^3$ denote the set of solutions to the equation $x^2 + y - z^2 = 1$.

a. Show that the map $\sigma: \mathbb{R}^2 \to \mathbb{R}^3$ given by

$$\sigma(u, v) = (u + v, 1 - 4uv, u - v)$$

is an injective parametrized surface with image $\mathcal{S}$.

b. Show that $\mathcal{S}$ contains no critical points for the function $f(x, y, z) = x^2 + y - z^2$, and determine a parametrization of $\mathcal{S}$ as a graph of a smooth function h.

6 Consider the equation $x^3y^3 - 3x + y = -1$, which is satisfied by $(x, y) = (1, 1)$. Show that it is possible to describe the level set as a graph $y = h(x)$ in a neighborhood of this point.

7 Denote by $\mathcal{C}$ the level set in $\mathbb{R}^2$ for the equation $4x^4 - 5x^2y^2 + y^4 = 0$.

a. Show by means of the implicit function theorem that for each point $(x_0, y_0) \in \mathcal{C} \setminus \{(0, 0)\}$ there exists a neighborhood in which $\mathcal{C}$ can be described as a graph $y = h(x)$.

b. Solve the equation and determine $\mathcal{C}$ explicitly.

8 The condition $f'_y(x_0, y_0) \neq 0$ in Theorem 1.5 is sufficient but not necessary. That is, if $f'_y(x_0, y_0) = 0$ it may still be possible to describe the level set as a graph $y = h(x)$ of a smooth function h in a neighborhood of (x_0, y_0). Give an example.

9 Let f be as in Theorem 1.5, but assume instead that $f'_y(x_0, y_0) = 0$. Show that if $f'_x(x_0, y_0) \neq 0$, then it is *not* possible to describe the level set as a graph $y = h(x)$, where h is smooth, in any neighborhood of (x_0, y_0).

10 The equation $xy + xz + \sin z = 0$ is solved by $(x, y, z) = (0, 0, 0)$. Show that the solution set in a neighborhood of this point allows a description as a graph. Show that this is the case in a neighborhood of all solutions.

11 Let $\mathcal{S} = \{(x, y, h(x, y))\} \subset \mathbb{R}^3$ be the graph of the function $h(x, y) = y - xy^3$. Show that in a neighborhood of each point of $\mathcal{S}$ in which $3xy^2 \neq 1$, it is possible to write $\mathcal{S}$ as a graph of the form $y = g(x, z)$ with g smooth.

12 Consider the system of equations in $\mathbb{R}^3$

$$2x^2 - x^2z^2 - y^2 = 0 \quad \wedge \quad xyz = 1$$

to which $(1, 1, 1)$ is a solution. Show that there exists a neighborhood in which the solution set can be described as a graph of the form $(x, y) = h(z)$, where x and y both are functions of z.

13 Let $f\colon \mathbb{R}^3 \to \mathbb{R}^2$ and $c \in \mathbb{R}^2$ be given by

$$f(x, y, z) = (x^2 + y^2 + z^2, (x - \frac{1}{2})^2 + y^2), \qquad c = (1, a^2)$$

where $a \geq 0$. Let $L \subset \mathbb{R}^3$ denote the set of solutions to the system $f(x, y, z) = c$.

a. Explain why L is the intersection of a sphere and a cylinder, and determine their radii.

b. Determine, in each of the following 6 cases, the set of points in L for which the rank of $Df(p)$ is < 2, that is, where the rows are linearly dependent.

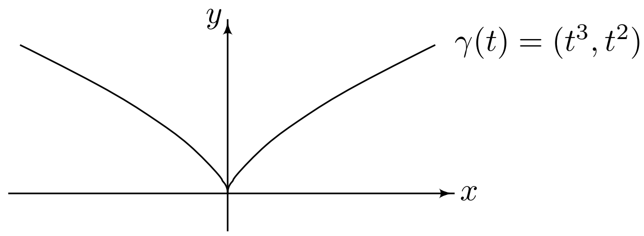
$$a = 0, \quad 0 < a < \tfrac{1}{2}, \quad a = \tfrac{1}{2}, \quad \tfrac{1}{2} < a < \tfrac{3}{2}, \quad a = \tfrac{3}{2}, \quad \tfrac{3}{2} < a.$$

c. What does the implicit function theorem tell about L in each case. Explain by means of the observation in a.

Chapter 2

Tangents

We have equipped parametrized curves and surfaces with the standing assumption that the parametrization is smooth. However, smoothness alone is not enough to ensure a simple geometric appearance. For example, the plane curve $\gamma(t) = (t^3, t^2)$ is perfectly smooth, but in $\gamma(0) = (0, 0)$ the trace of the curve has a sharp fold (a so-called 'cusp'), which conflicts with the intuitive notion of smooth. Another striking example will be given in Example 2.1.4 below.



In this chapter we will define a notion of regularity for parametrized curves and surfaces, which is motivated by the desire to exclude anomalies as the one just mentioned. The geometric significance of the regularity condition will be that it allows us to define notions of tangent lines and tangent planes.

2.1 Regular curves and tangent lines

Let $\gamma: I \to \mathbb{R}^n$ be a parametrized curve and let $t_0 \in I$ be given.

Definition 2.1.1. The curve γ is called *regular* in t_0 if $\gamma'(t_0) \neq 0$. Otherwise it is called *singular*. If γ is regular in all points of I we call it a *regular parametrized curve* or just a *regular curve*.

For example, the plane curve mentioned above is regular for $t \neq 0$ but it is singular at $t = 0$. The standard parametrizations of line and sphere (see Section 1.1), and the curves described in Examples 1.1.2, 1.1.3 and 1.1.4 are all regular curves.

Example 2.1.1 A constant curve (Example 1.1.1) is everywhere singular. Conversely, an everywhere singular parametrized curve is constant, since $\gamma' = 0$ implies that γ is constant.

Example 2.1.2 A graph $\gamma(t) = (t, h(t))$ (Section 1.3) is a regular curve in $\mathbb{R}^2$, since $\gamma'(t) = (1, h'(t)) \neq (0,0)$ (also if $h'(t) = 0$). Hence, by Corollary 1.5, a level set $f(x,y) = c$ can be parametrized as a regular curve in a neighborhood of each point which is not critical.

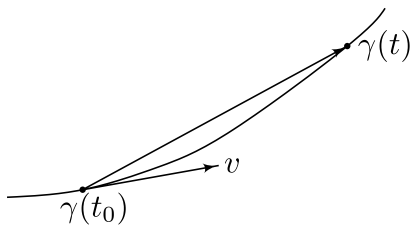
Definition 2.1.2. The vector $\gamma'(t_0)$ is called the *tangent vector* to γ at t_0. If γ is regular at t_0, the line through $p = \gamma(t_0)$ with direction $\gamma'(t_0)$ is called the *tangent line* of the curve.

The latter definition is motivated by the following result, which describes the tangent vector geometrically. The notation $\|v\|$ for vectors $v \in \mathbb{R}^n$ is defined in Appendix A.

Theorem 2.1. *Assume that γ is regular at t_0, and let $v = \gamma'(t_0)/\|\gamma'(t_0)\|$ be the unit vector in the direction of the tangent vector. Then*

$$v = \lim_{t \to t_0^+} \frac{\gamma(t) - \gamma(t_0)}{\|\gamma(t) - \gamma(t_0)\|} = \lim_{t \to t_0^-} \frac{\gamma(t_0) - \gamma(t)}{\|\gamma(t_0) - \gamma(t)\|}. \tag{1}$$

In other words, the unit tangent vector v is the limit position of the direction from $\gamma(t_0)$ to $\gamma(t)$, as t approaches t_0 from above, and the limit position of the direction from $\gamma(t)$ to $\gamma(t_0)$, as t approaches t_0 from below.



Proof. By definition

$$\gamma'(t_0) = \lim_{t \to t_0} \frac{\gamma(t) - \gamma(t_0)}{t - t_0}.$$

In particular, since $\gamma'(t_0) \neq 0$ it follows that $\gamma(t) \neq \gamma(t_0)$ for all $t \in I$ sufficiently close to (but different from) t_0. Thus the denominator of the fraction in (1) is not zero. Moreover for $t > t_0$,

$$\frac{\gamma(t) - \gamma(t_0)}{\|\gamma(t) - \gamma(t_0)\|} = \frac{1}{\|\frac{\gamma(t) - \gamma(t_0)}{t - t_0}\|} \frac{\gamma(t) - \gamma(t_0)}{t - t_0} \to \frac{1}{\|\gamma'(t_0)\|} \gamma'(t_0),$$

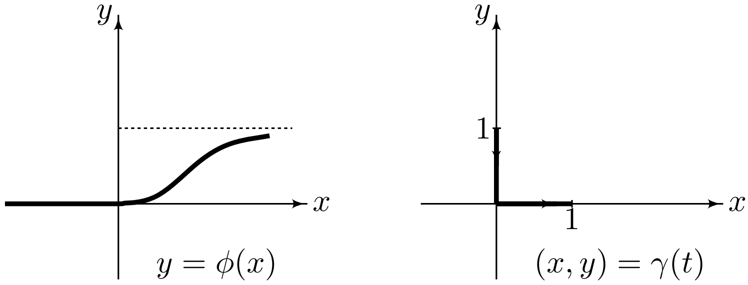
and similarly for $t < t_0$. $\square$

Example 2.1.3 According to the theorem, regularity of γ is a sufficient condition for (1) to hold. It is not a necessary condition. For example, the curve $\gamma(t) = (t^3, 0)$ which has the x-axis as its trace, is singular at $t_0 = 0$, but nevertheless both limits in (1) exist and are equal to the unit vector $v = (1, 0)$.

Example 2.1.4 A sophisticated example of a non-regular point on a smooth curve can be constructed as follows. Let $\phi: \mathbb{R} \to \mathbb{R}$ be the function defined by

$$\phi(t) = \begin{cases} \exp(-\frac{1}{t}) & \text{if } t > 0 \\ 0 & \text{otherwise} \end{cases}$$

then it can be shown that ϕ is smooth (the derivatives up to all orders vanish at 0). The graph of ϕ is shown to the left in the figure below.



Define $\gamma(t) = (\phi(t), \phi(-t))$ for $t \in \mathbb{R}$. This is a smooth curve whose trace consists of the line segment from 1 to 0 on the y-axis followed by the line segment from 0 to 1 on the x-axes. It is not regular at the origin, which is in accordance with the sharp turn of the curve in that point.

2.2 The tangent line of a level set

We have seen in Example 2.1.2 that the level set given by $f(x, y) = c$ can be parametrized as a regular curve $\gamma(t)$ in a neighborhood of each non-critical point p. We will determine the tangent line of such a parametrization.

Theorem 2.2. *Let $\mathcal{C} \subset \mathbb{R}^2$ be a level set of a smooth function f, and let $p = (x_0, y_0) \in \mathcal{C}$ be non-critical. Let $\gamma: I \to \mathbb{R}^2$ be any parametrized curve with trace $\gamma(I) \subset \mathcal{C}$ and with $\gamma(t_0) = p$ for some $t_0 \in I$, in which γ is regular. Then the tangent line of γ at t_0 is characterized by the equation*

$$\frac{\partial f}{\partial x}(p)(x - x_0) + \frac{\partial f}{\partial y}(p)(y - y_0) = 0.$$

Proof. We shall be using the following simple fact from plane geometry. The line with normal vector $(a, b) \neq (0, 0)$ through (x_0, y_0) is given by the equation

$$a(x - x_0) + b(y - y_0) = 0.$$

We thus have to prove that the tangent line has $(\frac{\partial f}{\partial x}(p), \frac{\partial f}{\partial y}(p))$, which is non-zero by assumption, as a normal vector.

Write $\gamma(t)$ in coordinates as $\gamma(t) = (x(t), y(t))$. Since γ maps into the level set we have $f(x(t), y(t)) = c$ for all t. By differentiation with the chain rule we obtain

$$x'(t_0)\frac{\partial f}{\partial x}(p) + y'(t_0)\frac{\partial f}{\partial y}(p) = 0,$$

which exactly shows that $(\frac{\partial f}{\partial x}(p), \frac{\partial f}{\partial y}(p))$ is perpendicular to the tangent vector $\gamma'(t_0) = (x'(t_0), y'(t_0))$. $\square$

Notice that it follows from the theorem that the tangent line depends on the level set through the function f, but it is independent of the chosen parametrization γ.

2.3 The tangent plane of a regular surface

Let $\sigma\colon U \to \mathbb{R}^3$ be a parametrized surface with image $\mathcal{S} = \sigma(U)$, and let a point $p = (u_0, v_0) \in U$ be given.

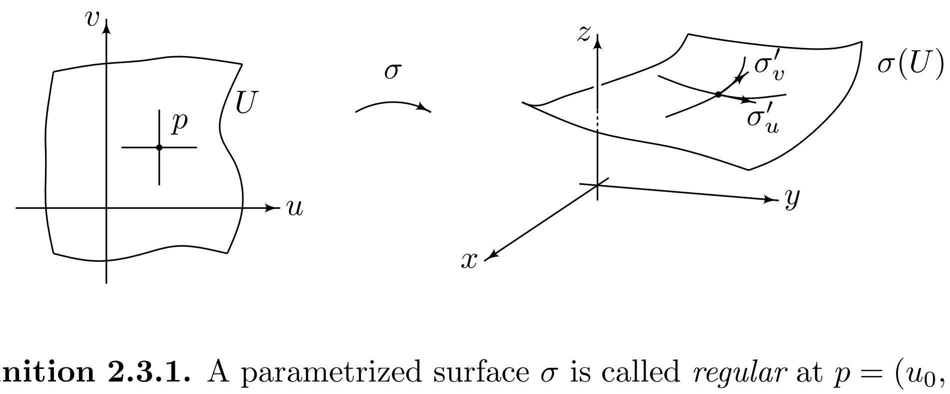
The notion of regularity for a parametrized surface is somewhat more complicated than that for a curve, because of the fact that we can differentiate with respect to both u and v. Let $\sigma = (\sigma_1, \sigma_2, \sigma_3)$ and put

$$\sigma'_u = \begin{pmatrix} \frac{\partial \sigma_1}{\partial u} \\ \frac{\partial \sigma_2}{\partial u} \\ \frac{\partial \sigma_3}{\partial u} \end{pmatrix} \quad \text{and} \quad \sigma'_v = \begin{pmatrix} \frac{\partial \sigma_1}{\partial v} \\ \frac{\partial \sigma_2}{\partial v} \\ \frac{\partial \sigma_3}{\partial v} \end{pmatrix}.$$

These vectors are the columns in the Jacobi matrix

$$D\sigma = \begin{pmatrix} \frac{\partial \sigma_1}{\partial u} & \frac{\partial \sigma_1}{\partial v} \\ \frac{\partial \sigma_2}{\partial u} & \frac{\partial \sigma_2}{\partial v} \\ \frac{\partial \sigma_3}{\partial u} & \frac{\partial \sigma_3}{\partial v} \end{pmatrix}.$$

Notice that $\sigma'_u(p)$ and $\sigma'_v(p)$ are the tangent vectors at $t = 0$ to the curves $t \mapsto \sigma(u_0 + t, v_0)$ and $t \mapsto \sigma(u_0, v_0 + t)$, respectively.



Definition 2.3.1. A parametrized surface σ is called *regular* at $p = (u_0, v_0)$ if the partial derivatives σ'_u and σ'_v, evaluated at p, are linearly independent.

Otherwise it is called *singular.* If σ is regular in all points of U we call it a *regular parametrized surface* or just a *regular surface.*

Recall (see Appendix C) that for two vectors $a = (a_1, a_2, a_3)$ and $b = (b_1, b_2, b_3)$ in $\mathbb{R}^3$ we define the *cross product* by

$$a \times b = \left(\begin{vmatrix} a_2 & b_2 \\ a_3 & b_3 \end{vmatrix}, - \begin{vmatrix} a_1 & b_1 \\ a_3 & b_3 \end{vmatrix}, \begin{vmatrix} a_1 & b_1 \\ a_2 & b_2 \end{vmatrix} \right).$$

Since a and b are linearly independent if and only if $a \times b \neq 0$, the regularity condition above is equivalent to $\sigma'_u \times \sigma'_v \neq 0$.

Example 2.3.1 For the standard spherical coordinates

$$\sigma(u, v) = (\cos u \cos v, \cos u \sin v, \sin u)$$

we derive

$$\sigma'_u = \begin{pmatrix} -\sin u \cos v \\ -\sin u \sin v \\ \cos u \end{pmatrix}, \qquad \sigma'_v = \begin{pmatrix} -\cos u \sin v \\ \cos u \cos v \\ 0 \end{pmatrix}$$

and hence

$$\sigma'_u \times \sigma'_v = (-\cos^2 u \cos v, -\cos^2 u \sin v, -\cos u \sin u) = -\cos u\, \sigma(u, v). \quad (2)$$

In particular, since $\sigma(u, v) \neq 0$ (it has length 1), we see that $\sigma'_u \times \sigma'_v = 0$ if and only if $\cos u = 0$, that is $u = \pm\frac{\pi}{2}$ (up to multiples of 2π). The points $\sigma(p)$ on the sphere, where σ is singular at p, are thus the two poles $(0, 0, \pm 1)$. Notice however that by choosing a different parametrization of the sphere, we can arrange that these points are in the regular range (at the cost of some other points becoming singular). For example with $(u, v) \mapsto (\cos u \cos v, \sin u, \cos u \sin v)$, which differs from σ by an interchange of y and z, the points $\sigma(p)$ with p singular are $(0, \pm 1, 0)$.

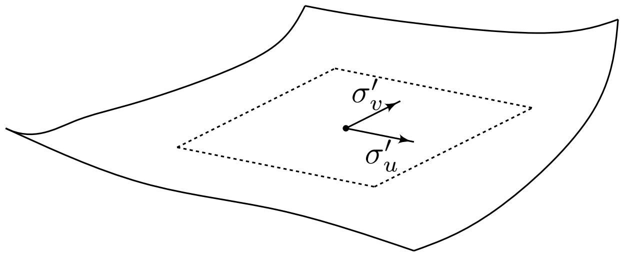
Example 2.3.2 The graph $\sigma(u, v) = (u, v, h(u, v))$ (Section 1.3) of a function of two variables is a regular surface in $\mathbb{R}^3$, since

$$\sigma'_u = (1, 0, \frac{\partial h}{\partial u}) \quad \text{and} \quad \sigma'_v = (0, 1, \frac{\partial h}{\partial v})$$

are linearly independent (also if the partial derivatives of h are 0). Hence, by Corollary 1.6, a level surface $f(x, y, z) = c$ *can be parametrized as a regular surface in a neighborhood of each point which is not critical.*

Definition 2.3.2. The linear subspace of $\mathbb{R}^3$ spanned by the partial derivatives $\sigma'_u(p)$ and $\sigma'_v(p)$ is called the *tangent space* of σ at p. It is denoted $T_p\sigma$.

If σ is regular at $p = (u_0, v_0)$, then the plane through $\sigma(p)$ and parallel to $T_p\sigma$ is called the *tangent plane* at p.



Notice that the tangent space $T_p\sigma$ is a two-dimensional linear subspace of $\mathbb{R}^3$ if and only if σ is regular. In this case the pair of vectors $\sigma'_u(p)$ and $\sigma'_v(p)$ form a basis for $T_p\sigma$, and the use of the word 'plane' for the tangent plane is justified.

Example 2.3.3 Let σ be the standard parametrization of the unit sphere, as in the Example 2.3.1. At $(u_0, v_0) = (0, 0)$ we have $\sigma(0, 0) = (1, 0, 0)$ and $\sigma'_u = e_3$ and $\sigma'_v = e_2$ (where e_1, e_2, e_3 are the standard basis vectors in $\mathbb{R}^3$). The tangent space at $(0, 0)$ is therefore the span of e_2 and e_3 (the yz-plane), and the tangent plane is the plane through $(1, 0, 0)$ parallel to this plane. On the other hand, if $u_0 = \frac{\pi}{2}$ (and v_0 is arbitrary) so that $\sigma(u_0, v_0) = (0, 0, 1)$, then $\sigma'_u = (-\cos v_0, -\sin v_0, 0)$ and $\sigma'_v = 0$, so in this singular case the tangent space at (u_0, v_0) is one-dimensional. However, the degeneracy of the tangent space at this point is caused by the singularity of the parametrization, and it has no geometric significance for the sphere.

It is convenient to have the notion of tangent space because of its structure as a linear space. On the other hand the tangent plane is more easy to visualize, because it passes through the given point on $\mathcal{S} = \sigma(U)$.

For a level set we have the following analogue of Theorem 2.2.

Theorem 2.3. *Let $\mathcal{S} \subset \mathbb{R}^3$ be a level set of a smooth function f, and let $p = (x_0, y_0, z_0) \in \mathcal{S}$ be non-critical. Let $\sigma\colon U \to \mathbb{R}^3$ be any parametrized surface with image $\sigma(U) \subset \mathcal{S}$ and with $\sigma(u_0, v_0) = p$ for some $(u_0, v_0) \in U$, in which σ is regular. Then the tangent plane of σ at (u_0, v_0) is characterized by the equation*

$$\frac{\partial f}{\partial x}(p)(x - x_0) + \frac{\partial f}{\partial y}(p)(y - y_0) + \frac{\partial f}{\partial z}(p)(z - z_0) = 0.$$

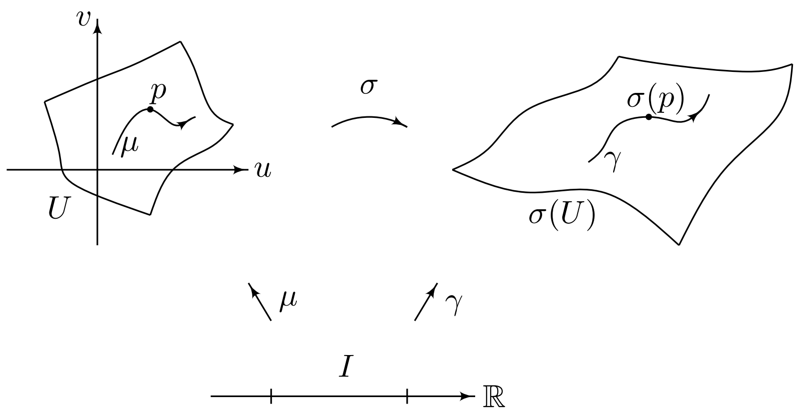
Proof. Entirely analogous to that of Theorem 2.2. $\square$

As in Section 2.2 we observe that the tangent plane in p of the level set $\mathcal{S}$ depends on f but is independent of the particular parametrization σ.

2.4 Curves on a surface

We shall now give a geometric characterization of the tangent space based on the following definition. Let $\sigma\colon U \to \mathbb{R}^3$ be a parametrized surface.

Definition 2.4. A *parametrized curve on* σ is a parametrized curve written in the form $\gamma = \sigma \circ \mu\colon I \to \mathbb{R}^3$ where $\mu\colon I \to U$ is a parametrized plane smooth curve.



Notice that μ is not uniquely determined by γ since we have not assumed σ to be injective. Furthermore, even if σ is injective, so that μ is uniquely determined by γ, then smoothness of μ is not ensured just by the smoothness of γ. It is therefore important to emphasize that in the above definition the smooth curve μ is assumed to be given together with γ.

Formally, a parametrized curve on σ is a *pair* of smooth curves μ and γ satisfying $\gamma = \sigma \circ \mu$. The plane curve μ is said to be the *coordinate curve* of the pair.

Example 2.4.1 The helix $\gamma(t) = (\lambda t, r\cos(\omega t), r\sin(\omega t))$ in Example 1.1.4 is realized as a curve $\sigma \circ \mu$ on the cylinder $\sigma(u, v) = (u, r\cos v, r\sin v)$ with coordinate curve $\mu(t) = (\lambda t, \omega t)$.

Lemma 2.4. *Let $\gamma = \sigma \circ \mu$ be a parametrized curve on σ. Then*

$$\gamma'(t) = u'(t)\sigma'_u(\mu(t)) + v'(t)\sigma'_v(\mu(t)) \tag{3}$$

where $u(t)$ and $v(t)$ are the coordinates of $\mu(t) = (u(t), v(t)) \in U$.

Proof. This follows from the chain rule for $D(G \circ F)$ with $G = \sigma\colon U \to \mathbb{R}^3$ and $F = \mu\colon I \to U$, see Appendix B, equation (B.2). The 3×2 matrix $DG = D\sigma$ has columns σ'_u, σ'_v, and the derivative $F'(t) = \mu'(t)$ has the elements $u'(t)$ and $v'(t)$. Their product $DG(F(t))F'(t)$ is exactly the linear combination $u'\sigma'_u + v'\sigma'_v$, as in (3). $\square$

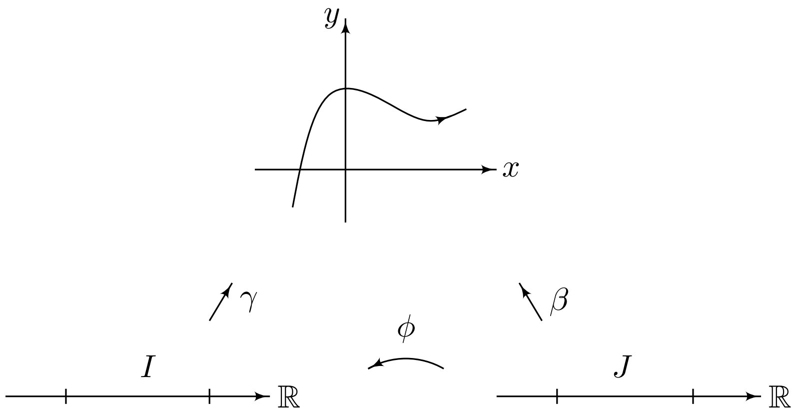
Theorem 2.4. *The tangent space $T_p\sigma$ is equal to the set of tangent vectors $\gamma'(t_0)$ of all parametrized curves $\gamma = \sigma \circ \mu$ on σ with $\mu(t_0) = p$ for some $t_0 \in I$.*

Proof. It follows from (3) that $\gamma'(t_0)$ belongs to the span $T_p\sigma$ of σ'_u and σ'_v for all parametrized curves on σ with $\mu(t_0) = p$.

Conversely, let a linear combination $a\sigma'_u + b\sigma'_v \in T_p\sigma$ be given. Let $p = (u_0, v_0)$ and define $\mu(t) = (u(t), v(t)) = (u_0 + at, v_0 + bt)$ for t sufficiently close to 0, so that $\mu(t) \in U$. Let $\gamma = \sigma \circ \mu$. Then $u'(t) = a$ and $v'(t) = b$, hence it follows from the expression (3) for $\gamma'(t)$ that $\gamma'(0) = a\sigma'_u + b\sigma'_v$. □

2.5 Reparametrization of curves

It can often be useful to change the way a given curve is parametrized. For example, one may prefer to parametrize the unit circle not by $(\cos t, \sin t)$, but by $(\cos(\omega t), \sin(\omega t))$ for some angular velocity ω. This concept is formalised in the following definition.



Definition 2.5.1. Let $\gamma: I \to \mathbb{R}^n$ be a parametrized curve, and let $\phi: J \to I$ be a smooth bijective map with a smooth inverse (I and J being open intervals in $\mathbb{R}$). The curve $\beta = \gamma \circ \phi: J \to \mathbb{R}^n$ is called a *reparametrization* of γ.

For the justification of the condition on ϕ we recall the following result from the calculus of functions of one variable.

Theorem 2.5. *Let $J \subset \mathbb{R}$ be an open interval and $\phi: J \to \mathbb{R}$ a smooth map. Let $I = \phi(J)$, then the following conditions are equivalent:*

(i) $\phi'(u) \neq 0$ *for all* $u \in J$,

(ii) $\phi: J \to I$ *is bijective,* I *is an open interval, and* $\phi^{-1}: I \to J$ *is smooth.*

Moreover, if these conditions hold, then

$$(\phi^{-1})'(t) = \frac{1}{\phi'(u)} \tag{4}$$

for all $t \in I$, where $u = \phi^{-1}(t)$.

Notice that if $\phi'(u) \neq 0$ for all $u \in J$ then, by continuity, either $\phi'(u) > 0$ for all u or $\phi'(u) < 0$ for all u. Thus ϕ is either monotonically increasing or monotonically decreasing.

Let $\beta = \gamma \circ \phi$ be a reparametrization. It follows from the chain rule that the tangent vectors of β are related to those of γ through

$$\beta'(u) = \phi'(u)\gamma'(\phi(u)). \tag{5}$$

Since $\phi'(u) \neq 0$ we see that β is regular if and only if γ is regular. Moreover, if $\phi' > 0$ the tangent vectors of β and γ have mutual directions, and if $\phi' < 0$ they have opposite directions. We say in the former case, where ϕ is increasing, that the reparametrization *preserves direction* and in the latter case, where ϕ is decreasing, that the reparametrization *reverses direction*.

Example 2.5.1 Let $p, q \in \mathbb{R}^n$ be fixed, $q \neq 0$. The curve $\beta(u) = p + \tan u\, q$, $u \in]-\frac{\pi}{2}, \frac{\pi}{2}[$, is a reparametrization of the line $\gamma(t) = p + tq$, $t \in \mathbb{R}$. The transformation between t and u is given by $t = \phi(u) = \tan u$. On the other hand, the curve $\alpha(v) = p + v^3 q$ is *not* a reparametrization, since $v \mapsto v^3$ does not have a differentiable inverse (and in fact, α is not regular).

2.6 Reparametrization of surfaces

We will now generalize some of these concepts to surfaces. The situation is considerably more complicated, because the higher dimensional Euclidean spaces $\mathbb{R}^n$ present some subtleties which do not show up in case $n = 1$. In particular, the theorem given above does not generalize directly to $\mathbb{R}^n$, as will be explained thoroughly later in this chapter (in Section 2.10).

Definition 2.6.1. Let $U, W \subset \mathbb{R}^n$ be open sets. A map $\phi: W \to U$ which is smooth, bijective and has a smooth inverse is called a *diffeomorphism*.

For example, a linear map $L: \mathbb{R}^n \to \mathbb{R}^n$ is a diffeomorphism if and only if the $n \times n$ matrix A that represents it (with respect to some basis for $\mathbb{R}^n$) is invertible. If A is invertible, then L is bijective and its inverse is the linear map represented by A^{-1}, hence this is a smooth map. If A is not invertible, then L is not bijective.

The expression (4) for the derivative of the inverse of a map $J \to I$, where $I, J \subset \mathbb{R}$, has the following generalization for a diffeomorphism $\phi: W \to U$:

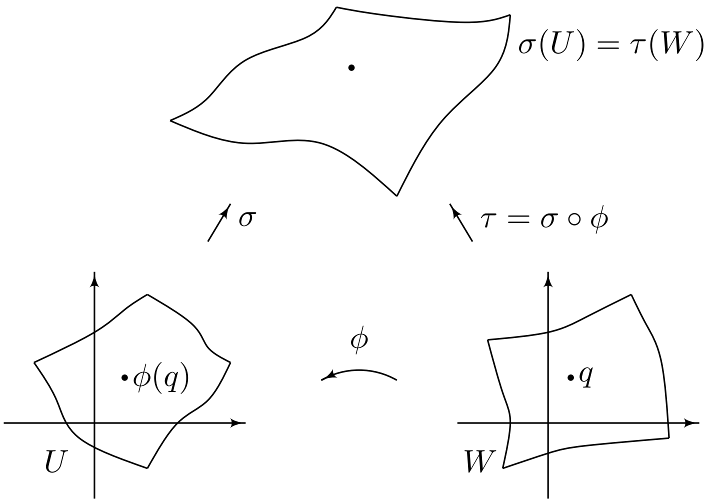
$$D(\phi^{-1})(p) = (D\phi(q))^{-1}$$

where $q = \phi^{-1}(p)$. Here $D\phi$ is the Jacobi matrix of ϕ, and the inverse on the right side is that of a matrix. This formula follows by application of the chain rule to the identity $\phi \circ \phi^{-1} = I$. In particular, the Jacobi matrix of a diffeomorphism is invertible, that is

$$\det(D\phi(q)) \neq 0$$

for all $q \in W$.

Definition 2.6.2. Let $\sigma: U \to \mathbb{R}^3$ be a parametrized surface, and let $\phi: W \to U$ be a diffeomorphism (U and W being open sets in $\mathbb{R}^2$)). The surface $\tau = \sigma \circ \phi: W \to \mathbb{R}^3$ is called a *reparametrization* of σ.



2.7 Invariance under reparametrization

A reparametrization of a curve is considered geometrically insignificant (at least if it is direction-preserving), and geometric properties of curves are required to be unchanged by such a reparametrization; otherwise they do not qualify for being 'geometric'. For example, it follows from (5) that the tangent vector in u of the reparametrized curve β differs by a multiple from that of γ in $t = \phi(u)$, hence the tangent vector is not 'geometric'. However, it also follows from (5) that the tangent *line* is unchanged, hence qualifies better as a 'geometric' object related to the curve. The corresponding result for surfaces is as follows.

Theorem 2.7. *Let $\tau = \sigma \circ \phi$ be a reparametrization of σ. Then the tangent spaces are identical:*

$$T_q\tau = T_{\phi(q)}\sigma, \qquad q \in W.$$

Moreover, τ is regular at q, if and only if σ is regular at $\phi(q)$.

We say that the tangent space is *invariant* under reparametrization. It therefore qualifies as a proper geometric object related to the surface.

Proof. It follows from the chain rule that the partial derivatives of τ are related to those of σ through

$$D\tau(q) = D\sigma(\phi(q)) \cdot D\phi(q),$$

where the dot denotes matrix multiplication.

Let (u, v) denote the coordinates in U and let (s, t) denote the coordinates in W. Let

$$D\phi(q) = \begin{pmatrix} a & c \\ b & d \end{pmatrix}.$$

Writing out the above matrix product in terms of the columns τ_s' and τ_t' of $D\tau(q)$ and the columns σ_u' and σ_v' of $D\sigma(\phi(q))$, it becomes

$$\tau_s' = a\sigma_u' + b\sigma_v', \qquad \tau_t' = c\sigma_u' + d\sigma_v'. \tag{6}$$

These identities show that τ_s' and τ_t' are linear combinations of σ_u' and σ_v', hence they belong to the tangent space $T_{\phi(q)}\sigma$. It follows that $T_q\tau \subset T_{\phi(q)}\sigma$. Since $\sigma = \tau \circ \phi^{-1}$ is also a reparametrization, the same argument with reversed roles of τ and σ shows that $T_{\phi(q)}\sigma \subset T_q\tau$. Thus the equality of the tangent spaces follows.

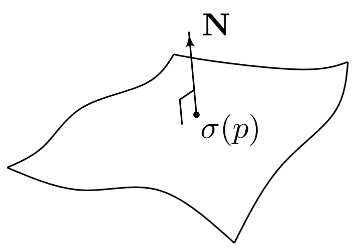
Now τ is regular in q if and only if $T_q\tau$ is two-dimensional, and σ is regular in $\phi(q)$ if and only if $T_{\phi(q)}\sigma$ is two-dimensional. The equivalence of the regularity of σ and τ follows. $\square$

2.8 The unit normal, orientation

Definition 2.8. If σ is regular at $p = (u, v)$, the vector

$$\mathbf{N} = \mathbf{N}(p) = \frac{\sigma_u' \times \sigma_v'}{\|\sigma_u' \times \sigma_v'\|}$$

is called the *unit normal* of the parametrization in p.



Example 2.8.1 It follows from (2) that the unit normal for the unit sphere with spherical coordinates is

$$\mathbf{N}(u, v) = -\sigma(u, v),$$

which is the unit vector pointing from $\sigma(u,v)$ towards the center of the sphere.

The unit normal is perpendicular to the tangent plane in p and has unit length. These properties determine it uniquely up to multiplication with ± 1. Let $\tau = \sigma \circ \phi$ be a reparametrization as in Theorem 2.7. Since $\sigma'_u \times \sigma'_u = \sigma'_v \times \sigma'_v = 0$ and $\sigma'_v \times \sigma'_u = -\sigma'_u \times \sigma'_v$, it follows from (6) that

$$\begin{aligned} \tau'_s \times \tau'_t &= (a\sigma'_u + b\sigma'_v) \times (c\sigma'_u + d\sigma'_v) \\ &= (ad - bc)\, \sigma'_u \times \sigma'_v, \end{aligned} \tag{7}$$

where $ad - bc = \det(D\phi(q)) \neq 0$. This equation shows that under reparametrisation the unit normal is multiplied with the sign of $ad - bc$. If $ad - bc > 0$ we say that the reparametrization *preserves the orientation* at $\phi(q)$, otherwise it *reverses orientation*. This notion is analogous to the notion of direction of a parametrized curve.

Example 2.8.2 Let $\sigma\colon U \to \mathbb{R}^3$ be a parametrized surface, and put $W = \{(v,u) \in \mathbb{R}^2 \mid (u,v) \in U\}$. The map $\phi\colon W \to U$ given by $\phi(v,u) = (u,v)$ is a diffeomorphism, and thus $\tau = \sigma \circ \phi$ is a reparametrization. The effect of this reparametrization is just that it reverses the order of u and v. The Jacobian of ϕ is $D\phi = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$, which has determinant -1. Therefore ϕ reverses orientation.

2.9 Regular curves as graphs

We have given three general descriptions of plane curves, namely as parametrized curves, as graphs of real functions, and as level sets of two-variable functions. In Section 1.5 it was seen that away from critical points, a level set is a graph. Conversely, the graph of a function $y = h(x)$ can be realized as the level set $f(x,y) = 0$ of the function $f(x,y) = h(x) - y$.

As remarked in Example 2.1.2 it is clear that all graphs are regular parametrized curves. We shall now establish the converse, that a regular parametrized curve can be reparametrized as a graph in a neighborhood of each of its points. This will complete the description of interconnections between these various types of curves, the conclusion being essentially that they are all the same.

For simplicity we limit our considerations to plane curves, although a completely similar result holds for curves in $\mathbb{R}^3$.

Theorem 2.9. *Assume that γ is a plane curve, regular at $t_0 \in I$. Then there exists a neighborhood of t_0 in which γ allows a reparametrization as the graph of a smooth function h, considered either as $y = h(x)$ or as $x = h(y)$.*

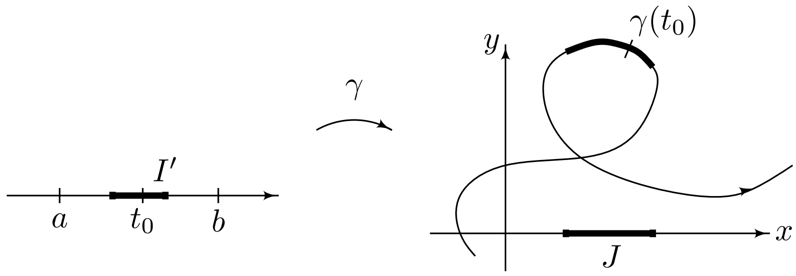
That is, there exists an open interval I' such that $t_0 \in I' \subset I$, an open interval J and a smooth bijective map $\phi\colon J \to I'$ with smooth inverse, such that

$$\gamma(\phi(u)) = (u, h(u))$$

for all $u \in J$*, or*

$$\gamma(\phi(u)) = (h(u), u)$$

for all $u \in J$*.*



Proof. Write $\gamma(t) = (x(t), y(t))$. The assumption is that $(x'(t_0), y'(t_0)) \neq (0,0)$. We are going to prove that if $x'(t_0) \neq 0$, so that the tangent vector is not vertical, then the curve allows a reparametrization as a graph of the form $y = h(x)$. An exchange of x and y then implies that if $y'(t_0) \neq 0$, then the curve allows a reparametrization as a graph of the form $x = h(y)$.

Assume $x'(t_0) \neq 0$. By continuity, there exists an open interval I' around t_0 in which $x'(t) \neq 0$. Let $J = \{x(t) \mid t \in I'\}$. It follows from Theorem 2.5 that the function $t \mapsto x(t)$ from I' to J is bijective with a smooth inverse. When we use this inverse function $\phi\colon J \to I'$ for reparametrization we obtain $\tau(u) = \gamma(\phi(u)) = (x(\phi(u)), y(\phi(u))) = (u, h(u))$ where $h(u) = y(\phi(u))$. □

Example 2.9.1 Let $\gamma(t) = (\cos t, \sin t)$ with $t \in \mathbb{R}$ be the standard parametrisation of the circle, then $\gamma'(t) = (-\sin t, \cos t)$. On the upper half circle, where $t \in]0, \pi[$, we have $x'(t) \neq 0$. Then $t \mapsto x(t) = \cos t$ is bijective $I' =]0, \pi[\to]-1, 1[$ and has a smooth inverse $\cos^{-1}\colon J \to I'$. The reparametrization of γ is then

$$\gamma(\phi(u)) = (u, \sin(\cos^{-1} u)) = (u, \sqrt{1-u^2}), \quad u \in J =]-1, 1[.$$

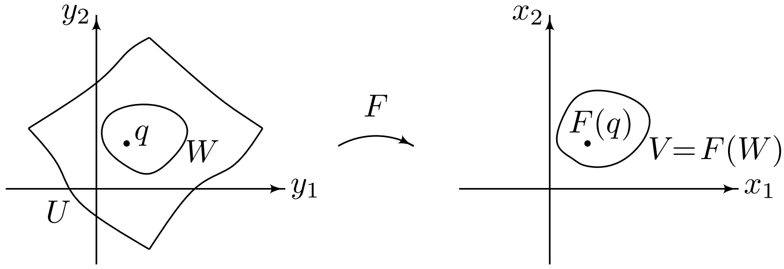
The following corollary is readily obtained, because the parametrization $t \mapsto (t, h(t))$ of a graph is injective.

Corollary 2.9. *A regular parametrized curve* γ *is locally injective, that is, there exist around each* $t_0 \in I$ *a neighborhood such that the restriction of* γ *to this neighborhood is injective.*

2.10 The inverse function theorem

The following fundamental result from multivariable calculus plays a very prominent role in differential geometry. We need it to obtain the analog of Theorem 2.9 for surfaces.

Theorem 2.10. *Let $F: U \to \mathbb{R}^m$ be smooth, where $U \subset \mathbb{R}^m$ is open, and let $q \in U$ be given. Suppose that $\det(DF(q)) \neq 0$. Then there exist an open set $W \subset U$ containing q and an open set $V \subset \mathbb{R}^m$ containing $F(q)$ such that $V = F(W)$ and such that the restriction of F is a diffeomorphism of W onto V (see Definition 2.6.1).*



Proof. It is convenient to distinguish the variables in the source space and the target space (both being $\mathbb{R}^m$) in the way that we view $x = F(y) \in \mathbb{R}^m$ as a function of $y \in U$. The inverse function that we are seeking will then give $y \in W$ as a function of $x \in V$.

We shall apply the implicit function theorem with $n = 2m$ to the map $f: \mathbb{R}^m \times U \to \mathbb{R}^m$ given by $f(x, y) = -x + F(y)$ where $x \in \mathbb{R}^m$, $y \in U$. Notice that $f(x, y) = 0$ if and only if $F(y) = x$. Therefore, if we can exhibit the solution set to the equation $f(x, y) = 0$ as the graph $y = h(x)$ of a function h, then $F(y) = x$ if and only if $y = h(x)$. This means exactly that h is inverse to F.

Let $y_0 \in \mathbb{R}^m$ denote the given point q, and let $x_0 = F(y_0)$. The matrix $A = \frac{\partial f}{\partial y}$ of Theorem 1.7 is exactly $DF(q)$, hence it has a non-vanishing determinant. Thus, according to the theorem there exist open intervals I and J around x_0 and y_0, respectively, and a smooth map $h: I \to J$ such that $f(x, y) = 0$ if and only if $y = h(x)$, for all $(x, y) \in I \times J$. Let $W = J \cap F^{-1}(I)$, then W is open (since F is continuous). It is now seen, as remarked above, that $F: W \to I$ and $h: I \to W$ are the inverse maps of each other. Hence F is a diffeomorphism of W onto $V = I$. $\square$

Remark The present theorem represents an analogue for functions of several variables of Theorem 2.5. There is, however, a fundamental difference between the two theorems. The theorem we have proved is *local*, as it only asserts the existence of an inverse to F in some neighborhood of $F(q)$. Even if the condition $\det(DF(q)) \neq 0$ holds for all $q \in U$, an inverse of F need not exist on all of $F(U)$. This is illustrated in the example below, and it contrasts the situation for $n = 1$: If $F'(x) \neq 0$ on an interval, then F is monotone on that interval, hence bijective, as also stated in Theorem 2.5.

Example 2.10.1 Let $F: U \to \mathbb{R}^2$ be given by

$$F(x, y) = (x^2 - y^2, 2xy),$$

where $U = \mathbb{R}^2 \setminus \{(0,0)\}$. The Jacobian of F,

$$DF(x,y) = \begin{pmatrix} 2x & -2y \\ 2y & 2x \end{pmatrix}$$

has non-zero determinant for all $q = (x,y) \neq (0,0)$, hence the inverse function theorem implies that for each $q \in U$, the restriction of F to a suitable neighborhood of q is invertible. However, since $F(-q) = F(q)$ for all q, F itself is not injective.

Corollary 2.10. *Let $F: U \to \mathbb{R}^m$ be smooth, where $U \subset \mathbb{R}^m$ is open, and suppose that $\det(DF(q)) \neq 0$ for each $q \in U$. Then $F(U)$ is open. If in addition F is injective, then F is a diffeomorphism of U onto $F(U)$.*

Proof. Let $p \in F(U)$ be given, and write $p = F(q)$. According to the theorem above there exists an open set $W \subset U$ around q such that $F(W)$ open. This open set $F(W)$ is then an open neighborhood of p in $F(U)$, hence $F(U)$ is open.

If F is injective, it has an inverse map $F^{-1}: F(U) \to U$. According to the theorem, F^{-1} is smooth in a neighborhood p. Since p was arbitrary, F^{-1} is smooth. $\square$

2.11 Regular surfaces as graphs

In this section we prove the following analogue for surfaces of Theorem 2.9. Let $\sigma: U \to \mathbb{R}^3$ be a parametrized surface.

Theorem 2.11. *Assume that σ is regular at $p \in U$. Then there exists a neighborhood of p in which σ allows a reparametrization such that it becomes the graph of a smooth function ψ, considered either as $z = \psi(x,y)$, $y = \psi(x,z)$ or as $x = \psi(y,z)$.*

As a consequence of this, together with Theorem 1.6, we see, as we saw for curves in Section 2.9, that there are simple connections between regular parametrized surfaces, graphs of two-variable functions and level sets of three-variable functions, away from critical points. Essentially these are different descriptions of the same kind of objects.

Proof. Write

$$\sigma(u,v) = (f(u,v), g(u,v), h(u,v)).$$

Since σ is regular at p the columns of the Jacobian

$$D\sigma = \begin{pmatrix} f'_u & f'_v \\ g'_u & g'_v \\ h'_u & h'_v \end{pmatrix}$$

are linearly independent at p. By changing the order of the coordinates on $\mathbb{R}^3$ if necessary, we may arrange that the two first rows of $D\sigma(p)$ are independent. Let $\pi\colon \mathbb{R}^3 \to \mathbb{R}^2$ denote the projection $(x, y, z) \mapsto (x, y)$ and put

$$F = \pi \circ \sigma\colon U \to \mathbb{R}^2.$$

Then $F(u, v) = (f(u, v), g(u, v))$ and

$$DF = \begin{pmatrix} f'_u & f'_v \\ g'_u & g'_v \end{pmatrix}.$$

It follows that $\det DF(p) \neq 0$. By the inverse function theorem there exists an open neighborhood W of p in U such that F is a diffeomorphism of W onto the open set $V = F(W) = \pi(\sigma(W)) \subset \mathbb{R}^2$.

Let $\phi = F^{-1}\colon V \to W$, then $\pi \circ \sigma \circ \phi = F \circ \phi$ is the identity map on V, that is, the first two coordinates of $\sigma(\phi(s, t))$ are exactly s and t. We define the function $\psi(s, t)$ as the third coordinate of $\sigma(\phi(s, t))$, then $\sigma(\phi(s, t)) = (s, t, \psi(s, t))$ as desired. $\square$

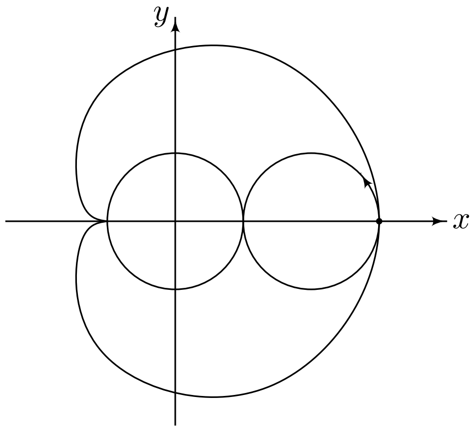
Corollary 2.11. *A regular parametrized surface σ is locally injective, that is, there exist around each point $p \in U$ a neighborhood such that the restriction of σ to this neighborhood is injective.*

2.12 Exercises

1 The following curve is called the *cardioid* (because of its heart-like shape):

$$\gamma(t) = (2\cos t + \cos 2t, 2\sin t + \sin 2t).$$

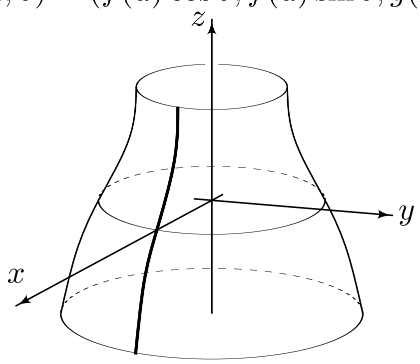
For which values of t is it regular? Find the point where t is singular in the figure below. The curve is constructed by a circle of radius 1 rolling without slipping on the outside of a fixed circle also of radius 1. The curve is the trace of a point on the circumference of the rolling circle.



2 Let $\gamma(t)$ be a parametrized curve which does not pass through the origin, and let $\gamma(t_0)$ be a point of the trace which is closest to the origin. Show that the position vector $\gamma(t_0)$ is orthogonal to the tangent vector $\gamma'(t_0)$.

3 Let $\gamma: I \to \mathbb{R}^3$ be a parametrized curve in the xz-plane, that is $\gamma(u) = (f(u), 0, g(u))$, and assume that $f(u) > 0$ for all $u \in I$. This curve, called the *profile curve*, is rotated around the z-axis. The result is a so-called *surface of revolution* :

$$\sigma(u,v) = (f(u)\cos v, f(u)\sin v, g(u)),$$



a. Explain how the parameter v describes the rotation around the z-axis.

b. Examples: $\gamma(u) = (1, 0, u)$ and $\gamma(u) = (u, 0, u)$ (the last case requires $u > 0$). Describe the corresponding surfaces of revolution.

c. Describe a sphere, minus two poles, as a surface of revolution. Which is the profile curve, and which coordinates on the sphere are obtained?

d. Assume that γ is a regular parametrized curve. Show that σ is a regular parametrized surface.

4 Let $\sigma(u,v) = (\cos u \cos v, \cos u \sin v, \sin u)$, the standard spherical coordinates.

a. Show that the tangent space of σ at $p = (0, \pi)$ is $T_p\sigma = \mathrm{Span}(e_2, e_3)$. Determine also the tangent space at $p = (\frac{\pi}{4}, 0)$.

c. Let $w = (1, 1, -1)$. Show that $w \in T_p\sigma$ where $p = (\frac{\pi}{4}, 0)$, and determine a curve on σ through p and with w as its tangent vector.

5 Consider the surface $\sigma(u,v) = (u^3, v^3, uv)$, $(u,v) \in \mathbb{R}^2$.

a. For which points $p = (u,v)$ is it regular?

b. Determine the tangent space $T_p\sigma \subset \mathbb{R}^3$ for each of the points $p_1 = (1,0)$, $p_2 = (1,1)$ and $p_3 = (0,0)$. Determine also the tangent *plane* in p_1 and p_2. Why not in p_3?

c. Show that σ is a bijection of $\mathbb{R}^2$ onto $\mathcal{S} = \{(x,y,z) \mid xy - z^3 = 0\}$.

d. Use Theorem 2.3 to determine the tangent plane in p_1.

e. The vector $v = (3, 6, 3)$ belongs to $T_p\sigma$ where $p = p_2$. Find a curve γ on σ with $\gamma(t_0) = \sigma(p)$ and $\gamma'(t_0) = v$ (it exists by Theorem 2.4).

6 Let $\sigma(u,v)$ be as in Exercise 5. In each of the following cases, determine whether γ can be realized as a parametrized curve on σ.
a) $\gamma(t)=(t^3,t^3,t^2)$, b) $\gamma(t)=(t^3,t^3,t^3)$, c) $\gamma(t)=(t,t^2,t)$.

7 Let

$$\gamma(t)=\Big(\frac{1-t^2}{1+t^2},\frac{2t}{1+t^2}\Big),\quad t\in\mathbb{R}.$$

Which curve is obtained through the reparametrization $\beta=\gamma\circ\phi$, where $\phi(u)=\tan\frac{u}{2}$ for $u\in]-\pi,\pi[$?

8 Let $\sigma(u,v)=(u,v,h(u,v))$ be the graph of a function $h\colon\mathbb{R}^2\to\mathbb{R}$. Show that the unit normal is given by

$$\mathbf{N}(u,v)=\frac{(-h'_u,-h'_v,1)}{\sqrt{1+(h'_u)^2+(h'_v)^2}}.$$

9 Let $\sigma(u,v)$, $(u,v)\in\mathbb{R}^2$ be a smooth surface and put $\tau(s,t)=\sigma(-t,s)$. Show that τ is obtained from σ by a reparametrization. Does it preserve or reverse orientations?

10 Let $\sigma(u,v)=(u,uv,\frac{1}{2}v^2)$, $(u,v)\in\mathbb{R}^2$. Determine σ'_u, σ'_v and $\sigma'_u\times\sigma'_v$. For which (u,v) is σ regular? Determine the unit normal $\mathbf{N}$ at $(u,v)=(4,2)$.

11 Let $\sigma(u,v)=(u,uv,\frac{1}{2}v^2)$ for $(u,v)\in U=\{(u,v)\in\mathbb{R}^2\mid u\neq 0\}$. Show that $(u,v)\mapsto(u,uv)$ is a diffeomorphism $U\to U$, and determine the inverse map $\phi\colon U\to U$. Show that the reparametrization $\sigma\circ\phi\colon U\to\mathbb{R}^3$ of σ is a graph of the form $z=h(x,y)$, $(x,y)\in U$.

12 Let again $\sigma(u,v)=(u,uv,\frac{1}{2}v^2)$. Find two open sets $U,W\subset\mathbb{R}^2$ (non-empty), and a diffeomorphism $\phi\colon W\to U$, such that the reparametrization $\sigma\circ\phi$ of $\sigma|_U$ is a graph of the form $x=h(y,z)$, where $(y,z)\in W$.

13 Let $U=\{(u,v)\in\mathbb{R}^2\mid u>v\}$ and $\sigma(u,v)=(\frac{1}{2}(u+v),\frac{1}{2}(u^2+v^2),uv)$ for $(u,v)\in U$. Let $p=(2,0)$.

a. Show that σ is regular at p, and determine $T_p\sigma$.

b. Let $W=\{(s,t)\in\mathbb{R}^2\mid s^2>t\}$ and define $\phi\colon W\to\mathbb{R}^2$ by

$$\phi(s,t)=(s+\sqrt{s^2-t},s-\sqrt{s^2-t}).$$

Show that ϕ is a diffeomorphism of W onto U, and determine whether it preserves or reverses orientation.

c. The surface $\tau=\sigma\circ\phi$ is the graph of a function. Which?

d. Find $q\in W$ such that $\phi(q)=p$, and determine then $\tau(q)$ and $T_q\tau$.

14 In this exercise we identify the set $M_{2,2}$ of 2×2 real matrices with $\mathbb{R}^4$ by numbering the entries in some (arbitrary) fashion. Let $F\colon M_{2,2} = \mathbb{R}^4 \to M_{2,2} = \mathbb{R}^4$ denote the map $A \mapsto A^2$ where the square is computed by matrix multiplication. Determine the 4×4 matrix $DF(I)$, where I is the identity matrix in $M_{2,2}$. Show that every matrix sufficiently close to I has a square root, which is unique if it is required to be sufficiently close to I.

15 Let $\sigma\colon U \to \mathbb{R}^3$ be an injective and regular parametrized surface, and assume that its image $\sigma(U)$ is contained in the xy-plane. Show that the set $V = \{(s,t) \mid (s,t,0) \in \sigma(U)\}$ is open in $\mathbb{R}^2$, and that the plane surface $\tau(s,t) = (s,t,0)$, $(s,t) \in V$, can be achieved as a reparametrization of σ (hint: apply Corollary 2.10 to $F = \pi \circ \sigma$ in the proof of Theorem 2.11).

Chapter 3

The first fundamental form

We shall introduce notions that allow us to treat metric questions on curves and surfaces, for example the determination of the length of a curve and the area of a subset of a surface. The notion of distance along a curve will be closely associated with the standard notion of the length of a vector in Euclidean space $\mathbb{R}^2$ or $\mathbb{R}^3$. The Euclidean notion of length is used on tangent vectors, and it can be viewed as a means to define the distance of very close points ('infinitesimal distances'). The ('global') distance between two points along the curve is then obtained by integration of these local distances.

Areas are defined similarly by multiple integrals. Certain topics in connection with the latter will be dealt with on a more intuitive level, because they are most efficiently treated by means of the Lebesgue measure, which we do not assume the reader to be acquainted with. The notion of area will only be used sporadically in the following chapters, but it is an important concept in the geometry of surfaces.

3.1 Arc length

Let $\gamma\colon I \to \mathbb{R}^n$ be a smooth curve. The *speed* of γ at $t \in I$ is defined to be the length $\|\gamma'(t)\|$ of the derivative $\gamma'(t) \in \mathbb{R}^n$, in accordance with the physical interpretation of γ as describing the motion of a particle in n-space. In this interpretation $\gamma'(t)$ is the velocity vector for the particle at time t.

The vector from $\gamma(t)$ to $\gamma(t + \Delta t)$ is approximately $\gamma'(t)\Delta t$, according to the first order (linear) approximation of γ, hence the distance between these points on the curve is approximately $\|\gamma'(t)\|\Delta t$. Adding up all these distances and taking the limit $\Delta t \to 0$, we are lead to the following formula for the distance along γ between $\gamma(t_1)$ and $\gamma(t_2)$:

$$\int_{t_1}^{t_2} \|\gamma'(t)\|\, dt. \tag{1}$$

The derivation we gave for this formula is not a rigorous proof. Rather than carrying out such a proof we will take the formula as a definition, and regard the derivation as motivation.

Definition 3.1. Let $\gamma\colon I \to \mathbb{R}^n$ be a smooth curve. The *arc-length* of γ from $t_1 \in I$ to $t_2 \in I$ is $\int_{t_1}^{t_2} \|\gamma'(t)\|\, dt$.

An *arc-length function* for γ is a primitive of $t \mapsto \|\gamma'(t)\|$, that is, a differentiable function $\ell\colon I \to \mathbb{R}$ with $\ell'(t) = \|\gamma'(t)\|$. The arc-length from t_1 to t_2 is then $\ell(t_2) - \ell(t_1)$. Notice that we do *not* require $t_1 \leq t_2$. If $t_2 < t_1$, then the arc-length is negative.

Example 3.1.1 Let $\gamma(t) = p + tq$ be a straight line (where $q \neq 0$). The arc length along γ from $p + t_1 q$ to $p + t_2 q$ is

$$\int_{t_1}^{t_2} \|q\|\, dt = \|q\|(t_2 - t_1) = \|\gamma(t_2) - \gamma(t_1)\|$$

if $t_1 < t_2$.

Example 3.1.2 A circle of radius r is parametrized by $\gamma(t) = (r\cos t, r\sin t)$, for which the speed $\|\gamma'(t)\| = \|(-r\sin t, r\cos t)\| = r$ is constant. Hence the arc-length from 0 to t is

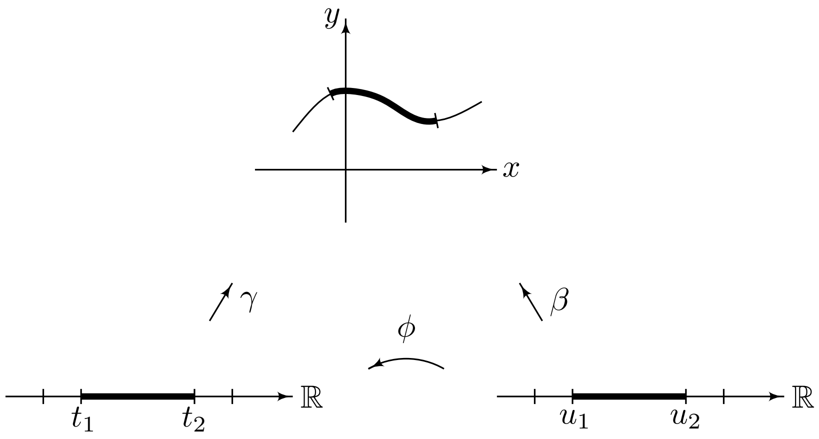
$$\int_0^t \|\gamma'(t)\|\, dt = rt.$$

Example 3.1.3 For the helix given by $\gamma(t) = (\lambda t, r\cos(\omega t), r\sin(\omega t))$ (see Example 1.1.4) we have $\gamma'(t) = (\lambda, -r\omega\sin(\omega t), r\omega\cos(\omega t))$ and the speed $\|\gamma'(t)\| = \sqrt{\lambda^2 + r^2\omega^2}$ is again constant. Hence the arc-length measured from 0 is this constant times t.

As explained in Section 2.7, reasonable geometric notions are invariant under reparametrizations that do not reverse orientation. The following result shows that arc length has this property.

Theorem 3.1. *Let $\gamma\colon I \to \mathbb{R}^n$ be a parametrized curve, and let $\beta = \gamma\circ\phi\colon J \to \mathbb{R}^n$ be a reparamatrization. Let $u_1, u_2 \in J$ and let $t_i = \phi(u_i)$ for $i = 1, 2$.*

If ϕ preserves the direction then the arc-length of β from u_1 to u_2 equals the arc-length of γ from t_1 to t_2. If ϕ reverses direction the arc-lengths are of the same absolute size but have opposite signs.



Proof. By the chain rule $\beta'(u) = \gamma'(\phi(u))\phi'(u)$. Hence

$$\int_{u_1}^{u_2} \|\beta'(u)\|\,du = \int_{u_1}^{u_2} \|\gamma'(\phi(u))\||\phi'(u)|\,du = \pm\int_{t_1}^{t_2} \|\gamma'(t)\|\,dt$$

where in the last step we have used the substitution $t = \phi(u)$. The sign in front is positive if ϕ' is positive, and otherwise negative. □

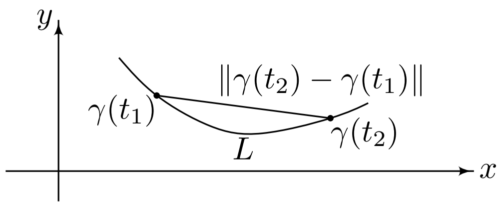
3.2 Lines as shortest curves

Let $P_1, P_2 \in \mathbb{R}^n$. The *linear curve* from P_1 to P_2 is parametrized by $t \mapsto P_1 + t(P_2 - P_1)$ where $t \in [0; 1]$. It has length $\|P_2 - P_1\|$ (see Example 3.1.1). Its trace, the *line segment* from P_1 to P_2, is denoted by $[P_1, P_2]$.

The geometric interpretation of the following theorem is that the linear curve is shortest from P_1 to P_2. Notice however that because of the possibility of reparametrization, the linear curve is not unique in this respect.

Theorem 3.2. *Let $\gamma\colon I \to \mathbb{R}^n$ be a parametrized curve. Let $t_1 < t_2$ in I and let L denote the arc length of γ from t_1 to t_2. Then*

$$L \geq \|\gamma(t_2) - \gamma(t_1)\|. \tag{2}$$



Proof. Let $P_1 = \gamma(t_1)$, $P_2 = \gamma(t_2)$ and $w = P_2 - P_1$, and consider the function

$$\varphi(t) = \gamma(t) \cdot w.$$

We have $\varphi(t_2) - \varphi(t_1) = (P_2 - P_1) \cdot w = \|w\|^2$, hence by the fundamental theorem of calculus

$$\|w\|^2 = \int_{t_1}^{t_2} \varphi'(t)\,dt.$$

It is easily seen that $\varphi'(t) = \gamma'(t) \cdot w$, hence

$$\varphi'(t) \leq |\varphi'(t)| \leq \|\gamma'(t)\|\,\|w\|.$$

We conclude

$$\|w\|^2 \leq \int_{t_1}^{t_2} \|\gamma'(t)\|\,\|w\|\,dt = L\|w\|$$

from which (2) follows. □

3.3 Unit speed parametrization

A parametrized curve γ is said to have *unit speed* if $\|\gamma'(t)\| = 1$ at all points. It is common practice to replace the symbol for the variable by s in this case. For a curve with unit speed, the determination of arc-lengths is particularly simple, because by (1) the arc-length from s_1 to s_2 is equal to the difference of the parameters $s_2 - s_1$. We say that the curve is *parametrized by arc-length.*

Theorem 3.3. *A regular parametrized curve γ allows a direction-preserving reparametrization with unit speed.*

Proof. Let $\ell(t)$ be an arbitrary arc-length function for γ, that is, a primitive of the speed function $t \mapsto \|\gamma'(t)\|$. The speed function is smooth since $\gamma'(t)$ is smooth and never zero. Hence ℓ is smooth. Notice that $\ell'(t) = \|\gamma'(t)\| > 0$.

We apply Theorem 2.5 to the function ℓ. It follows that ℓ is bijective onto its image. Furthermore, the inverse function $\phi = \ell^{-1}$ is smooth, and its derivative is given by

$$\phi'(s) = \frac{1}{\ell'(t)} = \frac{1}{\|\gamma'(t)\|} > 0$$

where $s = \ell(t)$. We use the function ϕ for the reparametrization. Then

$$(\gamma \circ \phi)'(s) = \gamma'(\phi(s))\phi'(s) = \frac{\gamma'(t)}{\|\gamma'(t)\|}$$

where $t = \phi(s)$. Hence $\gamma \circ \phi$ has unit speed. $\square$

Example 3.3.1 For a curve γ with constant speed $c \neq 0$, the function $\ell(t) = ct$ is a primitive of the speed function. The inverse of the map $t \mapsto ct$ is $\phi(s) = \frac{s}{c}$, hence a unit speed reparametrization is obtained by inserting $t = \frac{s}{c}$ in the expression for γ. For example a unit speed reparametrization of the circle $\gamma(t) = (r\cos t, r\sin t)$ (see Example 3.1.2) is

$$\beta(s) = \gamma(\frac{s}{r}) = (r\cos\frac{s}{r}, r\sin\frac{s}{r}),$$

and the helix in Example 3.1.3 is reparametrized with unit speed in

$$\beta(s) = (\lambda\frac{s}{c}, r\cos(\omega\frac{s}{c}), r\sin(\omega\frac{s}{c}))$$

where $c = \sqrt{\lambda^2 + r^2\omega^2}$.

3.4 The first fundamental form

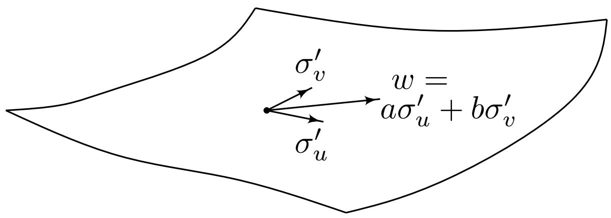
Let $\sigma: U \to \mathbb{R}^3$ be a parametrized surface. We define the following three functions on U, associated with σ:

$$E(p) = \|\sigma'_u(p)\|^2, \quad F(p) = \sigma'_u(p) \cdot \sigma'_v(p), \quad G(p) = \|\sigma'_v(p)\|^2,$$

where $p \in U$. These functions together should be seen as the analogue for surfaces of the speed $\|\gamma'(t)\|$ of a curve (or rather, of $\|\gamma'(t)\|^2$).

The functions E, F and G are useful for the computation of lengths of tangent vectors. If a vector $w \in T_p\sigma$ has coordinates a, b with respect to the basis $\sigma'_u(p), \sigma'_v(p)$, that is $w = a\sigma'_u + b\sigma'_v$, then its length is given by

$$\|w\|^2 = (a\sigma'_u + b\sigma'_v) \cdot (a\sigma'_u + b\sigma'_v) = Ea^2 + 2Fab + Gb^2.$$



Definition 3.4. The map $I_p: T_p\sigma \to \mathbb{R}$ that associates to a tangent vector at p the square of its length,

$$w \mapsto I_p(w) = \|w\|^2 = E(p)a^2 + 2F(p)ab + G(p)b^2,$$

is called the *first fundamental form* of σ in p. The coefficients E, F and G are called the *component functions*.

The component functions E, F and G are conveniently arranged as the entries of a symmetric matrix

$$\begin{pmatrix} E & F \\ F & G \end{pmatrix}.$$

By noting that σ'_u and σ'_v are the columns of the Jacobian matrix $D\sigma$, we see that the definition of E, F and G amounts to the matrix identity

$$\begin{pmatrix} E & F \\ F & G \end{pmatrix} = (D\sigma)^t D\sigma \tag{3}$$

where t denotes transposition. The formula for the first fundamental form can also be put in matrix form

$$I_p(w) = \begin{pmatrix} a \\ b \end{pmatrix}^t \begin{pmatrix} E(p) & F(p) \\ F(p) & G(p) \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix}.$$

By definition, a *quadratic form* on a two dimensional real vector space V with basis vectors v_1, v_2 is a map $Q\colon V \to \mathbb{R}$, which has the form

$$w = av_1 + bv_2 \mapsto Q(w) = ea^2 + 2fab + gb^2$$

for some numbers $e, f, g \in \mathbb{R}$. The first fundamental form I_p is a quadratic form on $T_p\sigma$, for each $p \in U$.

Example 3.4.1 For the plane parametrized by $\sigma(u,v) = p + uq_1 + vq_2$, where q_1, q_2 are linearly independent vectors in $\mathbb{R}^3$, we have $\sigma'_u = q_1$ and $\sigma'_v = q_2$. It follows that the component functions are constant:

$$E = \|q_1\|^2, \quad F = q_1 \cdot q_2, \quad G = \|q_2\|^2.$$

In particular, if q_1, q_2 is an orthonormal pair, we have $E = G = 1$, $F = 0$.

Example 3.4.2 For the parametrization $\sigma(u,v) = (r\cos v, r\sin v, u)$ of the cylinder, we obtain $\sigma'_u = (0,0,1)$ and $\sigma'_v = (-r\sin v, r\cos v, 0)$, so that

$$E = 1, \quad F = 0, \quad G = r^2.$$

As before, the component functions are constant.

Example 3.4.3 For the unit sphere with spherical coordinates we determined σ'_u and σ'_v in Example 2.3.1. An easy computation shows that

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

Notice that in this case the component function $G(p)$ is not constant.

The following theorem illustrates how the first fundamental form enters in the computation of arc lengths on surfaces. Recall from Section 2.4 that $\gamma(t)$ is called a parametrized curve on σ if it has the form $\gamma(t) = \sigma(u(t), v(t))$ for a pair of smooth functions with $(u(t), v(t)) \in U$, and that in this case (see Lemma 2.4)

$$\gamma' = u'\,\sigma'_u + v'\,\sigma'_v. \tag{4}$$

Theorem 3.4. *The arc length of a parametrized curve $\gamma(t) = \sigma(u(t), v(t))$ on σ is given with respect to the coordinates $(u(t), v(t))$ as follows:*

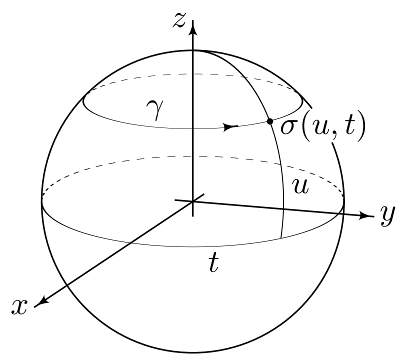
$$\int_{t_1}^{t_2} (Eu'^2 + 2Fu'v' + Gv'^2)^{1/2}\,dt$$

where the component functions E, F, G are evaluated in $(u(t), v(t))$ and the derivatives u', v' are evaluated in t.

Proof. This is immediate from Definition 3.1, since by (4)

$$\|\gamma'(t)\|^2 = I_p(\gamma'(t)) = Eu'(t)^2 + 2Fu'(t)v'(t) + Gv'(t)^2. \quad \square$$

Example 3.4.4 On the unit sphere consider the circle $\gamma(t) = \sigma(u,t)$ with a fixed latitude u. Since u is constant, we have $u' = 0$, and since $v(t) = t$, we have $v' = 1$.



With the values of E, F and G from Example 3.4.3 we obtain the total length of γ:

$$\int_0^{2\pi} (Eu'^2 + 2Fu'v' + Gv'^2)^{1/2}\,dt = \int_0^{2\pi} \cos u\,dt = 2\pi\cos u.$$

The first fundamental form can also be used to determine the angle between (non-zero) tangent vectors, say between

$$w = a\sigma'_u(p) + b\sigma'_v(p) \quad \text{and} \quad \tilde{w} = \tilde{a}\sigma'_u(p) + \tilde{b}\sigma'_v(p)$$

in $T_p\sigma$. If the angle is $\theta \in [0,\pi]$, then it is well-known from Euclidean geometry (see Appendix A) that $\cos\theta = \frac{w\cdot\tilde{w}}{\|w\|\,\|\tilde{w}\|}$, from which we obtain

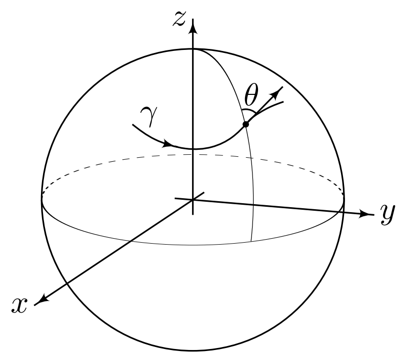
$$\cos\theta = \frac{Ea\tilde{a} + F(a\tilde{b} + b\tilde{a}) + Gb\tilde{b}}{(Ea^2 + 2Fab + Gb^2)^{1/2}(E\tilde{a}^2 + 2F\tilde{a}\tilde{b} + G\tilde{b}^2)^{1/2}}. \tag{5}$$

Although not particularly simple this formula allows the computation of θ from knowledge of the coordinates a, b, $\tilde{a}$ and $\tilde{b}$. In particular, the angle between σ'_u and σ'_v is given by

$$\cos\theta = \frac{F}{\sqrt{EG}}.$$

A parametrized surface $\sigma(u,v)$ is called *orthogonal*, if $F(p) = 0$ for all $p \in U$, or equivalently, if $\sigma'_u(p)$ and $\sigma'_v(p)$ are perpendicular for all p.

Example 3.4.5 Let $\gamma(t)$ be a curve on the unit sphere, which in spherical coordinates is described by $\gamma(t) = \sigma(u(t), v(t))$. We will determine the angle θ between the tangent vector $\gamma'(t)$ and the direction (North) of the meridians.



The coordinates of $\gamma'(t)$ with respect to (σ'_u, σ'_v) are determined from (4). They are $a = u'(t)$ and $b = v'(t)$. The meridians are characterized by having a fixed longitude v, hence the tangent vector of a meridian has direction σ'_u (with coordinates $\tilde{a} = 1$, $\tilde{b} = 0$). With the values of E, F and G from Example 3.4.3 inserted in (5) we obtain

$$\cos\theta = \frac{u'}{((u')^2 + \cos^2 u(v')^2)^{1/2}}.$$

3.5 Introduction to areas and plane integrals

In this section we will give a short introduction to the theory of plane integrals of continuous functions. Not all proofs will be given.

Consider a plane set $D \subseteq \mathbb{R}^2$. If $D = [a, b] \times [c, d]$, where $a \le b$, $c \le d$, we call it a *rectangle*, and we define that it has the area $A(D) = (b - a)(d - c)$. Moreover, in this case if $f \colon D \to \mathbb{R}$ is continuous we define the *integral* of f over D by

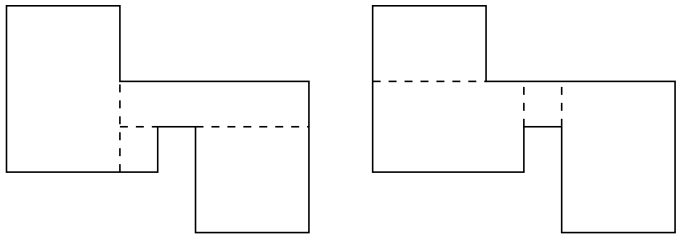
$$\int_D f\,dA = \int_a^b \int_c^d f(u, v)\,dv\,du. \tag{6}$$

It can be shown that the inner integral, $\int_c^d f(u, v)\,dv$, depends continuously on u, so that the outer integral makes sense. One can also prove that we have as well

$$\int_D f\,dA = \int_c^d \int_a^b f(u, v)\,du\,dv,$$

that is, the order of the integrations can be interchanged.

If D is not a rectangle, it is more complicated to define its area, and to define integrals over it. By a *block-set* we will understand a set K which is a finite union of closed rectangles. Notice that by decomposing further the rectangles used, such a set K can always be written as a finite union of closed rectangles, which only overlap on the boundaries. Such a decomposition will be called a *partition* of the block set. In general, the same block-set may have several different partitions, as in the following figure.



The area $A(K)$ is defined as the sum of the areas of the rectangles in a chosen partition, and the integral $\int_K f\,dA$ of a continuous function f over K is defined as the sum of the integrals over these rectangles. Since the partition of K is not unique, a proper treatment would require that it is verified that these notions are independent of the choice of partition. Intuitively this is quite clear, and we are not going to verify it here. Notice that it follows from these definitions that the area of K is the integral over K of the constant function 1, and that in general

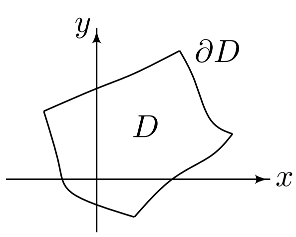
$$|\int_K f\,dA\,| \leq A(K)\sup_{p\in K}|f(p)|,$$

since this inequality holds for each of the subrectangles in K. Moreover, if K_1, K_2 are block sets which only overlap on their boundaries, then

$$\int_{K_1\cup K_2} f\,dA = \int_{K_1} f\,dA + \int_{K_2} f\,dA.$$

We will now consider more general sets $D \subset \mathbb{R}^2$. In the following definition, we consider smooth curves defined on *closed* intervals. That is, $\gamma\colon [a,b] \to \mathbb{R}^2$, where $-\infty < a < b < \infty$. This means that γ is smooth on (a,b) and that γ and all its derivatives have continuous extensions to $[a,b]$ (that is, they have limits for $t \to a$ from the right and for $t \to b$ from the left).

Definition 3.5.1. A set $D \subset \mathbb{R}^2$ is called an *elementary domain* if it is closed and bounded, and if its boundary ∂D is a finite union of (the trace of) smooth curves defined on closed intervals, as above.



An elementary domain

In particular a block-set is an elementary domain, since its boundary is a union of line segments.

Definition 3.5.2. Let $D \subset \mathbb{R}^2$ be an elementary domain. The *area* of D is defined by

$$A(D) = \sup_{K \subset D} A(K),$$

where the supremum is taken over all block-sets $K \subset D$. The *integral* of a continuous function $f\colon D \to \mathbb{R}$ with $f(p) \geq 0$ for all p, is defined by

$$\int_D f\,dA = \sup_{K \subset D} \int_K f\,dA.$$

It should be noticed that the supremums are finite. Since D is bounded, it is contained in a square of sufficiently large side length, say N. Hence the area $A(K)$ of any block-set K inside D is bounded above by the area N^2 of the square, and hence the same bound is valid for the supremum of the $A(K)$. The integral $\int_K f\,dA$ is bounded by $A(K)\sup_{p\in K} f(p)$, which in turn is bounded by $A(D)\sup_{p\in D} f(p)$, which is finite since f is continuous. The same bound is then valid for the supremum in the definition of the integral.

The assumption $f \geq 0$ is now removed. Let $f\colon U \to \mathbb{R}$ be continuous, and put

$$f_+(x) = \max\{0, f(x)\} \quad \text{and} \quad f_-(x) = \max\{0, -f(x)\},$$

so that $f_+ \geq 0$, $f_- \geq 0$, and $f = f_+ - f_-$. We define

$$\int_D f\,dA = \int_D f_+\,dA - \int_D f_-\,dA.$$

It is easily seen that if D is already a block set, these definitions of area and integral amount to the same as was already defined. Moreover, plane integrals share the following familiar properties of ordinary integrals (with obvious notation), of which we shall give no proof:

$$\int_D f + g\,dA = \int_D f\,dA + \int_D g\,dA$$

$$\int_D cf\,dA = c\int_D f\,dA$$

$$\left|\int_D f\,dA\right| \leq \int_D |f|\,dA$$

$$\int_{D_1\cup D_2} f\,dA = \int_{D_1} f\,dA + \int_{D_2} f\,dA,$$

where in the last line D_1 and D_2 are assumed to intersect only with their boundaries.

3.6 Null sets

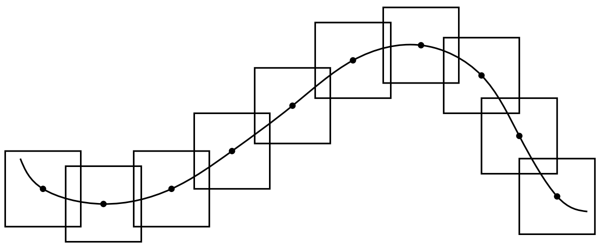
In this section we will prove a theorem which serves as motivation for the preceding definitions of area and integrals over an elementary domain D. In that definition we only considered block sets which were inside D, and the legitimate question is whether we 'miss' a substantial part of D by this. The theorem below shows that this is not the case, and thus the definitions are reasonable.

We say that a closed bounded set D is a *null set* if for each $\epsilon > 0$ there exists a block-set K of area $< \epsilon$ such that $D \subset K$.

As an example, consider a smooth curve $\gamma\colon [a, b] \to \mathbb{R}^2$, where $-\infty < a < b < \infty$. This means that γ is smooth on (a, b) and that all derivatives have a continuous extension to $[a, b]$.

Lemma 3.6. *Let $\gamma\colon [a, b] \to \mathbb{R}^2$ be smooth. The trace $\gamma([a, b])$ is a null set.*

Proof. Using the continuous arc-length function $s(t)$, we can divide γ in N pieces of equal length ℓ/N, where ℓ is the total length. Each piece is contained in the disk of radius $\ell/2N$ centered in the mid-point of the piece (this follows from Theorem 3.2). Hence the piece is also contained in the square of side length ℓ/N with the same center.



The union of these N squares has area at most $N(\ell/N)^2 = \ell^2/N$, which is $\leq \epsilon$ for N sufficiently large. $\square$

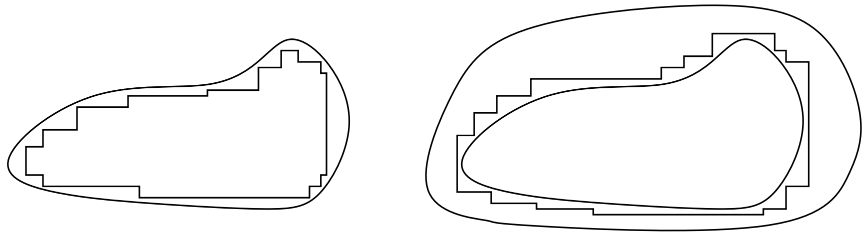
Since a finite union of null sets is a null set, it follows from the preceding lemma that the boundary of an elementary domain is a null set.

Theorem 3.6. *Let $U \subset \mathbb{R}^2$ be an open set, and let $f\colon U \to [0, \infty[$ be a continuous function. Let $D \subset U$ be an elementary domain. Then*

$$\int_D f\, dA = \inf_{D \subset K \subset U} \int_K f\, dA \tag{7}$$

where the infimum is taken over block-sets K.

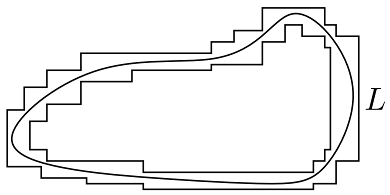
Proof. We first observe that there exist block sets K such that $D \subset K \subset U$. The proof of this depends on the fact, that D is closed and bounded and U is open (details are omitted). Thus the infimum on the right is not vacuous. For later use, we choose a fixed block set K_0 with $D \subset K_0 \subset U$.



$k \subset D$ $\qquad$ $D \subset K \subset U$

If $k \subset D$ and $K \supset D$ are block-sets, then $k \subset K$ and hence $\int_k f\,dA \leq \int_K f\,dA$. It then follows from Definition 3.5.2 that $\int_D f\,dA \leq \int_K f\,dA$, and hence the inequality $\leq$ holds in (7).

Let $\epsilon > 0$ be given. The boundary ∂D is a null set, according to Lemma 3.6. Hence there exists a block-set L around ∂D with area $A(L) \leq \epsilon$.



We may assume that $L \subset K_0$ (otherwise we replace L by its intersection with K_0). Let K denote the union $D \cup L$ and let k be the difference $D \setminus L$ together with its boundary. Then k and K are block-sets with $k \subset D \subset K \subset K_0$, and since $\int_k f\,dA \leq \int_D f\,dA \leq \int_K f\,dA$ we obtain

$$0 \leq \int_K f\,dA - \int_D f\,dA \leq \int_K f\,dA - \int_k f\,dA = \int_L f\,dA \leq \epsilon M$$

where $M = \sup_{K_0} f$. Since ϵ was arbitrary, (7) follows. $\square$

Thus for functions f as above the integral over D, which was defined by an approximation from the inside of D, can be approximated as well from the outside.

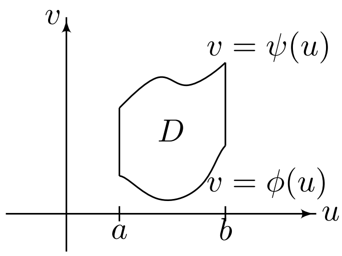
3.7 Double integrals

In the preceding section we have defined the notion of a plane integral over an elementary domain. In the simplest case when the elementary domain happens to be a rectangle, the integral was defined by two consecutive integrals (see equation (6)). In fact a similar formula can be given for a much larger class of elementary domains.

Let $\phi, \psi\colon [a, b] \to \mathbb{R}$ be smooth functions with $\phi(u) < \psi(u)$ for $u \in (a, b)$. The set

$$D = \{(u, v) \mid a \leq u \leq b,\ \phi(u) \leq v \leq \psi(u)\}$$

of points between the graphs of ϕ and ψ,



is an elementary domain.

Theorem 3.7. *The set D has the area*

$$A(D) = \int_a^b [\psi(u) - \phi(u)]\, du,$$

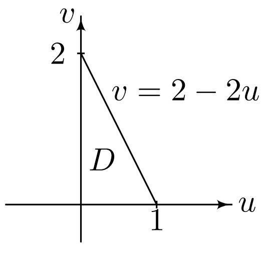
and the plane integral of a continuous function f over D is

$$\int_D f\, dA = \int_a^b \int_{\phi(u)}^{\psi(u)} f(u, v)\, dv\, du$$

We will not prove this. The formula for the area is well known from elementary calculus. When it comes to computation of plane integrals in practice, it is this formula which is used (not the definition given earlier). More complicated sets are treated by means of a disjoint division into subsets of this form (possibly with u and v interchanged).

Example 3.7.1 The triangle $D = \{(u, v) \mid 0 \leq u,\ 0 \leq v,\ 2u + v \leq 2\}$, has the form as above with inequalities

$$0 \leq u \leq 1, \quad 0 \leq v \leq 2 - 2u.$$



The set D is bounded above and below by the graphs of $\psi(u) = 2 - 2u$ and $\phi(u) = 0$. The area is then

$$A(D) = \int_0^1 (2 - 2u)\, du = 1.$$

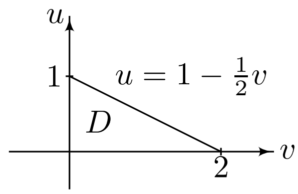
Furthermore, with $f(u,v)=v$, then

$$\int_D v\,dA = \int_0^1 \int_0^{2-2u} v\,dv\,du = \int_0^1 \frac{1}{2}(2-2u)^2\,du = \frac{2}{3}.$$

Notice that D can also be regarded as a set of the form as before, but with the inequalities

$$0 \le v \le 2, \quad 0 \le u \le 1 - \frac{1}{2}v$$

(that is, with interchanged roles of u and v).



Of course, the corresponding formulas for the area and the integral lead to the same results as above,

$$A(D) = \int_0^2 (1 - \frac{1}{2}v)\,dv = 1$$

and

$$\int_D v\,dA = \int_0^2 \int_0^{1-\frac{1}{2}v} v\,du\,dv = \int_0^2 v(1 - \frac{1}{2}v)\,dv = \frac{2}{3}.$$

3.8 Transformation of integrals

We shall need the important theorem of transformation of plane integrals, which is a generalization of the formula for substitution of variables in ordinary integrals. Let $\phi\colon W \to U$ be a diffeomorphism (see Definition 2.6.1), where $U, W \subset \mathbb{R}^2$ are open.

Theorem 3.8. *Assume that $D \subset \mathbb{R}^2$ is closed and bounded and contained in W. If D is an elementary domain, then so is its image $\phi(D) \subset U$. Moreover,*

$$\int_{\phi(D)} f\,dA = \int_D (f \circ \phi)\,|\det(D\phi)|\,dA$$

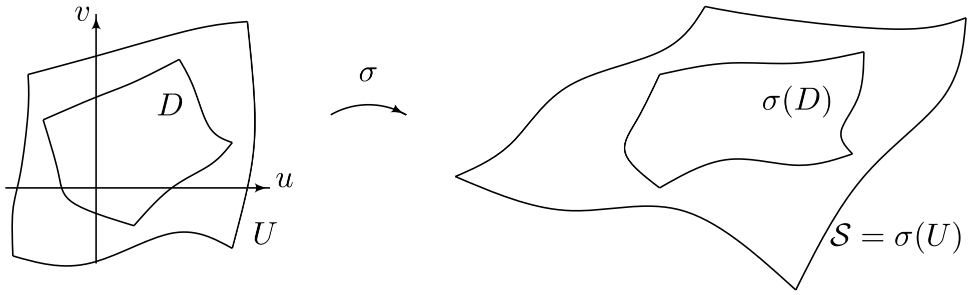
for $f\colon U \to \mathbb{R}$ continuous.

We shall not prove this theorem here. In particular, with $f = 1$ we obtain the following formula for the area

$$A(\phi(D)) = \int_D |\det(D\phi)|\,dA.$$

3.9 Surface area

Let $\sigma: U \to \mathbb{R}^3$ be a parametrized surface, and let $D \subset \mathbb{R}^2$ be an elementary domain, which is contained in U.



Definition 3.9. The *area* of the surface σ over D is

$$A(\sigma, D) = \int_D \|\sigma'_u \times \sigma'_v\| \, dA. \tag{8}$$

Recall that $\sigma'_u \times \sigma'_v$ is a normal vector to the tangent plane. Its length can be expressed by means of the first fundamental form as follows

$$\|\sigma'_u \times \sigma'_v\| = (EG - F^2)^{1/2}. \tag{9}$$

This identity is an immediate consequence of the following general rule of vector calculus:

$$\|a \times b\|^2 = \|a\|^2 \|b\|^2 - (a \cdot b)^2,$$

(see Appendix C).

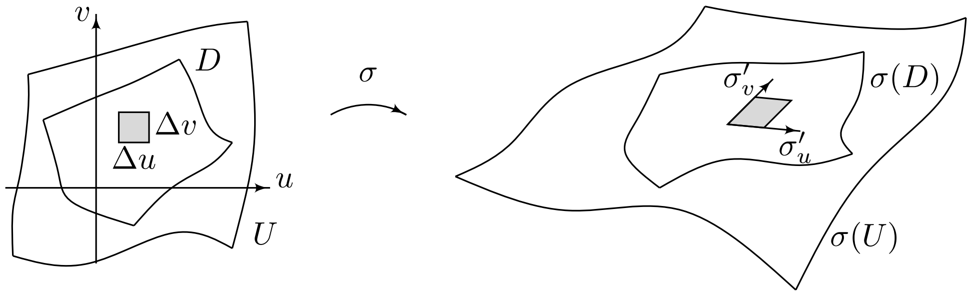
We often denote the area by $A(\sigma(D))$, although this is not quite legitimate, because in general the area depends on both σ and D, and not just their image $\sigma(D)$, unless some injectivity is assumed of σ.

Notice that if we consider the (x, y)-plane as the surface parametrized by $\sigma(u, v) = (u, v, 0)$, then $E = G = 1$ and $F = 0$ (see Example 3.4.1) and hence (8) reads

$$A(\sigma, D) = \int_D 1 \, dA,$$

by which we see that the new notion of area coincides with the previous one for plane sets.

The definition of area can be motivated by the following geometrical consideration, which is analogous to the motivation that was given for the definition of arc length. Consider a small rectangle in D with (u, v) as its lower left corner and with sides of length Δu and Δv. This rectangle is mapped approximately to the parallelogram in $\mathbb{R}^3$ placed at $\sigma(u, v)$ and with the vectors $\Delta u\,\sigma'_u$ and $\Delta v\,\sigma'_v$ as its sides, according to the first order (linear) approximation of σ.



The area of this parallelogram is

$$\|\Delta u\,\sigma'_u \times \Delta v\,\sigma'_v\| = \|\sigma'_u \times \sigma'_v\|\Delta u \Delta v.$$

Adding up all these areas and taking the limit $(\Delta u, \Delta v) \to (0,0)$ leads to the formula (8).

Further justification that our definition of surface area is reasonable can be found in the following theorem, which is analogous to Theorem 3.1.

Theorem 3.9. *Surface area is invariant under reparametrization.*

Proof. Let $\tau = \sigma \circ \phi \colon W \to \mathbb{R}^3$ be a reparametrization (see Section 2.5), and let $E \subset U$ be an elementary domain. Then $D = \phi^{-1}(E) \subset W$ is an elementary domain. The statement of the theorem amounts to the identity $A(\tau, D) = A(\sigma, E)$.

Since $\tau = \sigma \circ \phi$ we have $\tau(D) = \sigma(E)$, and the statement that these sets have the same area thus appears to be a tautology. However, as we noted earlier, in the definition (8) of the area, reference is made to both the parametrization and the domain, not just their image. For the area of $\tau(D)$, we have

$$A(\tau, D) = \int_D \|\tau'_s \times \tau'_t\|\, dA.$$

The claim is that this equals

$$A(\sigma, E) = \int_E \|\sigma'_u \times \sigma'_v\|\, dA.$$

We have from equation (7) in Section 2.8 that for $q \in W$

$$\tau'_s(q) \times \tau'_t(q) = \det(D\phi)(q)\ \sigma'_u(\phi(q)) \times \sigma'_v(\phi(q)).$$

Inserting this expression in the formula for $A(\tau(D))$ and using the substitution of variables in Theorem 3.8, we see that $A(\tau, D) = A(\sigma, E)$. $\square$

Example 3.9.1 As an illustration, let us compute the surface area of the sphere with radius 1. It is, as usual, parametrized by the spherical coordinates $\sigma(u,v)$ (see Example 1.2.2), where $(u,v) \in U = \mathbb{R}^2$. Let D be the rectangle where $-\frac{\pi}{2} \le u \le \frac{\pi}{2}$ and $-\pi \le v \le \pi$, then σ is injective on the interior of D. We found in Example 3.4.3 that the first fundamental form is given by $E = 1$, $F = 0$ and $G = \cos^2 u$, so that $(EG - F^2)^{1/2} = \cos u$. We therefore obtain the area

$$\int_{-\pi}^{\pi} \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \cos u \, du \, dv = 4\pi.$$

3.10 Exercises

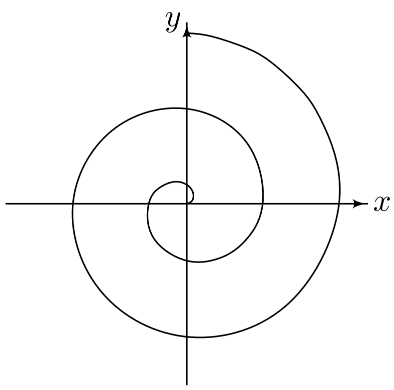
1 Let $\gamma(t) = (3t, 3t^2, 2t^3)$. Show that the speed of the curve is $\|\gamma'(t)\| = 3(1 + 2t^2)$, and determine the arc length of γ from $t = 0$ to t.

2 Let $\gamma(t) = (t, \frac{4}{3}t^{3/2}, t^2)$ for $t > 0$. Determine that value t_0 for which the length of γ from $t = t_0$ to $t = 1$ is equal to the length from $t = 1$ to $t = \frac{3}{2}$.

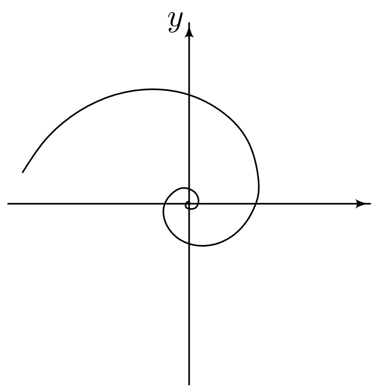
3 Let $\gamma(t) = (t \cos t, t \sin t)$, $t \in \mathbb{R}$. The section of the curve where $t \ge 0$ (drawn below) is called the *spiral of Archimedes*, because it was described in a book by Archimedes. Determine the arc length of the curve, measured from $t = 0$. The following formula can be used

$$\int \sqrt{1 + x^2} \, dx = \frac{x}{2}\sqrt{1 + x^2} + \frac{1}{2} \ln(x + \sqrt{1 + x^2}) + c.$$

4 The parametrized curve $\gamma(t) = (e^{ct} \cos t, e^{ct} \sin t)$, $t \in \mathbb{R}$, where $c > 0$ is a constant, is called a *logarithmic spiral.* Determine an arc length function $s(t)$ for γ, and show that $s(t)$ has a limit s_0 for $t \to -\infty$. Show that $s(t) - s_0$, which can be interpreted as the arc length from $\gamma(-\infty) = (0,0)$ to $\gamma(t)$, is proportional to $\|\gamma(t)\|$. This curve appears in nature, for example in the shape of snail shells. The natural appearance is explained by proportionality in the growth of the diameter of the shell and the length of the snail.



spiral of Archimedes



logarithmic spiral

5 Determine a unit speed parametrization of the line through $(0, 1, -3)$ and $(3, 3, 3)$.

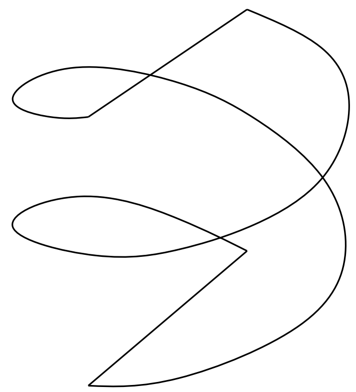
6 Show that the curve $\gamma(t) = (\cos t \sin t, \sin^2 t, \frac{3}{4}t)$ has constant speed, and determine a constant k for which the reparametrization $t \mapsto \gamma(kt)$ has unit speed.

7 Let $\gamma(t) = (\frac{2}{3}\cos^3 t, \frac{2}{3}\sin^3 t)$ for $t \in \mathbb{R}$. For which values of t is γ regular? Determine a direction preserving reparametrization with unit speed of the segment where $0 < t < \frac{\pi}{2}$. (Use the formula $\int \cos x \sin x \, dx = \frac{1}{2}\sin^2 x + c$.)

8 Let $\gamma(t) = (e^t \cos t, e^t \sin t)$, $t \in \mathbb{R}$, be the logarithmic spiral with $c = 1$ (see exercise 4). Determine a reparametrization $\beta(s)$, $s > 0$, with unit speed such that $\beta(s) \to (0, 0)$ for $s \to 0$ (the solution explains the name of the curve).

9 Let $\beta = \gamma \circ \phi \colon J \to \mathbb{R}^n$ be a direction preserving reparametrization of $\gamma \colon I \to \mathbb{R}^n$, where I and J are open intervals, and assume that *both* curves γ and β have unit speed. Show that there exists a constant c such that $\phi(s) = s + c$ for all $s \in J$. If $I =]a, b[$, then what is J? State and prove similar statements for a direction reversing reparametrization.

10 The surface $\sigma(u, v) = (u \cos v, u \sin v, av)$, $(u, v) \in \mathbb{R}^2$, with $a \neq 0$ constant, is called a *helicoid*. It resembles a (double) spiral staircase. The following figure shows one winding of the surface (u from -1 to 1, v from 0 to 2π)



Show that σ is regular at all $(u, v) \in \mathbb{R}^2$, and that the coefficients of the first fundamental form are $E = 1$, $F = 0$ and $G = a^2 + u^2$.

11 Consider a parametrized surface $\sigma \colon \mathbb{R}^2 \to \mathbb{R}^3$ for which $E = 1$, $F = 0$, $G = 1 + u^2$. Determine the arc length of the curve $t \mapsto \sigma(\frac{3}{4}, \frac{4}{5}t)$ from $t = 0$ to $t = 1$. Determine also the angle between the tangent vector at $t = 0$ of this curve and the tangent vector at $t = 0$ of the curve $t \mapsto \sigma(\frac{3}{4} + t, \frac{4}{5}t)$.

12 Show that the coefficients E, F and G for the surface of revolution (see page 35) $\sigma(u,v) = (f(u)\cos v, f(u)\sin v, g(u))$ are given by

$$E = f'(u)^2 + g'(u)^2, \quad F = 0, \quad G = f(u)^2$$

a. The curve $t \mapsto \sigma(u_0, t)$ on σ, where u_0 is constant, is called a *parallel curve*. Show that it has constant speed.

b. The curve $t \mapsto \sigma(t, v_0)$ on σ, where v_0 is constant, is called a *meridian*. Show that it has unit speed if the profile curve has unit speed.

13 Let $\sigma(u,v) = (\sqrt{1-u^2}\cos v, \sqrt{1-u^2}\sin v, u)$ for $(u,v) \in U =]-1,1[\times\mathbb{R}$. Show that σ is a regular parametrization of the unit sphere, minus the two poles. A map of the Earth based on this parametrization is called a *cylinder projection*. Explain! Determine E, F and G, and show that σ is *area preserving*, that is, the area $A(\sigma, D)$ equals the area of D for all elementary domains $D \subset]-1,1[\times[-\pi,\pi]$.

14 Draw the following sets in $\mathbb{R}^2$ and verify that they are elementary domains:
a. $D = [1,2] \times [0,1]$.
b. $D = \{(u,v) \mid 1 \le u \le 2, 0 \le v \le u-1\}$.
c. $D = \{(u,v) \mid 0 \le v \le 1, v+1 \le u \le 2\}$.
d. $D = \{(u,v) \mid 1 \le u \le 2, 0 \le v \le 2-u\}$.
Compute in each case the integral $\int_D u\,dA$. Repeat the computations but with the opposite order of the integrations with respect to u and v.

15 Let $\gamma(t) = (3t, 4t, 5\sqrt{1-t^2})$ for $t \in]-1,1[$.

a. Determine a reparametrization of γ with unit speed. (Use the formula $\int (1-t^2)^{-1/2}\,dt = \sin^{-1} t + c$, where $\sin^{-1}\colon]-1,1[\to]-\frac{\pi}{2}, \frac{\pi}{2}[$ is the inverse function of $\sin\colon]-\frac{\pi}{2}, \frac{\pi}{2}[\to]-1,1[$.)

b. Let

$$\sigma(u,v) = (3u, 4u, 5\sqrt{1-v^2}),$$

for $u \in \mathbb{R}$ and $-1 < v < 1$. Verify that γ can be realized as a parametrized curve on σ, and determine the coefficients of the tangent vector $\gamma'(t)$ with respect to the basis (σ'_u, σ'_v) for $T_{\mu(t)}\sigma$ when $v \neq 0$.

c. Determine E, F and G for σ, and write down a formula for the area $A(\sigma, D)$ where D is the rectangle $D = [0,1] \times [0, \frac{1}{2}]$.

16 Let σ be a surface of revolution (see Exercise 12). Let

$$D = \{(u,v) \mid a \le u \le b, -\pi \le v \le \pi\}$$

and assume that $[a,b]$ is contained in the interval on which the profile curve is defined.

Verify that the area of σ over D is given by

$$2\pi \int_a^b (f'(u)^2 + g'(u)^2)^{1/2} f(u)\, du.$$

Determine the area of the belt on a sphere of radius 1, where the latitude satisfies $|u| \leq \frac{\pi}{4}$. Determine also the area of the cap, where $\frac{\pi}{4} \leq u \leq \frac{\pi}{2}$.

17 Let σ denote the graph of a smooth function $z = h(x, y)$, and let $D \subset \mathbb{R}^2$ be an elementary domain. Verify the formula

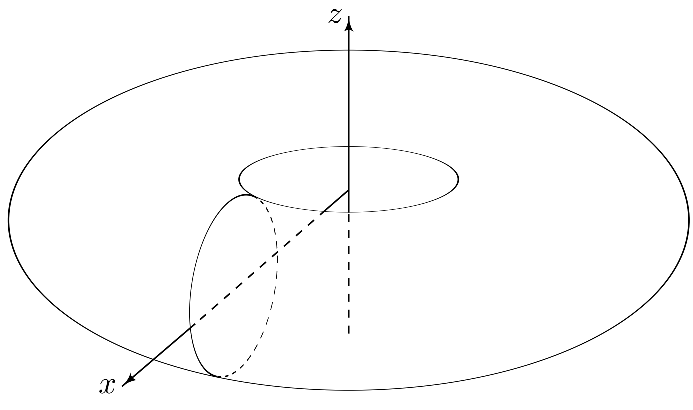
$$A(\sigma, D) = \int_D \sqrt{1 + (h'_x)^2 + (h'_y)^2}\, dA$$

(assuming that D is contained in the open set where h is defined).
Write down an integral formula for the area of that part of a sphere of radius 1 and centered at the origin, where $|x|$ and $|y|$ both are $\leq \frac{1}{\sqrt{2}}$ (disregard the assumption above about D). The computation of the integral is not quite simple. Instead the area can be determined from area of the cap (see Exercise 16) by a simple geometric consideration. How?

18 The *torus* is the surface of revolution whose profile curve is the circle in the xz-plane with radius r and center $(R, 0, 0)$, where $R > r$). It is parametrized by

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

Determine its total area.



19 Let $\sigma(u, v) = (u\cos v, u\sin v, v)$ (the helicoid, see Exercise 10). Determine the area of σ over $D = \{(u, v) \mid 0 \leq v \leq 1, v \leq u \leq 1\}$.

Chapter 4

Curvature

In this chapter we introduce and study a quantity, called curvature, which describes the shape of a curve in a given point. More precisely, it is a measure of the rate at which the curve is turning in the point. The number is associated with the second derivative $\gamma''(t)$ of a parametrization.

We shall also discuss the curvature of curves on a given surface. In particular, we introduce the so-called geodesic curvature, which describes the turning of a curve relative to the given surface containing the curve.

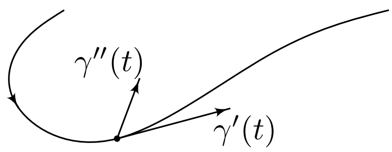
4.1 Curvature of plane curves

Let $\gamma\colon I \to \mathbb{R}^2$ be a regular parametrized curve.

Definition 4.1. The real number

$$\kappa(t) = \frac{\det[\gamma'(t)\,\gamma''(t)]}{\|\gamma'(t)\|^3} \tag{1}$$

is called the *curvature* of γ at t. Here $[\gamma'(t)\,\gamma''(t)]$ denotes the 2×2 matrix with columns $\gamma'(t)$ and $\gamma''(t)$.



The idea behind the definition is that the turning at t is described by the position and size of the vector $\gamma''(t)$ relative to $\gamma'(t)$. This relative position of the two vectors is described through their determinant, which measures the area of the parallelogram that they span. For example, if $\gamma''(t)$ has the same direction as $\gamma'(t)$, then the curve is not turning at all, and the determinant is zero. The power 3 in the denominator will be explained shortly by our desire to have a quantity independent of reparametrization (see Theorem 4.1).

Example 4.1.1 For a straight line with arbitrary parametrization, the vectors γ' and γ'' will both have the same direction as the line, hence their determinant is zero. Thus $\kappa = 0$ for a line.

Example 4.1.2 For a circle of radius r with counter clockwise parametrization $\gamma(t) = (r\cos t, r\sin t)$ we have

$$\gamma'(t) = (-r\sin t, r\cos t), \quad \gamma''(t) = (-r\cos t, -r\sin t)$$

and

$$\kappa(t) = \frac{\det[\gamma'(t)\,\gamma''(t)]}{\|\gamma'(t)\|^3} = \frac{1}{r}.$$

Similar computations show that the circle with the clockwise parametrization $\gamma(t) = (r\cos t, -r\sin t)$ has curvature $\kappa = -\frac{1}{r}$.

Example 4.1.3 For an ellipse

$$\gamma(t) = (a\cos t, b\sin t)$$

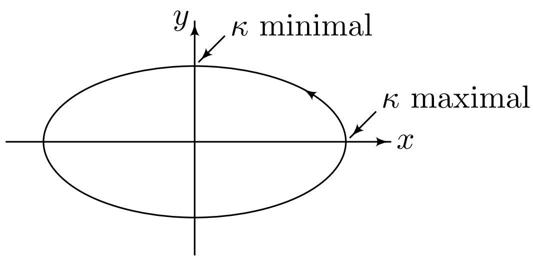
we have

$$\gamma'(t) = (-a\sin t, b\cos t), \quad \gamma''(t) = (-a\cos t, -b\sin t)$$

and

$$\kappa(t) = \frac{ab}{(a^2\sin^2 t + b^2\cos^2 t)^{3/2}}.$$

Say for example that $a > b$. Then κ attains its maximal value $\frac{a}{b^2}$ when $\sin t = 0$ (where the denominator is minimal), and it attains its minimal value $\frac{b}{a^2}$ when $\cos t = 0$ (where the denominator is maximal).



Example 4.1.4 Let $\gamma(t) = (t, h(t))$ be the graph of a smooth function h, defined on an open interval $I \subset \mathbb{R}$. Then $\gamma'(t) = (1, h'(t))$ and $\gamma''(t) = (0, h''(t))$, and we obtain

$$\kappa(t) = \frac{h''(t)}{(1 + h'(t)^2)^{3/2}}.$$

In particular if $h'(t) = 0$ then $\kappa(t) = h''(t)$.

Theorem 4.1. *The curvature of a plane curve is unchanged under a direction-preserving reparametrisation, and it is multiplied by* -1 *under a direction-reversing reparametrization.*

Proof. Let $\beta(u) = \gamma(\phi(u))$ denote the reparametrization, and let $\epsilon = \pm 1$ denote the sign of ϕ'. Then

$$\beta'(u) = \phi'(u)\gamma'(\phi(u)) \tag{2}$$

and

$$\beta''(u) = \phi''(u)\gamma'(\phi(u)) + \phi'(u)^2\gamma''(\phi(u)). \tag{3}$$

Hence

$$\det[\beta'(u)\,\beta''(u)] = \phi'(u)^3 \det[\gamma'(\phi(u))\,\gamma''(\phi(u))]$$

and

$$\|\beta'(u)\| = |\phi'(u)|\,\|\gamma'(\phi(u))\|.$$

By insertion in the definition (1), applied to the curve β, we see that the curvature of β at u is $\epsilon\kappa(\phi(u))$. $\square$

Notice that the power 3 in the denominator of (1) was crucial in the preceding proof.

4.2 Curvature of unit speed curves

For a unit speed curve the expression (1) for the curvature becomes simpler. Notice that unit speed is not a serious limitation because of Theorems 3.3 and 4.1.

Let $\gamma\colon I \to \mathbb{R}^2$ be a unit speed curve. As usual, the variable is then denoted by s. Let $\widehat{\gamma'}(s)$ denote the normal vector of $\gamma'(s)$ (see Appendix C), which is the unit vector perpendicular to $\gamma'(s)$ and pointing to the left.

Theorem 4.2. *For a curve with unit speed*

$$\gamma'' = \kappa\widehat{\gamma'}. \tag{4}$$

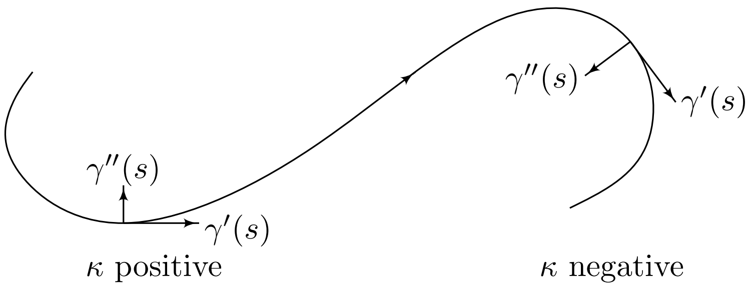
In particular, it follows that $\kappa = \pm\|\gamma''\|$, where the sign is $+$ if γ'' and $\widehat{\gamma'}$ have the same direction, and $-$ if they have opposite directions.

Proof. According to the lemma below $\gamma''(s)$ is perpendicular to $\gamma'(s)$, hence a scalar multiple of $\widehat{\gamma'}(s)$. The scalar is given by

$$\widehat{\gamma'} \cdot \gamma'' = \det[\gamma'\,\gamma''] = \kappa.$$

This proves (4). $\square$

Notice that if $\kappa > 0$ then γ'' and $\widehat{\gamma}'$ have the same direction and the curve turns towards the left, and if $\kappa < 0$ they have opposite direction and the curve turns to the right.



Lemma 4.2. *Let $F(t) \in \mathbb{R}^n$ be a smooth function of $t \in I \subset \mathbb{R}$ with $\|F(t)\| = 1$ for all t. Then $F(t) \cdot F'(t) = 0$ for all t.*

Proof. We shall differentiate the expression

$$F(t) \cdot F(t) = 1.$$

Observe that the ordinary rule for differentiation of products also holds for the differentiation of a dot product, that is, if f and g are differentiable maps $I \to \mathbb{R}^n$, then

$$(f \cdot g)' = f' \cdot g + f \cdot g'.$$

Applying this rule we obtain $F' \cdot F + F \cdot F' = 0$ and hence $F \cdot F' = 0$ as claimed. □

Notice that Theorem 4.2 suggests a way to determine a plane unit speed curve from its curvature function $\kappa(s)$. With $\gamma(s) = (x(s), y(s))$, equation (4) is equivalent with the system of differential equations $x'' = -\kappa y'$ and $y'' = \kappa x'$. By solving this system we can determine x' and y' (up to some constants), and after an integration we obtain x and y (up to further constants). A simple example of this procedure is given in the following proof.

Corollary 4.2. *A regular parametrized curve is part of a line if and only if its curvature is zero everywhere.*

Proof. We may assume that the curve has unit speed. Assume that $\kappa(s) = 0$ for all s, then $\gamma''(s) = 0$ by (4). Integrating twice we obtain $\gamma(s) = p + sq$ where p and q are constant vectors.

The statement 'only if' was seen in Example 4.1.1. □

4.3 The tangent angle

Any unit vector $w \in \mathbb{R}^2$ can be written in the form $w = (\cos\theta, \sin\theta)$, where the angle $\theta \in \mathbb{R}$ is determined up to addition of integral multiples of 2π. In particular, if w is the tangent direction $\gamma'(t)/\|\gamma'(t)\|$ of a regular plane curve, we call θ a *tangent angle* at t. Viewed as a function of t, we call θ a *tangent angle function*.

Example 4.3.1 The parametrized circle $\gamma(t) = (r\cos t, r\sin t)$ has tangent angle function $\theta(t) = t + \frac{\pi}{2}$, because

$$\gamma'(t)/\|\gamma'(t)\| = (-\sin t, \cos t) = (\cos(t + \frac{\pi}{2}), \sin(t + \frac{\pi}{2})).$$

Example 4.3.2 Consider the curve $\gamma(t) = (t, t^2)$, where $t \in \mathbb{R}$. It has the tangent vector $\gamma'(t) = (1, 2t)$. Since the first coordinate is positive, we can determine a tangent angle as $\theta(t) = \tan^{-1}(2t)$.

Because of the ambiguity in the choice of θ, it is not obvious that a tangent angle can be chosen which depends smoothly on t. The following lemma, when applied to $w(t) = \gamma'(t)/\|\gamma'(t)\|$, shows that this is the case.

Lemma 4.3. *Let $w(t)$ be a unit vector in $\mathbb{R}^2$ depending smoothly on a parameter t in an open interval $I \subset \mathbb{R}$. There exists a smooth map $\theta\colon I \to \mathbb{R}$ such that*

$$w(t) = (\cos\theta(t), \sin\theta(t)) \tag{5}$$

for all $t \in I$.

Proof. Write $w(t) = (u(t), v(t))$ and notice for motivation that if (5) is valid for some function θ, then $(u', v') = (-\theta'\sin\theta, \theta'\cos\theta)$ and hence $uv' - vu' = \theta'(\cos^2\theta + \sin^2\theta) = \theta'$.

Choose an arbitrary initial value $t_0 \in I$, and an angle θ_0 such that $w(t_0) = (\cos\theta_0, \sin\theta_0)$. Define a smooth function by

$$\theta(t) = \theta_0 + \int_{t_0}^{t} uv' - vu'\,dt,$$

then $\theta(t_0) = \theta_0$ and $\theta' = uv' - vu'$. We claim that this function satisfies (5). In order to show this identity of unit vectors in $\mathbb{R}^2$, it suffices to show that $w \cdot (\cos\theta, \sin\theta) = 1$, since otherwise the dot product would be strictly smaller (see (A.1)).

From the identity $u^2 + v^2 = 1$ we obtain $uu' + vv' = 0$. By simple computations we then derive

$$(u\cos\theta)' = u'\cos\theta - u\sin\theta(uv' - vu') = u'\cos\theta - v'\sin\theta$$
$$(v\sin\theta)' = v'\sin\theta + v\cos\theta(uv' - vu') = v'\sin\theta - u'\cos\theta.$$

It follows that $(u\cos\theta + v\sin\theta)' = 0$, hence the expression in the bracket is constant. At $t = t_0$ its value is 1. Hence $w \cdot (\cos\theta, \sin\theta) = 1$ as desired. □

Theorem 4.3. *Assume that $\theta(s)$ is a smooth tangent angle for a plane curve $\gamma(s)$ with unit speed. Then the curvature of γ at s is given by*

$$\kappa(s) = \theta'(s).$$

Proof. From $\gamma'(s) = (\cos\theta(s), \sin\theta(s))$ we derive

$$\gamma''(s) = (-\theta'(s)\sin\theta(s), \theta'(s)\cos\theta(s))$$

and $\kappa(s) = \det[\gamma'(s)\,\gamma''(s)] = \theta'(s)$. ☐

Thus the curvature is the rate of change of the tangent angle. In Example 4.3.1 with $r = 1$ (so that there is unit speed), we have $\theta'(t) = 1$, which matches with the curvature 1 of the circle.

4.4 Curvature of space curves

Let $\gamma\colon I \to \mathbb{R}^3$ be a regular parametrized curve.

Definition 4.4. The non-negative number

$$\kappa(t) = \frac{\|\gamma'(t) \times \gamma''(t)\|}{\|\gamma'(t)\|^3}$$

is called the *curvature* of γ at t. For a unit speed curve it is

$$\kappa(s) = \|\gamma''(s)\|. \tag{6}$$

The simpler expression for a curve with unit speed is derived from the fact that in this case $\gamma''(s)$ is perpendicular to the unit vector $\gamma'(s)$ (by Lemma 4.2) and hence

$$\|\gamma' \times \gamma''\| = \|\gamma''\|.$$

Note that $\|\gamma'(t) \times \gamma''(t)\|$ is easily computed by means of Appendix C (iii).

The motivation is similar to the one given in Section 4.1 for plane curves. It will be shown below that the curvature κ is unchanged by reparametrisation. For a unit speed curve (6) shows that κ describes the rate of change of the direction of the curve. Notice that the conclusions of Example 4.1.1 and Corollary 4.2 are valid for space curves as well, with similar proofs.

Notice however that in contrast with the situation for plane curves in Section 4.1, the curvature of a space curve is always ≥ 0. This is related to the fact that the curvature for a space curve does not contain information about the direction to which the curve is turning. For a plane curve there are only two possibilities, left and right, which can be determined by the sign of the curvature, but for a space curve there are infinitely many possibilities, and it would be impossible to distinguish them just by a sign.

In connection with this, it should be remarked that if we apply the present definition to a plane curve, viewed as a space curve in the xy-plane, we obtain the absolute value of the previous definition. Indeed, if $\gamma(t) = (x(t), y(t), 0)$ then

$$\gamma' \times \gamma'' = (x', y', 0) \times (x'', y'', 0) = (0, 0, \det \begin{pmatrix} x' & x'' \\ y' & y'' \end{pmatrix})$$

and hence $\|\gamma' \times \gamma''\| = \left| \det \begin{pmatrix} x' & x'' \\ y' & y'' \end{pmatrix} \right|$.

The notion of curvature for space curves is thus more primitive than that for plane curves. This is also reflected when the following theorem is compared with Theorem 4.1.

Theorem 4.4. *The curvature of a space curve is unchanged under reparametrisation.*

Proof. We use the notation in the proof of Theorem 4.1 (but now applied to a space curve). It follows from (2) and (3) that

$$\beta'(u) \times \beta''(u) = \phi'(u)^3\, \gamma'(\phi(u)) \times \gamma''(\phi(u)). \tag{7}$$

This equation together with (2) implies the theorem. □

Example 4.4.1 Let

$$\gamma(t) = (\lambda t, r\cos(\omega t), r\sin(\omega t))$$

be a helix, as in Example 3.1.3. We find

$$\gamma'(t) = (\lambda, -r\omega\sin(\omega t), r\omega\cos(\omega t))$$
$$\gamma''(t) = (0, -r\omega^2\cos(\omega t), -r\omega^2\sin(\omega t))$$

with $\|\gamma'(t)\| = \sqrt{\lambda^2 + r^2\omega^2}$. Furthermore

$$\gamma'(t) \times \gamma''(t) = (r^2\omega^3, \lambda r\omega^2\sin(\omega t), -\lambda r\omega^2\cos(\omega t))$$

with $\|\gamma'(t) \times \gamma''(t)\| = r\omega^2\sqrt{r^2\omega^2 + \lambda^2}$. Hence

$$\kappa(t) = \frac{r\omega^2}{r^2\omega^2 + \lambda^2}.$$

Notice that the curvature is constant, which is reasonable from a geometric point of view, because the helix has the same shape everywhere.

4.5 Torsion

For space curves with non-zero curvature it is possible to define a 'higher curvature' called torsion, which is associated with the third derivative γ'''. It describes the 'twisting' of the curve. For a plane curve, regarded as a curve in $\mathbb{R}^3$, the torsion is 0.

Let $\gamma\colon I \to \mathbb{R}^3$ be a regular parametrized curve with curvature $\kappa(t)$.

Definition 4.5. Let $t \in I$ and assume that $\kappa(t) \neq 0$. The number

$$\tau(t) = \frac{\det[\gamma'(t)\,\gamma''(t)\,\gamma'''(t)]}{\|\gamma'(t) \times \gamma''(t)\|^2}$$

is called the *torsion* of γ at t.

Notice the resemblance of this formula with (1). The denominator is $\kappa(t)^2\|\gamma'(t)\|^6$, which is non-zero by assumption. Motivation for the definition will be given in the following section.

Example 4.5.1 For a curve which is contained in a fixed plane in $\mathbb{R}^3$, the three vectors $\gamma'(t)$, $\gamma''(t)$ and $\gamma'''(t)$ will all be parallel to this plane. Hence they are linearly dependent and their determinant is zero. Therefore $\tau = 0$ (if it is defined).

Example 4.5.2 For the helix of Example 4.4.1 we find

$$\gamma'''(t) = (0, r\omega^3 \sin(\omega t), -r\omega^3 \cos(\omega t))$$

and hence the determinant $\det[\gamma'\,\gamma''\,\gamma''']$ is

$$\det \begin{pmatrix} \lambda & 0 & 0 \\ -r\omega \sin(\omega t) & -r\omega^2 \cos(\omega t) & r\omega^3 \sin(\omega t) \\ r\omega \cos(\omega t) & -r\omega^2 \sin(\omega t) & -r\omega^3 \cos(\omega t) \end{pmatrix} = \lambda r^2 \omega^5.$$

Hence

$$\tau = \frac{\lambda\omega}{r^2\omega^2 + \lambda^2}.$$

Again we obtain a constant, which is reasonable by the same argument as in Example 4.4.1.

Theorem 4.5. *The torsion of a space curve is unchanged under a reparametrisation.*

Proof. It follows by differentiation of equation (3) in the proof of Theorem 4.1 that

$$\beta'''(u) = \phi'''(u)\gamma'(\phi(u)) + 3\phi''(u)\phi'(u)\gamma''(\phi(u)) + \phi'(u)^3\gamma'''(\phi(u)). \qquad (8)$$

From (2), (3) and (8) we see that

$$\det[\beta'(u)\,\beta''(u)\,\beta'''(u)] = \phi'(u)^6 \det[\gamma'(\phi(u))\,\gamma''(\phi(u))\,\gamma'''(\phi(u))].$$

The theorem follows from this, combined with (7). □

Notice that the torsion of a curve is unchanged also when the direction of the curve is reversed. The sign of the torsion allows us to separate space curves with non-zero curvature and torsion in two kinds, 'right' and 'left'. For example, a helix for which λ and ω have the same sign is called a *right helix* (compare the thread of a conventional screw) and a helix for which they have opposite signs is called a *left helix*.

4.6 The osculating plane and the binormal vector

The geometric significance of the torsion will now be explained. As before, let $\gamma: I \to \mathbb{R}^3$ be a regular parametrized curve with non-zero curvature $\kappa(t)$. Then $\gamma'(t)$ and $\gamma''(t)$ are linearly independent vectors in $\mathbb{R}^3$. The plane through $\gamma(t)$ with directions spanned by these two vectors is called the *osculating* plane. It can be viewed as the plane in $\mathbb{R}^3$ to which the curve comes closest in the vicinity of $\gamma(t)$ (osculare in Latin means to kiss), because of the Taylor approximation of order two

$$\gamma(t+\Delta t) \simeq \gamma(t) + \Delta t\gamma'(t) + \frac{1}{2}(\Delta t)^2\gamma''(t),$$

where the right hand side belongs to the osculating plane for all Δt. We will show that the torsion describes the rate of change of the osculating plane.

It follows from equations (2) and (3) that the osculating plane is unchanged if the curve is reparametrized. Because of Theorem 4.5, we may therefore assume that the given curve has unit speed. We introduce the notation $\mathbf{t}(s) = \gamma'(s)$ for the unit tangent vector. Keeping the assumption that $\kappa(s) \neq 0$, let

$$\mathbf{n}(s) = \frac{\gamma''(s)}{\|\gamma''(s)\|}.$$

This unit vector, called the *principal normal*, is orthogonal to $\mathbf{t}(s)$ by Lemma 4.2. The unit vector

$$\mathbf{b}(s) = \mathbf{t}(s) \times \mathbf{n}(s),$$

which is called the *binormal* of the curve, is orthogonal to $\mathbf{t}(s)$ as well.

Notice that the vectors $\mathbf{t}(s)$ and $\mathbf{n}(s)$ span the directions of the osculating plane, and that the binormal $\mathbf{b}(s)$ is normal to the osculating plane. It follows that the rate of change of the osculating plane is expressed by the size of the derivative $\mathbf{b}'(s)$. The following result shows that this is exactly what the torsion τ measures.

Theorem 4.6. *For a curve in $\mathbb{R}^3$ with unit speed and non-zero curvature we have*

$$\mathbf{b}' = -\tau\mathbf{n}.$$

In particular, $\tau = \pm\|\mathbf{b}'\|$.

Proof. We first show that $\mathbf{b}'$ is proportional to $\mathbf{n}$. For this it suffices to show that it is orthogonal to $\mathbf{t}$ and $\mathbf{b}$. That $\mathbf{b}' \perp \mathbf{b}$ is immediate from Lemma 4.2. By differentiation of the equation $\mathbf{b} \cdot \mathbf{t} = 0$ we obtain that $\mathbf{b}' \cdot \mathbf{t} + \mathbf{b} \cdot \mathbf{t}' = 0$. Hence $\mathbf{b}' \perp \mathbf{t}$ if and only if $\mathbf{b} \perp \mathbf{t}'$. By the definition of $\mathbf{n}$ we have $\mathbf{t}' = \kappa\mathbf{n}$, hence $\mathbf{b} \perp \mathbf{t}'$ follows from $\mathbf{b} \perp \mathbf{n}$.

We thus conclude that $\mathbf{b}' = c\mathbf{n}$ for some constant c, which we now claim is $-\tau$. Since $\gamma'' = \kappa\mathbf{n}$ we have $\gamma''' = (\kappa\mathbf{n})' = \kappa'\mathbf{n} + \kappa\mathbf{n}'$. Then

$$\det[\gamma' \, \gamma'' \, \gamma'''] = (\gamma' \times \gamma'') \cdot \gamma''' = (\mathbf{t} \times \kappa\mathbf{n}) \cdot (\kappa'\mathbf{n} + \kappa\mathbf{n}') = \kappa^2(\mathbf{t} \times \mathbf{n}) \cdot \mathbf{n}'.$$

It follows that

$$\tau = (\mathbf{t} \times \mathbf{n}) \cdot \mathbf{n}' = \mathbf{b} \cdot \mathbf{n}'. \tag{9}$$

From $\mathbf{b} \cdot \mathbf{n} = 0$ we obtain by differentiation that $\mathbf{b}' \cdot \mathbf{n} + \mathbf{b} \cdot \mathbf{n}' = 0$, hence $\mathbf{b} \cdot \mathbf{n}' = -\mathbf{b}' \cdot \mathbf{n} = -c\mathbf{n} \cdot \mathbf{n} = -c$, and the proof is finished. □

We read from Theorem 4.6 that the absolute size of $\tau(s)$ measures the rate of change of the osculating plane. Moreover, the sign determines the direction to which the osculating plane is turning, according to the following rule. Follow the curve with your right hand such that the index finger is in the tangent direction $\mathbf{t}$ and the thumb is in the normal direction $\mathbf{n}$, then if $\tau > 0$, the hand will be turning as a right screw (the middle finger, pointing in direction $-\mathbf{b}$, turns towards the thumb $\mathbf{n}$).

Corollary 4.6. *A regular space curve with $\kappa \neq 0$ is contained in a fixed plane if and only if $\tau = 0$ everywhere.*

Proof. Assume $\tau = 0$. From the preceding theorem we have $\mathbf{b}'(s) = 0$, hence $\mathbf{b}$ is a constant vector. Since $\mathbf{t}(s) \perp \mathbf{b}$ we have $\gamma'(s) \cdot \mathbf{b} = 0$ for all s. Since $(\gamma \cdot \mathbf{b})' = \gamma' \cdot \mathbf{b}$ we conclude that $\gamma \cdot \mathbf{b}$ is a constant c. Hence $\gamma(s)$ belongs to the plane $\{\xi \in \mathbb{R}^3 \mid \xi \cdot \mathbf{b} = c\}$ for all s. The converse implication was seen in Example 4.5.1. □

4.7 The Frenet formulas

The three vectors $\mathbf{t}(s), \mathbf{n}(s), \mathbf{b}(s)$ constitute a positively ordered orthonormal basis for $\mathbb{R}^3$ (depending on s), which is called the *moving frame of Frenet* for the curve. We have seen that $\mathbf{t}' = \kappa\mathbf{n}$ and $\mathbf{b}' = -\tau\mathbf{n}$. It is of interest also to determine $\mathbf{n}'$. We collect all three formulas in a single theorem.

Theorem 4.7. *For a curve with unit speed and non-zero curvature*

$$\begin{aligned} \mathbf{t}' &= \kappa \mathbf{n} \\ \mathbf{n}' &= -\kappa \mathbf{t} + \tau \mathbf{b} \\ \mathbf{b}' &= -\tau \mathbf{n} \end{aligned}$$

Proof. Since $\mathbf{t}, \mathbf{n}, \mathbf{b}$ is an orthonormal basis for $\mathbb{R}^3$ we have

$$\mathbf{n}' = (\mathbf{n}' \cdot \mathbf{t})\, \mathbf{t} + (\mathbf{n}' \cdot \mathbf{n})\, \mathbf{n} + (\mathbf{n}' \cdot \mathbf{b})\, \mathbf{b}.$$

By Lemma 4.2, $\mathbf{n}' \cdot \mathbf{n} = 0$, and in (9) we saw that $\mathbf{b} \cdot \mathbf{n}' = \tau$. Finally, from $\mathbf{n} \cdot \mathbf{t} = 0$ we derive by differentiation that

$$\mathbf{n}' \cdot \mathbf{t} = -\mathbf{n} \cdot \mathbf{t}' = -\mathbf{n} \cdot \kappa \mathbf{n} = -\kappa. \quad \square$$

The formulas in this theorem are called the *formulas of Frenet* (or of Frenet-Serret). Since each of the functions $\mathbf{t}$, $\mathbf{n}$ and $\mathbf{b}$ have three coordinates, this is essentially a linear system of 9 first order differential equations in these coordinates. By solving this system one can (at least in principle) determine a curve from its curvature $\kappa(s)$ and torsion $\tau(s)$, up to integration constants.

4.8 Curvature of curves on a surface

We will now study some refined notions of curvature for curves which are contained in a given surface. Let $\sigma: U \to \mathbb{R}^3$ be a parametrized surface, and let $\gamma: I \to \mathbb{R}^3$ be a parametrized curve on σ. This means (see Section 2.4) that $\gamma = \sigma \circ \mu$ where $\mu: I \to U$ is a plane curve. Assume that γ is regular and that σ is regular at $\mu(t)$ for all $t \in I$. We denote by

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

the unit normal vector of σ (see Section 2.8), and put

$$\mathbf{m}(t) = \mathbf{N}(\mu(t)),$$

the unit normal vector of σ *along* γ.

Definition 4.8. The numbers

$$\kappa_g(t) = \frac{\det[\gamma'(t)\, \gamma''(t)\, \mathbf{m}(t)]}{\|\gamma'(t)\|^3} \quad \text{and} \quad \kappa_n(t) = \frac{\gamma''(t) \cdot \mathbf{m}(t)}{\|\gamma'(t)\|^2}$$

are called, respectively, the *geodesic* (or *tangential*) *curvature* and the *normal curvature* of γ at t with respect to σ. For a unit speed curve, they are

$$\kappa_g(s) = \gamma''(s) \cdot \mathbf{u}(s) \quad \text{and} \quad \kappa_n(s) = \gamma''(s) \cdot \mathbf{m}(s), \tag{10}$$

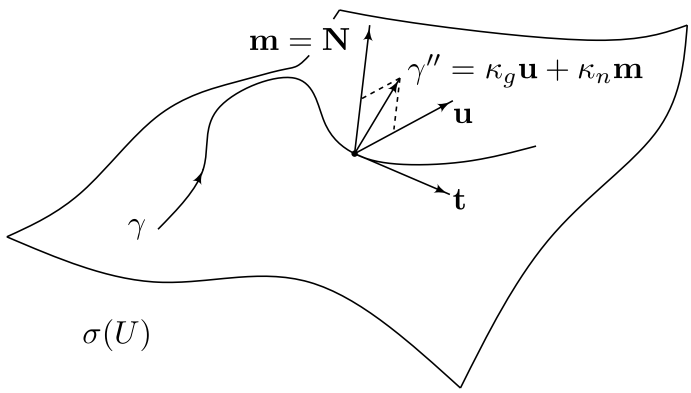
where $\mathbf{u}(s) = \mathbf{m}(s) \times \mathbf{t}(s)$ with $\mathbf{t}(s) = \gamma'(s)$. The vector $\mathbf{u}(s)$ is called the *tangent normal* of γ with respect to σ.

The formulas (10) for unit speed curves are easily obtained from the general definitions. For a unit speed curve on σ, the three vectors

$$\mathbf{t}(s), \mathbf{u}(s), \mathbf{m}(s)$$

again consitute a positively ordered orthonormal basis for $\mathbb{R}^3$, this is called the *moving frame of Darboux.* The first two basis vectors span the tangent space $T_{\mu(s)}\sigma$. Since in this case $\gamma''(s)$ is orthogonal to $\mathbf{t}(s) = \gamma'(s)$, it follows from (10) that the decomposition of $\gamma''(s)$ according to this basis reads

$$\gamma''(s) = \kappa_g(s)\,\mathbf{u}(s) + \kappa_n(s)\mathbf{m}(s). \tag{11}$$



Since $\kappa_g(s)\,\mathbf{u}(s) \in T_{\mu(s)}\sigma$ and $\kappa_n(s)\mathbf{m}(s) \perp T_{\mu(s)}\sigma$, this explains the reason behind the terms 'tangential' and 'normal' curvature for κ_g and κ_n.

Theorem 4.8. *The geodesic curvature κ_g is unchanged under a direction-preserving reparametrisation of γ, and multiplied by -1 under a direction-reversing reparametrization. The normal curvature κ_n is unchanged in both cases.*

Both κ_g and κ_n are unchanged under orientation-preserving reparametrisations of σ, and multiplied by -1 under orientation-reversing reparametrizations.

Proof. The statements concerning reparametrization of γ are easily seen from (2) and (3), and the statements concerning reparametrization of σ are straightforward, since σ is only represented in the definitions through the presence of $\mathbf{N}$ (see Section 2.8 for the notion of orientation). $\square$

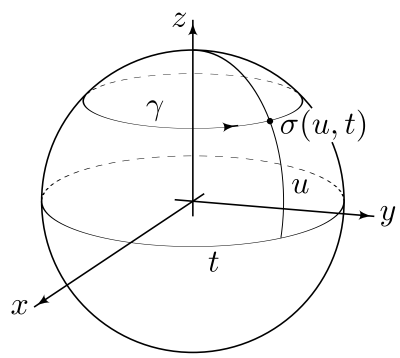
Corollary 4.8. *The curvature, the geodesic curvature and the normal curvature of γ satisfy*

$$\kappa^2 = \kappa_g^2 + \kappa_n^2. \tag{12}$$

Proof. We may assume the curve has unit speed by the preceding theorem. The equation then follows from (11) by the theorem of Pythagoras. $\square$

Example 4.8.1 A plane curve regarded as a space curve $\gamma(t) = (x(t), y(t), 0)$ (as in Section 4.3) is a curve on the surface $\sigma(u, v) = (u, v, 0)$ (the xy-plane). The normal vector of this surface is $\mathbf{N} = (0, 0, 1)$. It is easily seen that the definition of κ_g in this case is identical with the original definition of curvature of plane curves in Section 4.1, and $\kappa_n = 0$.

Example 4.8.2 We will compute the curvatures κ_g and κ_n of a circle on a sphere of radius 1. Such a curve is the intersection of the sphere and a plane. It is called a *great circle* if the plane passes through the center of the sphere, otherwise a *small circle.* For simplicity we assume that the plane is horizontal (this is not a serious restriction, as it can be arranged by a spatial rotation around the center of the sphere).



With standard spherical coordinates the circle can be parametrized by

$$t \mapsto \gamma(t) = \sigma(u, t) = (\cos u \cos t, \cos u \sin t, \sin u)$$

with a fixed latitude u. The radius of the circle is $\cos u$, hence it has curvature $\kappa = \frac{1}{\cos u}$ (see Example 4.1.2). We find

$$\gamma'(t) = (-\cos u \sin t, \cos u \cos t, 0), \quad \gamma''(t) = (-\cos u \cos t, -\cos u \sin t, 0),$$

and from Example 2.8.1 we have

$$\mathbf{m}(t) = -\sigma(u, t).$$

Hence

$$\det[\gamma'(t)\,\gamma''(t)\,\mathbf{m}(t)]$$
$$= \det\begin{pmatrix} -\cos u\sin t & -\cos u\cos t & -\cos u\cos t \\ \cos u\cos t & -\cos u\sin t & -\cos u\sin t \\ 0 & 0 & -\sin u \end{pmatrix} = -\cos^2 u\sin u$$

and

$$\gamma''(t)\cdot\mathbf{m}(t) = \cos^2 u.$$

We conclude that

$$\kappa_g(t) = -\tan u \quad \text{and} \quad \kappa_n(t) = 1.$$

We can verify $\kappa^2 = \kappa_g^2 + \kappa_n^2$ by the formula $\frac{1}{\cos^2 u} = \tan^2 u + 1$.

4.9 Interpretation of normal curvature

A curve which is required to be on a given surface has to follow the shape of the surface, and is therefore forced to some amount of curvature. The interpretation of the normal curvature κ_n is exactly, that it is this part of κ (in the decomposition (12)) which the curve is forced to have by being on σ. This interpretation is supported by the following theorem.

Theorem 4.9. *Given a point $p = (u_0, v_0) \in U$ and a non-zero vector $w_0 \in T_p\sigma$. All parametrized curves $\gamma = \sigma\circ\mu$ on σ with $\mu(t_0) = p$ and $\gamma'(t_0) = w_0$ have the same normal curvature $\kappa_n(t_0)$.*

Notice that by (12) this common value of κ_n is then a lower bound for the curvature κ for all such curves.

Part of the proof of the theorem is separated in the following lemma.

Lemma 4.9. *Let $\gamma = \sigma\circ\mu$ be a parametrized curve on σ and let $\mathbf{m}(t) = \mathbf{N}(\mu(t))$ for $t \in I$. Let $t \in I$ be given, and let $p = \mu(t) \in U$ and $(a, b) = \mu'(t) \in \mathbb{R}^2$. Then*

$$\gamma'(t) = a\sigma'_u(p) + b\sigma'_v(p) \tag{13}$$
$$\mathbf{m}'(t) = a\mathbf{N}'_u(p) + b\mathbf{N}'_v(p). \tag{14}$$

Proof of the lemma. Equation (13) was established by means of the chain rule in Lemma 2.4, and Equation (14) is obtained in exactly the same manner. □

Proof of the theorem. Since $\gamma'(t)$ belongs to the tangent space at $\mu(t)$, we have $\gamma'(t)\cdot\mathbf{m}(t) = 0$ for all t. By differentiation $\gamma''(t)\cdot\mathbf{m}(t)+\gamma'(t)\cdot\mathbf{m}'(t) = 0$, from which it follows that

$$\kappa_n(t) = \frac{\gamma''(t)\cdot\mathbf{m}(t)}{\|\gamma'(t)\|^2} = -\frac{\gamma'(t)\cdot\mathbf{m}'(t)}{\|\gamma'(t)\|^2}. \tag{15}$$

From (13) we see that the coordinates a and b are the same for all curves with tangent vector $\gamma'(t_0) = w_0$, and from (14) we then see that $\mathbf{m}'(t_0)$ is the same for all such curves. It then follows from (15) that $\kappa_n(t_0)$ is the same for all such curves. $\square$

Example 4.9.1 For circles on the unit sphere we found in Example 4.8.2 that $\kappa_n = 1$. Since every tangent direction w_0 in every point is the tangent direction of some circle on the sphere (in fact, of a unique great circle), we conclude from the preceding theorem that $\kappa_n = 1$ at all points on all curves on the sphere.

It follows from Theorem 4.9 that the normal curvature is a property of the surface rather than of the curve γ, and the following definition makes sense.

Definition 4.9. Let p and w_0 be as in Theorem 4.9. The *normal curvature* of σ in p with direction w_0 is the normal curvature $\kappa_n(t_0)$ of any parametrized curve $\gamma = \sigma \circ \mu$ on σ with $\mu(t_0) = p$ and $\gamma'(t_0) = w_0$.

It follows from Theorem 4.8 that the normal curvature of σ is unchanged under reparametrizations, except for the sign which changes if orientation is reversed.

4.10 Geodesics

Definition 4.10. A *geodesic* on a surface is a regular parametrized curve on the surface whose geodesic curvature is identically zero.

Thus by (12) together with Theorem 4.9, a geodesic on a surface is a curve which in each point is as straight as possible, in the sense that it has the *least possible curvature* of a curve on the surface in that point and with that direction. For this reason (among others) we regard geodesics on a surface as the analogues of straight lines on a plane.

The property of being a geodesic is unchanged under reparametrizations of γ as well as σ, also those which revert direction or orientation (since $\kappa_g = 0$ if and only if $-\kappa_g = 0$).

Example 4.10.1 It follows from Examples 4.8.1 and Corollary 4.2 that the geodesics on the xy-plane are the straight lines contained in the plane.

Example 4.10.2 It follows from Example 4.8.2 that great circles on the unit sphere $\mathcal{S} = S^2$ are geodesics, and that small circles are not. In fact, the great circles are the only geodesics on the sphere (up to reparametrization and restriction to subsets). This can be verified as follows. Assume $\gamma(s)$ is a unit speed geodesic on $\mathcal{S}$. From Example 2.8.1 we have $\mathbf{m}(s) = -\gamma(s)$. Since $\kappa_g = 0$ for a geodesic and $\kappa_n = 1$ for all curves on a sphere (see Example 4.9.1) we conclude from (11) that $\gamma''(s) = -\gamma(s)$ for all s. Hence $\gamma''' = -\gamma'$ and it follows that $\det[\gamma' \, \gamma'' \, \gamma'''] = 0$. Hence the torsion τ is zero, and by

Corollary 4.6 the curve is contained in a fixed plane. Being in the intersection of a plane and the sphere, the curve is contained in a great circle or a small circle. However, the latter possibility was already excluded.

Theorem 4.10. *Let $\gamma = \sigma \circ \mu$ be a regular parametrized curve on σ. Then γ is a geodesic and has constant speed if and only if $\gamma''(t)$ is normal to $T_{\mu(t)}\sigma$ for all t.*

Proof. If γ has constant speed, a unit speed reparametrization is obtained by multiplying t with a constant. The second derivative of γ is proportional to the second derivative of the reparametrized curve. Hence if the curve is a geodesic, it follows from (11) that $\gamma''(t)$ is normal to the surface.

Conversely, if $\gamma''(t)$ is normal to $T_{\mu(t)}\sigma$ for all t, then $\gamma''(t) \cdot \gamma'(t) = 0$, and hence $\frac{d}{dt}\|\gamma'(t)\|^2 = 0$, from which we conclude there is constant speed. After reparametrization to unit speed we conclude from (11) that $\kappa_g = 0$. □

According to this theorem, a particle which moves on the surface with no acceleration in the tangent directions of the surface, follows a geodesic. The only acceleration is that which is necessary to keep the particle on the surface, and it is normal to the surface.

4.11 Exercises

1 Determine the curvature of the following curves in $\mathbb{R}^2$:
a. $\gamma(t) = (2t, t^2)$, b. $\gamma(t) = (e^t \cos t, e^t \sin t)$, (see page 55).

2 Let $\gamma(s)$ be a unit speed curve in $\mathbb{R}^2$, about which it is assumed that the curvature κ is a non-zero constant. Prove that the curve β defined by

$$\beta(s) = \gamma(s) + \frac{1}{\kappa}\widehat{\gamma'(s)}$$

is a constant curve, that is, a point p. Conclude that the trace of γ is contained in a circle centered in p.

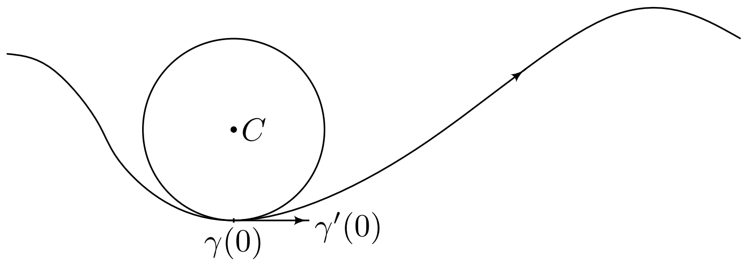
3 Let $\gamma(s)$ be a unit speed curve in $\mathbb{R}^2$, and assume that the curvature κ is non-zero at $s = 0$. Let $k = \kappa(0)$ and put

$$C = \gamma(0) + \frac{1}{k}\widehat{\gamma'(0)}.$$

Prove that the circle parametrized by

$$\beta(s) = C + \frac{1}{k}(-\cos(ks)\widehat{\gamma'(0)} + \sin(ks)\gamma'(0)),$$

satisfies $\beta(0) = \gamma(0), \beta'(0) = \gamma'(0), \beta''(0) = \gamma''(0)$. Its trace, which has radius $1/|k|$, is called the *osculating circle*. The center C is called the *center of curvature* of γ at $t = 0$. See the following figure.



4 Let $\gamma\colon I \to \mathbb{R}^2$ be a regular parametrized curve, and assume that $\|\gamma(t)\|$ has a local maximum in a given value $t_0 \in I$. Prove that $|\kappa(t_0)| \geq 1/\|\gamma(t_0)\|$. Hint: Assume unit speed. The condition on γ implies that the second derivative of $t \mapsto \|\gamma(t)\|^2$ is ≤ 0 at t_0. Conclude that $\gamma(t_0) \cdot \gamma''(t_0) \leq -1$ and employ the Cauchy-Schwarz inequality (see Appendix A)

5 Let $\gamma(s) = (\sinh^{-1}(s), \sqrt{1+s^2})$. Determine $\gamma'(s)$ and show that the curve has unit speed. Determine $\gamma''(s)$ and the curvature $\kappa(s)$. Determine a tangent angle $\theta(s)$, and verify Theorem 4.3 for this curve.
The following formula for the inverse function $\sinh^{-1}\colon \mathbb{R} \to \mathbb{R}$ can be used

$$\frac{d}{dy}\sinh^{-1} y = \frac{1}{\sqrt{1+y^2}}.$$

6 Let $\alpha\colon I \to \mathbb{R}^2$ and $\beta\colon I \to \mathbb{R}^2$ be two unit speed curves with a common interval of definition I, and with smooth tangent angles $\theta\colon I \to \mathbb{R}$ and $\varphi\colon I \to \mathbb{R}$. Assume that they have equal curvature $\kappa(s)$ for all $s \in I$, and that there exists some value $s_0 \in I$ for which $\alpha(s_0) = \beta(s_0)$ and $\alpha'(s_0) = \beta'(s_0)$. Prove that then $\alpha(s) = \beta(s)$ for all $s \in I$.

7 Determine the arc length $s(t)$, the curvature $\kappa(t)$ and the torsion $\tau(t)$ for the curve $\gamma(t) = (3t, 3t^2, 2t^3)$.

8 The curve $\gamma(t) = (t, \cosh t, \sinh t)$ is called a *hyperbolic helix*. Determine its curvature and torsion.

9 Let $\gamma(s) = (3\sin\frac{s}{5}, 4\sin\frac{s}{5}, 5\cos\frac{s}{5})$. Find $\mathbf{t}$, $\mathbf{n}$ and $\mathbf{b}$ for this curve. Find also the curvature and torsion, and show that the curve is contained in a fixed plane. Give a normal vector for this plane.

10 Let $\gamma\colon I \to \mathbb{R}^3$ be a regular parametrized curve on a regular parametrized surface $\sigma\colon U \to \mathbb{R}^3$. Assume the image of γ is a (segment of) a straight line. Prove that γ is a geodesic on σ.

11 Let σ denote the cylinder $\sigma(u, v) = (\cos v, \sin v, u)$ where $(u, v) \in \mathbb{R}^2$.

a. Let $\gamma(t) = \sigma(a\cos t, t)$ for $t \in \mathbb{R}$, where $a \in \mathbb{R}$ is constant. Determine κ_n and κ_g for γ. For which value of a is this a geodesic?

b. Instead, let $\gamma(t) = \sigma(at + b, \omega t)$ for $t \in \mathbb{R}$, where a, b and ω are constants. Describe the curve and show it is a geodesic on σ.

c. Determine two geodesic curves on σ which both have end points $(1,0,0)$ and $(1,0,1)$, but which have different trace in between these two points. Are there other geodesics between the same two points?

12 Let $\sigma(u,v) = (f(u)\cos v, f(u)\sin v, g(u))$ be a surface of revolution (see pages 35 and 57).

a. Show that the meridians $t \mapsto \sigma(t,v)$ are geodesics.

b. Verify that for the parallel curve $t \mapsto \sigma(u,t)$

$$\kappa_g(t) = \frac{f'(u)}{f(u)(f'(u)^2 + g'(u)^2)^{1/2}}, \quad \kappa_n(t) = \frac{g'(u)}{f(u)(f'(u)^2 + g'(u)^2)^{1/2}}.$$

Give a necessary and sufficient condition for it to be a geodesic.

13 Let $\gamma = \sigma \circ \mu \colon I \to \mathbb{R}^3$ be a regular parametrized curve on a regular parametrized surface σ. Assume that there exists a fixed plane Π in $\mathbb{R}^3$ containing the image of γ. If for some $t_0 \in I$ the plane Π is orthogonal to the tangent plane $T_{\mu(t_0)}\sigma$ of σ at $\mu(t_0)$, we call γ a *normal section* of σ at this point (two planes in $\mathbb{R}^3$ are orthogonal if their normal vectors are orthogonal). For example, a great circle on a sphere is a normal section at each of its points, because it belongs to a plane that intersects orthogonally with the tangent spaces.

a. Show that a normal section at t_0 has $\kappa_g(t_0) = 0$.

b. Use part a to verify Exercise 12a, that the meridians of a surface of revolution are geodesics. Verify also the geodesics found in Exercise 12b.

14 Let $\gamma \colon I \to \mathbb{R}^3$ be a unit speed curve with curvature $\kappa(t) \neq 0$ for all t. Let $\mathbf{b}(t)$ be the binormal of the curve at t. Put $\sigma(u,v) = \gamma(v) + u\mathbf{b}(v)$ for $(u,v) \in \mathbb{R} \times I$.

a. Show that σ is a regular parametrized surface. It is called the *binormal surface* of the curve.

b. Show that γ is a geodesic on the binormal surface.

15 Consider a cone $\sigma(u,v) = (u\cos v, u\sin v, au)$, with $u > 0$ and with $a > 0$ a fixed number. A sphere of radius 1 is inserted in the cone such that it exactly touches (like a scoop of ice cream in a cone).

a. Determine the center of the sphere, and parametrize the intersection of the surfaces as a smooth curve.

b. Give an argument, showing that this curve has the same geodesic curvature κ_g and the same normal curvature κ_n with respect to the two surfaces (the sphere is assumed oriented with its normal pointing towards the center). Determine $|\kappa_g|$ and κ_n.

Chapter 5

The second fundamental form

We shall now extend the notion of curvature from curves to surfaces. However the description is considerably more complicated, and the curvature of a surface in a given point p will not be completely described by a single number. The description of curvature will be based on the concept of normal curvature (see Definition 4.9), which associates infinitely many numbers to each point p, namely one for each unit tangent vector at p, describing the curvature of the surface in that direction. One of the central goals of this chapter will be to organize these numbers in an efficient way.

5.1 The shape operator

In order to treat the notion of curvature efficiently, we will use some concepts from linear algebra. The main object that describes the curvature at p will be a linear map W from the tangent space at p to itself. The map W is called the *shape operator*, or the *Weingarten map*. It will be explained in Section 5.2 how W relates to the normal curvature of Definition 4.9.

For a plane unit speed curve the description of its curvature is embodied in the formula $\mathbf{t}' = \kappa\hat{\mathbf{t}}$ (see Theorem 4.2), which expresses that the curvature is given by the rate of change of the direction $\mathbf{t}$ of the tangent line. For surfaces we will take a similar view, and our definition of W at p will reflect the rate of change of the tangent space at p.

Let $\sigma: U \to \mathbb{R}^3$ be a regular parametrized surface, and let $p = (u_0, v_0) \in U$ be given. The position of the tangent space $T_p\sigma$ in $\mathbb{R}^3$ is completely determined by the unit normal vector

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

at p. We will regard the derivative of $\mathbf{N}$ as a measure for the curvature of the surface.

In fact, it will be more convenient to use the *negative* of the derivative of $\mathbf{N}$. That this is actually in accordance with the description of the curvature of a plane curve will be explained below in Example 5.1.3.

However, since $\mathbf{N}$ is a function of the two variables u and v, both partial derivatives $-\mathbf{N}'_u$ and $-\mathbf{N}'_v$ have to be taken into account. This could be done by considering the Jacobian matrix for $-\mathbf{N}: U \to \mathbb{R}^3$, whose columns are exactly the two vectors $-\mathbf{N}'_u$ and $-\mathbf{N}'_v$, but for reasons to be explained, we prefer to proceed somewhat differently.

Observe that it follows from Lemma 4.2 that $-\mathbf{N}'_u$ and $-\mathbf{N}'_v$ are perpendicular to $\mathbf{N}$, hence they both belong to the tangent space $T_p\sigma$.

Definition 5.1. Let $p = (u_0, v_0) \in U$. The linear map

$$W = W_p\colon T_p\sigma \to T_p\sigma,$$

defined by $W(\sigma'_u) = -\mathbf{N}'_u$ and $W(\sigma'_v) = -\mathbf{N}'_v$, and hence

$$W(a\sigma'_u + b\sigma'_v) = -a\mathbf{N}'_u - b\mathbf{N}'_v \tag{1}$$

for all $a, b \in \mathbb{R}$, is called the *shape operator* of the surface at p (the derivatives σ'_u, σ'_v, $\mathbf{N}'_u$ and $\mathbf{N}'_v$ are all evaluated in p).

It follows from the fact that the pair (σ'_u, σ'_v) is a basis for $T_p\sigma$, that W is a well defined linear map $T_p\sigma \to T_p\sigma$. The motivation for studying this map rather than just the vectors $\mathbf{N}'_u, \mathbf{N}'_v$ is to obtain a notion which behaves nicely under reparametrizations. The idea is that a reparametrization will change $\mathbf{N}'_u$ and $\mathbf{N}'_v$, but also σ'_u and σ'_v, and it turns out that these changes are always so related that the map remains essentially the same. This will be seen in the theorem below, and the conclusion is that the shape operator is more directly related to a geometric property of the surface than the vectors $\mathbf{N}'_u$ and $\mathbf{N}'_v$.

Example 5.1.1 Let $\sigma(u, v) = p + uq_1 + vq_2$ be the plane through p spanned by two linearly independent vectors $q_1, q_2 \in \mathbb{R}^3$. Then $\mathbf{N} = \frac{q_1 \times q_2}{\|q_1 \times q_2\|}$ is constant, and the shape operator W is the zero operator.

Example 5.1.2 For the unit sphere with standard spherical coordinates $\sigma(u, v)$ we have seen in Example 2.8.1 that $\mathbf{N}(u, v) = -\sigma(u, v)$. Hence $\mathbf{N}'_u = -\sigma'_u$ and $\mathbf{N}'_v = -\sigma'_v$, and it follows that the shape operator W_p is the identity operator on $T_p\sigma$ for all p.

Example 5.1.3 Let $\gamma\colon I \to \mathbb{R}^2$ be a plane curve with unit speed and tangent vector $\mathbf{t}(s) = \gamma'(s)$. Since $(\hat{\mathbf{t}})' = \widehat{\mathbf{t}'}$ it follows from Theorem 4.2 that

$$(\hat{\mathbf{t}})' = \widehat{\mathbf{t}'} = \widehat{\kappa\hat{\mathbf{t}}} = \kappa\hat{\hat{\mathbf{t}}} = -\kappa\mathbf{t},$$

where we used that $\hat{\hat{\mathbf{t}}} = -\mathbf{t}$. Thus it is the *negative* of the derivative of the normal vector $\hat{\mathbf{t}}$ which describes the curvature κ.

If we view $\mathbf{N}$ and $\hat{\mathbf{t}}$ as analogues of each other, the derivatives of $\mathbf{N}$ are analogous to $(\hat{\mathbf{t}})'$. The analogue of the map W defined by (1) is therefore the linear map $a\gamma' \mapsto -a(\hat{\mathbf{t}})'$ (where $a \in \mathbb{R}$ is arbitrary) of the 1-dimensional tangent space to itself. By the equation found above this map is multiplication by κ. In this sense W is a higher dimensional version of κ.

Recall from Section 2.4 that a parametrized curve γ on σ by definition is a curve of the form $\gamma = \sigma \circ \mu$ where $\mu\colon I \to U$ is a parametrized plane curve.

Lemma 5.1. *Let $\gamma = \sigma \circ \mu$ be a parametrized curve on σ with $\mu(t_0) = p$. Then*

$$W(\gamma'(t_0)) = -\mathbf{m}'(t_0) \tag{2}$$

where $\mathbf{m} = \mathbf{N} \circ \mu$ is the surface normal along the curve.

Proof. Immediate from (1) and the two expressions in Lemma 4.9. $\square$

In other words, the shape operator associates to a tangent vector w the derivative of $-\mathbf{N}$ along any curve on the surface which has w as its tangent.

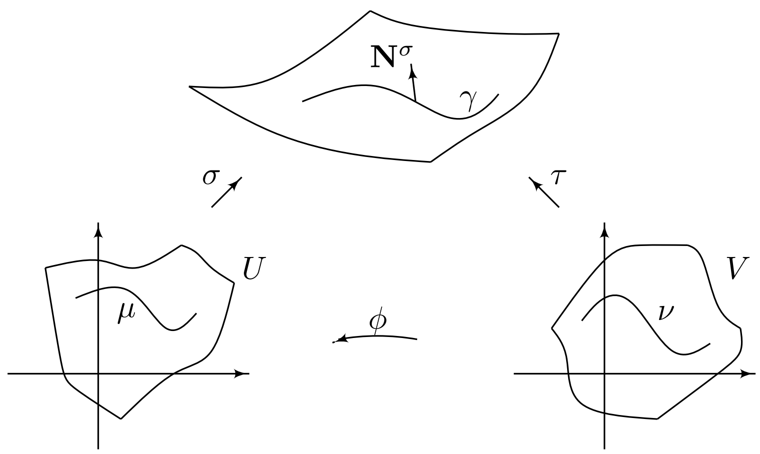
Theorem 5.1. *The shape operator W is unchanged under reparametrizations which preserve orientation, and it changes to $-W$ under reparametrizations which reverse orientation.*

Proof. Let $\tau = \sigma \circ \phi \colon V \to \mathbb{R}^3$ denote a reparametrization with diffeomorphism $\phi \colon V \to U$. We denote the shape operator of σ by W^σ, and the corresponding map for τ by W^τ. The claim is that $W^\tau = \pm W^\sigma$.

According to Theorem 2.4 each tangent vector $w \in T_p\sigma$ is of the form $w = \gamma'(t_0)$ for some parametrized curve $\gamma = \sigma \circ \mu$ on σ. We can then use the formula (2) to determine W^σ:

$$W^\sigma(\gamma'(t_0)) = -(\mathbf{N}^\sigma \circ \mu)'(t_0).$$

The curve $\gamma = \sigma \circ \mu$ can also be written as $\gamma = \tau \circ \nu$, where $\nu = \phi^{-1} \circ \mu \colon I \to V$ (see the figure below). Hence γ may be considered as a parametrized curve on τ as well.



Hence we can use (2) to determine also W^τ:

$$W^\tau(\gamma'(t_0)) = -(\mathbf{N}^\tau \circ \nu)'(t_0).$$

The unit normals for σ and τ are related by $\mathbf{N}^\tau = \pm \mathbf{N}^\sigma \circ \phi \colon V \to \mathbb{R}^3$ with sign $+$ if and only if ϕ preserves orientation (see Section 2.8). Hence

$$-(\mathbf{N}^\tau \circ \nu)'(t_0) = -(\pm(\mathbf{N}^\sigma \circ \phi) \circ (\phi^{-1} \circ \mu))'(t_0) = \pm(-\mathbf{N}^\sigma \circ \mu)'(t_0).$$

The theorem follows immediately. $\square$

5.2 The second fundamental form

We shall now introduce another fundamental object through which we wish to describe the curvature of a surface in a given point. It is closely related to the shape operator W, and it serves to relate this map to the normal curvature which was defined in Section 4.9. Let $\sigma: U \to \mathbb{R}^3$ be a regular parametrized surface, and let $p \in U$ be given.

Definition 5.2. The map $w \in T_p\sigma \mapsto I\!I_p(w) = w \cdot W(w) \in \mathbb{R}$ is called the *second fundamental form* of σ in p.

It follows from Theorem 5.1 that the second fundamental form does not change under reparametrizations, except by a sign if the orientation is reversed.

Let a tangent vector $w_0 \in T_p\sigma$ be given. Recall from Definition 4.9 that the normal curvature κ_n of σ in p with direction w_0 is the normal curvature of any curve on σ through p with that tangent vector.

Theorem 5.2. *The normal curvature in direction w_0, is*

$$\kappa_n = \frac{I\!I_p(w_0)}{\|w_0\|^2} \tag{3}$$

Proof. Let $\gamma = \sigma \circ \mu$ be a curve on σ with $\mu(t_0) = p$ and $\gamma'(t_0) = w_0$. It follows from Section 4.9, equation (15), that $\kappa_n = -\gamma'(t_0) \cdot \mathbf{m}'(t_0)/\|\gamma'(t_0)\|^2$. Hence $\kappa_n = w_0 \cdot W_p(w_0)/\|w_0\|^2$ follows from (2). $\square$

Thus, if we assume $\|w_0\| = 1$, then $I\!I_p(w_0)$ is the normal curvature at p of any curve on σ, which has the tangent vector w_0 in this point. The relation (3) describes the geometric content of the second fundamental form.

5.3 Coordinate expressions for the second fundamental form.

In the following theorem we give an explicit expression by which the second fundamental form can be computed for a given parametrization.

Theorem 5.3. *The second fundamental form is given by*

$$I\!I_p(a\sigma'_u + b\sigma'_v) = La^2 + 2Mab + Nb^2, \qquad a, b \in \mathbb{R}, \tag{4}$$

with respect to the basis σ'_u, σ'_v. Here

$$\begin{aligned} L &= \mathbf{N} \cdot \sigma''_{uu} = \frac{\det[\sigma'_u\, \sigma'_v\, \sigma''_{uu}]}{\|\sigma'_u \times \sigma'_v\|} \\ M &= \mathbf{N} \cdot \sigma''_{uv} = \frac{\det[\sigma'_u\, \sigma'_v\, \sigma''_{uv}]}{\|\sigma'_u \times \sigma'_v\|} \\ N &= \mathbf{N} \cdot \sigma''_{vv} = \frac{\det[\sigma'_u\, \sigma'_v\, \sigma''_{vv}]}{\|\sigma'_u \times \sigma'_v\|}, \end{aligned} \tag{5}$$

where all the terms are evaluated in the given point $p \in U$.

Recall the analogous expression $I_p(a\sigma'_u + b\sigma'_v) = Ea^2 + 2Fab + Gb^2$ for the first fundamental form. The two fundamental forms are quadratic forms on the tangent space (see page 44)

Proof. We shall derive the following expression for all $a, b, \tilde{a}, \tilde{b} \in \mathbb{R}$,

$$(a\sigma'_u + b\sigma'_v) \cdot W(\tilde{a}\sigma'_u + \tilde{b}\sigma'_v) = La\tilde{a} + M(a\tilde{b} + b\tilde{a}) + Nb\tilde{b}. \tag{6}$$

Taking $a = \tilde{a}$, $b = \tilde{b}$ we then obtain (4).

By linearity of W the left side of (6) equals

$$a\tilde{a}\,\sigma'_u \cdot W(\sigma'_u) + a\tilde{b}\,\sigma'_u \cdot W(\sigma'_v) + b\tilde{a}\,\sigma'_v \cdot W(\sigma'_u) + b\tilde{b}\,\sigma'_v \cdot W(\sigma'_v).$$

The expression (6) follows if we prove

$$\begin{aligned} \sigma'_u \cdot W(\sigma'_u) &= L, \qquad & \sigma'_u \cdot W(\sigma'_v) &= M, \\ \sigma'_v \cdot W(\sigma'_u) &= M, \qquad & \sigma'_v \cdot W(\sigma'_v) &= N, \end{aligned} \tag{7}$$

with L, M and N defined by (5). By definition of W (see (1)),

$$\begin{aligned} \sigma'_u \cdot W(\sigma'_u) &= -\sigma'_u \cdot \mathbf{N}'_u, \qquad & \sigma'_u \cdot W(\sigma'_v) &= -\sigma'_u \cdot \mathbf{N}'_v, \\ \sigma'_v \cdot W(\sigma'_u) &= -\sigma'_v \cdot \mathbf{N}'_u, \qquad & \sigma'_v \cdot W(\sigma'_v) &= -\sigma'_v \cdot \mathbf{N}'_v. \end{aligned} \tag{8}$$

From $\sigma'_u \cdot \mathbf{N} = 0$ we derive by differentiation with respect to u and v that

$$\sigma''_{uu} \cdot \mathbf{N} + \sigma'_u \cdot \mathbf{N}'_u = 0 \quad \text{and} \quad \sigma''_{uv} \cdot \mathbf{N} + \sigma'_u \cdot \mathbf{N}'_v = 0, \tag{9}$$

and from $\sigma'_v \cdot \mathbf{N} = 0$ we derive similarly that

$$\sigma''_{vu} \cdot \mathbf{N} + \sigma'_v \cdot \mathbf{N}'_u = 0 \quad \text{and} \quad \sigma''_{vv} \cdot \mathbf{N} + \sigma'_v \cdot \mathbf{N}'_v = 0. \tag{10}$$

The expressions in (7) then follow from (8) and (5) (since $\sigma''_{uv} = \sigma''_{vu}$). $\square$

The coefficients L, M and N are conveniently arranged as the entries of a symmetric matrix

$$\begin{pmatrix} L & M \\ M & N \end{pmatrix},$$

so that the formula for the second fundamental form can be put in matrix form

$$I\!I_p(a\sigma'_u + b\sigma'_v) = \begin{pmatrix} a \\ b \end{pmatrix}^t \begin{pmatrix} L & M \\ M & N \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix}.$$

Example 5.3 Let $\sigma(u, v) = (r \cos u \cos v, r \cos u \sin v, r \sin u)$ be the standard parametrization of a sphere with radius $r > 0$. A straightforward computation shows that the first fundamental form has coefficients $E = r^2$, $F = 0$ and $G = r^2 \cos^2 u$. Moreover, $\mathbf{N} = -(\cos u \cos v, \cos u \sin v, \sin u)$ and

$$L = \mathbf{N} \cdot \sigma''_{uu} = r, \quad M = \mathbf{N} \cdot \sigma''_{uv} = 0, \quad N = \mathbf{N} \cdot \sigma''_{vv} = r \cos^2 u.$$

Hence the second fundamental form at $p = (u, v)$ is the map

$$a\sigma'_u + b\sigma'_v \mapsto r(a^2 + b^2 \cos^2 u).$$

5.4 A formula for the shape operator

With the aid of the coefficients L, M and N we can establish a formula by which the shape operator can be computed in a given parametrization.

Theorem 5.4. *The matrix for the shape operator $W_p\colon T_p\sigma \to T_p\sigma$ with respect to the basis (σ'_u, σ'_v) is*

$$\begin{pmatrix} E & F \\ F & G \end{pmatrix}^{-1} \begin{pmatrix} L & M \\ M & N \end{pmatrix}.$$

Recall that for a 2×2 matrix,

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}^{-1} = \frac{1}{ad-bc} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}.$$

Proof. We express $W(\sigma'_u)$ and $W(\sigma'_v)$ as linear combinations

$$W(\sigma'_u) = h\sigma'_u + j\sigma'_v, \qquad W(\sigma'_v) = i\sigma'_u + k\sigma'_v$$

with coefficients i, j, k, l to be determined. The matrix for W_p will then be

$$\begin{pmatrix} h & i \\ j & k \end{pmatrix}.$$

From (7) we obtain

$$\begin{aligned} \begin{pmatrix} L & M \\ M & N \end{pmatrix} &= \begin{pmatrix} \sigma'_u \cdot (h\sigma'_u + j\sigma'_v) & \sigma'_u \cdot (i\sigma'_u + k\sigma'_v) \\ \sigma'_v \cdot (h\sigma'_u + j\sigma'_v) & \sigma'_v \cdot (i\sigma'_u + k\sigma'_v) \end{pmatrix} \\ &= \begin{pmatrix} Eh + Fj & Ei + Fk \\ Fh + Gj & Fi + Gk \end{pmatrix} \\ &= \begin{pmatrix} E & F \\ F & G \end{pmatrix} \begin{pmatrix} h & i \\ j & k \end{pmatrix}. \end{aligned}$$

Hence

$$\begin{pmatrix} h & i \\ j & k \end{pmatrix} = \begin{pmatrix} E & F \\ F & G \end{pmatrix}^{-1} \begin{pmatrix} L & M \\ M & N \end{pmatrix}. \quad \square$$

Notice that it does *not* follow that W_p is represented by a symmetric matrix (the product of two symmetric matrices need not be symmetric). In fact, this will not be the case in general.

Example 5.4 For the sphere with radius r we obtain from Example 5.3 that the matrix for W with respect to σ'_u and σ'_v is

$$\begin{pmatrix} r^2 & 0 \\ 0 & r^2\cos^2 u \end{pmatrix}^{-1} \begin{pmatrix} r & 0 \\ 0 & r\cos^2 u \end{pmatrix} = \begin{pmatrix} 1/r & 0 \\ 0 & 1/r \end{pmatrix}.$$

5.5 Diagonalization of the second fundamental form

We shall now introduce a general method by which one can deduce from the operator W (and its matrix expression) some detailed information about the shape of a surface. The information is obtained through diagonalization of W (see Appendix D).

Definition 5.5. An eigenvector for the shape operator W_p is called a *principal vector* in $T_p\sigma$, and the corresponding eigenvalue is called the corresponding *principal curvature* at p.

Notice that if $w \in T_p\sigma$ is a principal vector with unit length and corresponding principal curvature λ, then by Theorem 5.2 the normal curvature at p in direction w is

$$\kappa_n = I\!I_p(w) = w \cdot W_p(w) = w \cdot \lambda w = \lambda.$$

This explains why λ is called a 'curvature'.

It follows from Theorem 5.1 that the notion of a principal vector is unchanged under a reparametrization, and that the corresponding principal curvatures are unchanged in absolute value, but with the opposite sign if the orientation is reversed.

We see from (6) that the shape operator $W\colon T_p\sigma \to T_p\sigma$ is *symmetric* with respect to the dot product, that is

$$w_1 \cdot W(w_2) = W(w_1) \cdot w_2 \tag{11}$$

for all $w_1, w_2 \in T_p\sigma$.

Theorem 5.5. *There exists for each $p \in U$ an orthonormal basis w_1, w_2 for $T_p\sigma$ consisting of principal vectors with corresponding principal curvatures $\kappa_1, \kappa_2 \in \mathbb{R}$.*

With respect to this basis the second fundamental form is given by

$$I\!I_p(aw_1 + bw_2) = \kappa_1 a^2 + \kappa_2 b^2 \tag{12}$$

for all $a, b \in \mathbb{R}$.

Proof. The first statement follows immediately from Corollary D.1 in Appendix D with $U = T_p\sigma \subset \mathbb{R}^3$.

The expression (12) now follows by evaluation of $w \cdot W(w)$ with $w = aw_1 + bw_2$. $\square$

Corollary 5.5.1. *Let w_1, w_2 and κ_1, κ_2 be as above, and let $\theta \in \mathbb{R}$. The normal curvature in direction*

$$w_0 = \cos\theta w_1 + \sin\theta w_2$$

is

$$\kappa_n = \kappa_1 \cos^2\theta + \kappa_2 \sin^2\theta. \tag{13}$$

In particular, κ_n belongs to the interval between κ_1 and κ_2, which are the extremal values of κ_n.

Proof. It follows from Theorem 5.2 that $\kappa_n = I\!I_p(w_0)$. Then (13) is obtained from (12). Furthermore, if for example $\kappa_1 \le \kappa_2$, then from (13)

$$\kappa_n = \kappa_1 \cos^2\theta + \kappa_2 \sin^2\theta \le \kappa_2 \cos^2\theta + \kappa_2 \sin^2\theta = \kappa_2$$

and similarly $\kappa_n \ge \kappa_1$. □

The principal curvatures and directions can be explicitly determined by means of the matrix for W. We summarize the conclusion:

Corollary 5.5.2. *The principal curvatures κ_1, κ_2 are the roots κ of the equation*

$$\det\left(\begin{pmatrix} E & F \\ F & G \end{pmatrix}^{-1}\begin{pmatrix} L & M \\ M & N \end{pmatrix} - \kappa\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\right) = 0.$$

The corresponding principal vectors are $a\sigma'_u + b\sigma'_v$ where $\binom{a}{b}$ is non-zero and solves

$$\begin{pmatrix} E & F \\ F & G \end{pmatrix}^{-1}\begin{pmatrix} L & M \\ M & N \end{pmatrix}\begin{pmatrix} a \\ b \end{pmatrix} = \kappa_i\begin{pmatrix} a \\ b \end{pmatrix}.$$

Proof. This follows from the fact that the shape operator is represented by the matrix $\begin{pmatrix} E & F \\ F & G \end{pmatrix}^{-1}\begin{pmatrix} L & M \\ M & N \end{pmatrix}$ according to Theorem 5.4. □

Example 5.5.1 Let $\sigma(u,v) = (\cos v, \sin v, u)$, then σ parametrizes a cylinder (Example 1.2.3). We will determine the principal curvatures and principal vectors at the point $\sigma(u,v)$. We find

$$\sigma'_u = (0,0,1), \quad \sigma'_v = (-\sin v, \cos v, 0)$$

and hence $E = G = 1$, $F = 0$, and $\mathbf{N} = (-\cos v, -\sin v, 0)$. Furthermore

$$\sigma''_{uu} = \sigma''_{uv} = 0, \quad \sigma''_{vv} = (-\cos v, -\sin v, 0)$$

and hence

$$L = M = 0, \quad N = 1.$$

The matrix of the shape operator with respect to σ'_u, σ'_v is therefore

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}^{-1}\begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}.$$

The principal curvatures are the eigenvalues of this matrix, $\kappa_1 = 0$ and $\kappa_2 = 1$. Principal vectors are σ'_u and σ'_v since the matrix is already diagonal. The normal curvature in direction σ'_u (vertical) is zero, and the normal curvature in direction σ'_v (horizontal) is 1.

5.6 The graph of a quadratic form

In order to illustrate the theory of the previous section, we will study the surface formed by the graph of a particularly simple function.

A *quadratic form* on $\mathbb{R}^2$ is a function $q\colon \mathbb{R}^2 \to \mathbb{R}$ of the form

$$q(x,y) = ax^2 + 2bxy + cy^2 \tag{14}$$

for some constants $a, b, c \in \mathbb{R}$. It is convenient to write the formula for q in matrix form

$$q(x,y) = \begin{pmatrix} x \\ y \end{pmatrix}^t \begin{pmatrix} a & b \\ b & c \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}, \tag{15}$$

where t denotes transposition. As in the previous section, the key to the analysis is the diagonalization known from linear algebra. Recall that every symmetric matrix A is *orthogonally diagonalizable*, that is, there exists an orthogonal 2×2 matrix C such that

$$D = C^{-1}AC$$

is a diagonal matrix with real entries (see Appendix D).

We apply the diagonalization to the matrix A of our quadratic form (15). As explained in Appendix D, the columns of C are chosen as an orthonormal basis of eigenvectors for A. Let $w = \left(\begin{smallmatrix} x \\ y \end{smallmatrix}\right) \in \mathbb{R}^2$ be given. The coordinates of w, with respect to the basis given by the columns of C, are denoted $\left(\begin{smallmatrix} x' \\ y' \end{smallmatrix}\right)$. Then

$$w = C \begin{pmatrix} x' \\ y' \end{pmatrix}.$$

Write $w' = \left(\begin{smallmatrix} x' \\ y' \end{smallmatrix}\right)$, then $w = Cw'$ and we obtain from (15) that

$$q(w) = w^t A w = (Cw')^t A (Cw') = {w'}^t C^t A C w' = {w'}^t D w'$$

since $C^t = C^{-1}$ and $C^{-1}AC = D$. Let λ_1, λ_2 be the eigenvalues in the diagonal of D. It follows from the preceding calculation that

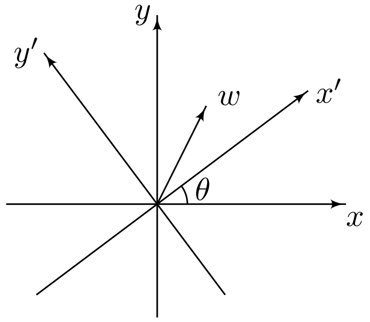
$$q(w) = \lambda_1 {x'}^2 + \lambda_2 {y'}^2.$$

Thus the change of variables from (x, y) to (x', y') results in a simplification of the expression for q, where the product term xy disappears.

Notice that $\det C = \pm 1$, and by changing the sign on one of the columns, if necessary, we can arrange that $\det C = 1$ (the columns will still be an orthonormal set of eigenvectors). Then C has the form

$$C = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$$

for some $\theta \in \mathbb{R}$, and it corresponds to a counterclockwise rotation by the angle θ. The basis vectors in the columns of C are obtained from the standard basis vectors e_i exactly by this rotation, and the new coordinates x' and y' are the coordinates of w with respect to the rotated basis.



$$\begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}$$

We have established the following theorem.

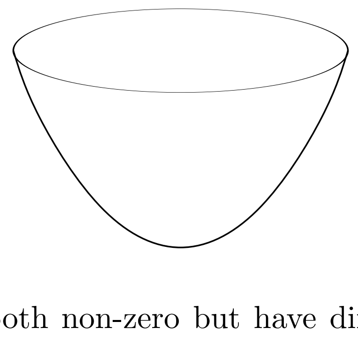
Theorem 5.6. *Let $q(w) = w^t A w$ be a quadratic form on $\mathbb{R}^2$ with symmetric 2×2 matrix A. There exists a rotation of $\mathbb{R}^2$ such that in the rotated $x'y'$-coordinates*

$$q(w) = \lambda_1 {x'}^2 + \lambda_2 {y'}^2,$$

where λ_1, λ_2 are the eigenvalues of A.

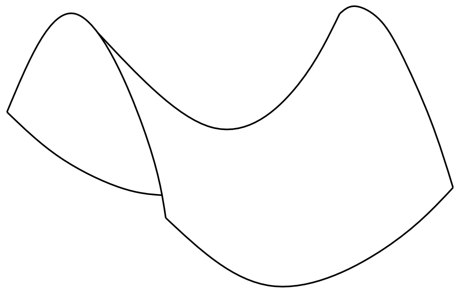
In these rotated coordinates we can easily describe the graph of q. Notice that the vertical cross section of the graph, obtained by taking the intersection with one of the two vertical coordinate planes ($x'z$-plane and $y'z$-plane respectively), is a parabola ($z = \lambda_1 {x'}^2$ and $z = \lambda_2 {y'}^2$, respectively). Therefore the surface is called a *paraboloid*.

The shape of the horizontal cross sections of the graph depend very much on the eigenvalues λ_1 and λ_2. If the eigenvalues are both positive or both negative, then each horizontal cross section of the graph is an ellipse, and the graph is called an *elliptic paraboloid*. The graph is shown below in the positive case (the negative case is similar, but upside down).



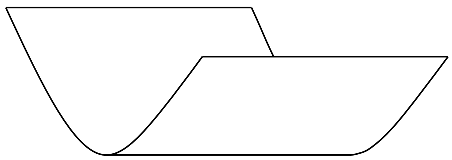
elliptic paraboloid

If λ_1 and λ_2 are both non-zero but have different signs, then the graph is called a *hyperbolic paraboloid*, because each horizontal cross section of the graph is a hyperbola. In this case the graph has the shape of a 'saddle', see below.



hyperbolic paraboloid

If one of the eigenvalues is zero, but not the other one, then the graph is called a *parabolic cylinder* (it is a 'cylinder' in which the cross section is a parabola instead of a circle). Finally, if $\lambda_1 = \lambda_2 = 0$ then q is the zero function and the graph is a *plane*.



parabolic cylinder

The relation to the theory in Section 5.5 is as follows. In the rotated coordinates we obtain a graph of the form $\sigma(u, v) = (u, v, \lambda_1 u^2 + \lambda_2 v^2)$. A simple calculation shows that at $(u, v) = (0, 0)$ we have

$$\begin{pmatrix} E & F \\ F & G \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad \begin{pmatrix} L & M \\ M & N \end{pmatrix} = \begin{pmatrix} 2\lambda_1 & 0 \\ 0 & 2\lambda_2 \end{pmatrix}.$$

We see that the rotation of coordinates exactly has the effect that the shape operator is diagonalized. The principal curvatures are $2\lambda_1$ and $2\lambda_2$, and principal vectors are along the two horizontal axes.

Example 5.6.1 To the quadratic form $q(x, y) = x^2 + xy + y^2$ corresponds the symmetric matrix

$$\begin{pmatrix} 1 & \frac{1}{2} \\ \frac{1}{2} & 1 \end{pmatrix}$$

which is diagonalized in Example D.1. The diagonalized matrix is

$$D = C^{-1}AC = \begin{pmatrix} \frac{1}{2} & 0 \\ 0 & \frac{3}{2} \end{pmatrix}$$

where

$$C = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix}.$$

The quadratic form $x^2 + xy + y^2$ thus becomes $\frac{1}{2}x'^2 + \frac{3}{2}y'^2$ in rotated coordinates. The graph is an elliptic paraboloid. Its axes are rotated from the x and y axes by the angle θ determined from $\cos\theta = \frac{1}{\sqrt{2}}$, $\sin\theta = -\frac{1}{\sqrt{2}}$, that is, clockwise by 45 degrees.

5.7 The type of a surface

The principal curvatures and vectors can be explained geometrically as follows. For simplicity we assume that the given point $\sigma(p)$ on the surface is the origin, and that the tangent plane in this point is exactly the xy-coordinate plane. This can always be arranged by a suitable translation followed by a suitable rotation of $\mathbb{R}^3$, and it can be shown that such a transformation does not alter κ_1 and κ_2. Furthermore, it follows from Theorem 2.11 (and its proof) that σ allows an orientation preserving reparametrization as a graph over the xy-plane. Observe that the principal curvatures are unchanged also by such a reparametrization. We therefore assume that σ is already of this form, that is

$$\sigma(u, v) = (u, v, h(u, v))$$

where $h(u, v)$ is smooth.

Since $\sigma(p) = (0, 0, 0)$ we have $p = (0, 0)$ and $h(0, 0) = 0$. Now

$$\sigma'_u = (1, 0, h'_u), \qquad \sigma'_v = (0, 1, h'_v)$$

and since $T_p\sigma$ is the xy-plane we conclude that $h'_u(0, 0) = h'_v(0, 0) = 0$. In particular, we see that the first fundamental form has

$$E = G = 1, \qquad F = 0$$

in p. The unit normal vector is $\mathbf{N} = (0, 0, 1)$, and since

$$\sigma''_{uu} = (0, 0, h''_{uu}), \quad \sigma''_{uv} = (0, 0, h''_{uv}), \quad \sigma''_{vv} = (0, 0, h''_{vv})$$

we obtain from Theorem 5.3 that at p

$$L = h''_{uu}(0, 0), \quad M = h''_{uv}(0, 0), \quad N = h''_{vv}(0, 0).$$

The Taylor expansion to order two of σ now reads (see Appendix B)

$$\begin{aligned}\sigma(u, v) &\simeq \sigma(0, 0) + u\sigma'_u(0, 0) + v\sigma'_v(0, 0)\\ &\qquad + \tfrac{1}{2}(u^2\sigma''_{uu}(0, 0) + 2uv\sigma''_{uv} + v^2\sigma''_{vv}(0, 0))\\ &= (u, v, \tfrac{1}{2}(u^2L + 2uvM + v^2N)) = (u, v, \tfrac{1}{2}I\!I_p(u\sigma'_u + v\sigma'_v)).\end{aligned}$$

We thus see that σ is approximated near p by the graph of $\frac{1}{2}I\!I_p$, and we can read off the shape of σ from the shape of this graph. Since $I\!I_p$ is a quadratic form, its shape was described in Section 5.6. The conclusion is that after a suitable rotation of the xy-plane, which brings the principal vectors in the direction of the axes, the surface will have an appearance like one of the figures in Section 5.6, depending on the signatures of the numbers κ_1, κ_2.

Definition 5.7. The *type* of a point $p \in U$ is defined as follows. It is called an *elliptic point* of the surface if the principal curvatures κ_1, κ_2 at p are non-zero and have the same sign, and a *hyperbolic point* if they are non-zero with opposite signs. If one of the principal curvatures is zero, but the other not, the point is called *parabolic*, and finally if $\kappa_1 = \kappa_2 = 0$ the point is called *planar*.

Notice that the type of a point does not change by reparametrization, since the principal curvatures are either unchanged or both change sign.

5.8 Exercises

1 Let σ denote the helicoid $\sigma(u,v) = (u\cos v, u\sin v, v)$.

a. Determine κ_g and κ_n for the helix $\gamma(t) = (a\cos t, a\sin t, t)$ on σ. Here $a \in \mathbb{R}$ is a constant (in the degenerate case $a = 0$, the helix is a line).

b. Determine $W(\gamma'(t))$, where W is the shape operator for σ at $p = (a,t)$.

c. Answer the same questions for the curve $\beta(t) = (t\cos b, t\sin b, b)$ on σ, with $b \in \mathbb{R}$ a constant.

d. Which of the mentioned curves are geodesics on the helicoid?

2 For the helicoid $\sigma(u,v) = (u\cos v, u\sin v, av)$, where $a \neq 0$ is a constant, the first fundamental form was determined in Exercise 10, page 56. Determine the coefficients L, M and N of the second fundamental form.

3 For a surface of revolution $\sigma(u,v) = (f(u)\cos v, f(u)\sin v, g(u))$ the first fundamental form was determined in Exercise 12, page 57. Verify the following expressions for the second fundamental form:

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

4 Let $\sigma\colon U \to \mathbb{R}^3$ be a regular parametrized surface. Show that if $\sigma(U)$ is contained in a fixed plane $\{x \in \mathbb{R}^3 \mid n \cdot x = c\}$, where $n \in \mathbb{R}^3$ is a unit vector and $c \in \mathbb{R}$, then $L = M = N = 0$.
Prove also the following converse. Assume that U is a rectangle $]a,b[\times]c,d[$ and that the second fundamental form is identically 0. Then $\sigma(U)$ is contained in a plane. (Hint: Use (9)-(10) to prove that $\mathbf{N}$ is constant. Prove next that $\mathbf{N} \cdot \sigma$ is constant).

5 Let σ be a regular parametrized surface for which the image is contained in a fixed sphere $\{x \in \mathbb{R}^3 \mid \|x-a\| = r\}$ where $a \in \mathbb{R}^3$ and $r > 0$. Show that then $\pm r\mathbf{N}(u,v) = \sigma(u,v) - a$ for all (u,v), and prove that the fundamental forms are proportional: $\mp(rL, rM, rN) = (E, F, G)$.
Prove also the following converse. Assume that U is a rectangle $]a,b[\times]c,d[$ and that there exists a constant $r \neq 0$ such that $(rL, rM, rN) = (E, F, G)$. Then $a = \sigma + r\mathbf{N}$ is constant and $\sigma(U)$ is contained in the sphere with this center and radius $|r|$.

6 Consider the parametrized surface $\sigma(u,v) = (u-v, u+v, u^2+v^2)$ for $(u,v) \in \mathbb{R}^2$.

a. Determine the coefficients E, F and G.

b. Let $p = (\frac{1}{2}, \frac{1}{2})$. Show that the vectors $e_1 = (1,0,0)$ and $e_2 + e_3 = (0,1,1)$ belong to $T_p\sigma$, and determine their coordinates with respect to $\sigma'_u(p)$, $\sigma'_v(p)$.

c. Determine L, M and N at $p = (\frac{1}{2}, \frac{1}{2})$.

d. Show that e_1 and $e_2 + e_3$ are principal vectors at $p = (\frac{1}{2}, \frac{1}{2})$, and determine the corresponding principal curvatures κ_1 and κ_2.

e. Let $\gamma(t) = (\frac{1}{\sqrt{2}}(\cos t - \sin t), \frac{1}{\sqrt{2}}(\cos t + \sin t), \frac{1}{2})$ for $t \in \mathbb{R}$. Show that γ can be realized as a curve on σ, and determine the curvatures κ_n and κ_g at $t = \frac{\pi}{4}$. One of them coincides with κ_1. Explain why.

7 Let $\sigma(u,v) = (u, v, uv)$ for $(u,v) \in \mathbb{R}^2$ and consider the point $p = (1,0)$. Compute E, F, G, L, M and N for σ at p, and determine the normal curvature of σ in the direction $w_0 = (\frac{1}{3}, \frac{2}{3}, \frac{2}{3}) \in T_p\sigma$.
Determine the principal curvatures and principal vectors for σ at p.

8 Let $\sigma(u,v)$ be a regular parametrized surface. Assume at a given point (u_0, v_0) that $F(u_0,v_0) = M(u_0,v_0) = 0$. Show that then σ'_u and σ'_v are principal vectors at this point, with corresponding principal curvatures $\kappa_1 = \frac{L}{E}$ and $\kappa_2 = \frac{N}{G}$.
In the converse direction, show also that if $\sigma'_u(u_0,v_0)$ and $\sigma'_v(u_0,v_0)$ are principal vectors with corresponding curvatures κ_1, κ_2, which *are different*, then $F = M = 0$ at this point. Give finally an example which shows that this converse conclusion cannot be reached if $\kappa_1 = \kappa_2$.

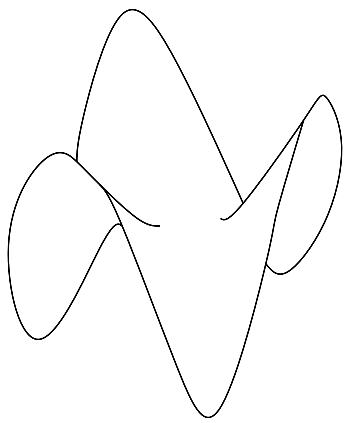
9 Let $F\colon \mathbb{R}^3 \to \mathbb{R}^3$ be a map of the form $F(x) = Ax + b$, where A is an orthogonal 3×3-matrix with $\det A = 1$, and $b \in \mathbb{R}^3$ a constant vector (such a map is called a *rigid motion*).
Prove that if $\sigma\colon U \to \mathbb{R}^3$ is a regular parametrized surface, then so is $\tau = F \circ \sigma$ (use Exercise C2 in Appendix C). Verify furthermore that the coefficients E, F, G, L, M, N are equal for σ and τ. Verify that if $w \in \mathbb{R}^3$ is a principal vector for σ, then Aw is a principal vector for τ with the same principal curvature κ.

10 Let $q(x, y) = 2x^2 + 4xy + 5y^2$. Determine a rotation of $\mathbb{R}^2$ which brings q in the form of Theorem 5.6. Of which type is the graph of q? Describe the level set $q(x, y) = 1$?
Answer the same questions for $q(x, y) = ax^2 + 24xy + (a + 7)y^2$, for all possible values of $a \in \mathbb{R}$.

11 Suppose a quadratic form $q(x, y) = ax^2 + 3xy + by^2$ can be brought to the form $4(x')^2 - (y')^2$ by a rotation of $\mathbb{R}^2$. Determine the possible values of a and b.

12 Consider the graph of $h(u, v) = uv - \cos u - \cos v$, where $u, v \in]-\pi, \pi[$. Show that each point $(u, v) \neq (0, 0)$ is hyperbolic, and that $(u, v) = (0, 0)$ is parabolic.

13 The graph of $h(u, v) = u^3 - 3uv^2$ is called the *monkey saddle* because the point $(0, 0, 0)$ is a saddle point with slopes for both two legs and a tail.



Determine E, F, G and L, M, N at $(u, v) = (0, 0)$. Determine also the principal curvatures κ_1, κ_2 in this point. Which is the type of the point $(0, 0, 0)$ on the monkey saddle?

14 Let $\sigma(u, v) = (u + v, v, \frac{1}{2}u^2 + uv + 2v^2)$. Compute E, F, G, L, M, N and the principal curvatures κ_1, κ_2 at $(u, v) = (0, 0)$. Determine also the corresponding principal vectors and the type of the point.

15 Find a function $h(u, v)$ of the form $h(u, v) = au + bv + cu^2 + duv + ev^2$, for which the graph has

$$E = 5/4,\ F = 1/2,\ G = 2,\ L = 3/4,\ M = -3/2,\ N = 3$$

at $(u, v) = (0, 0)$. Determine the principal curvatures, corresponding principal vectors, and the type of the point $(0, 0, 0)$ on the graph.

16 Let $\sigma : U \to \mathbb{R}^3$ be a regular parametrized surface. Let $\gamma = \sigma \circ \mu : I \to \mathbb{R}^3$ be a regular parametrized curve on σ, and assume the image of γ is contained in a straight line. Let κ_1 and κ_2 be the principal curvatures for σ at some point on the curve, say $\mu(t_0) \in U$ where $t_0 \in I$. Prove that $\kappa_1 \leq 0 \leq \kappa_2$ or $\kappa_2 \leq 0 \leq \kappa_1$.

Chapter 6

Teorema egregium

In the investigation of the geometry of surfaces one of the central issues is to determine which geometric quantities of the surface can be determined solely on the basis of computations involving measurements of arc lengths on the surface. Such a quantity is called *intrinsic*. The point of the notion is that an intrinsic quantity is an 'internal' property of the surface, independent of the surrounding space. For example, the distance between two opposite poles on a sphere of radius 1 is 2, but the shortest distance that can be measured on the surface is π, along a great circle. The distance measured through the surrounding space is not intrinsic.

In this chapter we will investigate some of the geometric notions we have introduced from this perspective. Most importantly, we shall prove a famous theorem of Gauss, which asserts that a particular measure for the curvature, called the Gaussian curvature, is intrinsic.

6.1 The Gaussian curvature

In the preceding chapter we have described the curvature of a surface in a given point either by means of a linear map or by means of a quadratic form, both being rather complicated objects. It would be tempting to try to reduce to a description by means of a single number. One such number is the following measure of curvature, which was introduced by Gauss.

Recall, that if $U \subset \mathbb{R}^n$ is an m-dimensional linear space and $L: U \to U$ a linear map, the *determinant* of L, denoted by $\det L$, is defined as the determinant of the $m \times m$ matrix that represents L in some basis for U. It is a theorem of linear algebra that the determinant is independent of the chosen basis (the matrix will be different in another basis, but the determinant will remain the same).

Definition 6.1. The *Gaussian curvature* (or *total curvature*) $K(p)$ of σ at p is the determinant of the map W. That is (compare Theorem 5.4)

$$K(p) = \det\left(\begin{pmatrix} E & F \\ F & G \end{pmatrix}^{-1} \begin{pmatrix} L & M \\ M & N \end{pmatrix} \right) = \frac{LN - M^2}{EG - F^2}.$$

Notice that the determinant $K(p)$ does not depend on the use of the basis (σ'_u, σ'_v) for $T_p\sigma$, which is used in the above expression. It follows that $K(p)$ is unchanged by reparametrizations, since by Theorem 5.1 the shape operator W is unchanged or changes to $-W$ (the latter change does not alter the determinant).

It will be seen in the examples below that there exist surfaces with quite different shapes, which have the same Gaussian curvature everywhere. Therefore, the Gaussian curvature does not hold complete information about the shape of the surface.

Example 6.1.1 For the plane we saw in Example 5.1.1 that W is the zero operator. Hence its Gaussian curvature is $K = 0$. For the unit sphere we determined W to be the identity operator (see Example 5.1.2), and we conclude that the Gaussian curvature is $K = 1$. More generally, it follows from Example 5.4 that the Gaussian curvature of a sphere of radius r is $K = 1/r^2$.

Example 6.1.2 Consider again the cylinder $\sigma(u, v) = (\cos v, \sin v, u)$ from Example 5.5.1. We will determine the Gauss curvature in the point $\sigma(u, v)$. We saw that $E = G = 1$, $F = 0$, and $L = M = 0$, $N = 1$. It follows that the Gaussian curvature is $K = 0$. Notice that the cylinder and the plane thus have the same Gaussian curvature, although they have different shapes.

The sign of the Gaussian curvature has a particular geometric significance, which is explained in the following result.

Theorem 6.1. *The Gauss curvature of σ at p is the product*

$$K(p) = \kappa_1 \kappa_2.$$

In particular, σ is elliptic at p if and only if $K(p) > 0$, it is hyperbolic at p if and only if $K(p) < 0$, and it is parabolic or planar at p if and only if $K(p) = 0$.

Proof. With respect to a basis of eigenvectors, the matrix of W is diagonal with κ_1, κ_2 in the diagonal. The determinant is then the product of these entries. □

We see that although the Gauss curvature $K(p)$ does not give the complete picture, it holds sufficient information to determine the type of the surface, except that it does not permit distinction between parabolic and planar points.

6.2 Intrinsic geometry

We shall now make the considerations in the introduction to this chapter more precise.

We can determine lengths of tangent vectors as follows. Let a tangent vector $w \in T_p\sigma$ be given. Choose a curve $\gamma(t)$ on σ with w as tangent vector

$\gamma'(t_0) = w$. Let $\ell(t)$ denote the arc length of γ from t_0 to t, then this function is determined by measurements of arc lengths. Since

$$\|w\| = \ell'(t_0)$$

we conclude that the length of w is intrinsic.

In particular, the coefficients $E = \|\sigma'_u\|^2$ and $G = \|\sigma'_v\|^2$ of the first fundamental form can thus be determined by measuring the arc lengths of the curves $t \mapsto \sigma(t, v)$ and $t \mapsto \sigma(u, t)$, to which σ'_u and σ'_v are the tangent vectors. By measuring arc lengths along $t \mapsto \sigma(t, t)$, whose tangent vector is $\sigma'_u + \sigma'_v$, we can determine $\|\sigma'_u + \sigma'_v\|$, and since $\|\sigma'_u + \sigma'_v\|^2 = E + G + 2F$ we can thus determine F as well. Therefore, any quantity that can be expressed in terms of E, F and G, can also be expressed in terms of lengths of curves. Conversely, the arc length of a parametrized curve on σ was expressed by means of E, F and G, in Theorem 3.4. The property of being expressible in terms of arc lengths is therefore equivalent with the property of being expressible in terms of the first fundamental form.

The following definition is a more concise version of what was explained above.

Definition 6.2. A quantity or property of a parametrized surface σ, which can be expressed purely in terms of the coefficient functions E, F and G of the first fundamental form for σ, is called *intrinsic*. If in addition it is invariant under reparametrizations of σ, it is called *intrinsic invariant*.

As discussed above, the arc length of a parametrized curve on σ is intrinsic invariant. Other examples are the angle between tangent vectors (see Section 3.4, eq. (5)) and the area of a subset (see Definition 3.9).

The coefficients E, F and G are intrinsic but not invariant, because they change when the surface is reparametrized. On the other hand, the coordinates in $\mathbb{R}^3$ of $\sigma(u, v)$ are not intrinsic since they cannot be determined from E, F and G alone. To see this, it suffices to notice that a translation of the surface will change these coordinates without changing E, F and G.

The coefficients L, M and N of the second fundamental form are not intrinsic either. For example, we have seen that the plane and the cylinder can both be parametrized such that $E = G = 1$ and $F = 0$, but the second fundamental forms do not agree.

The shape operator W and the principal curvatures κ_1 and κ_2 are invariant under reparametrization (up to $\pm$), but the same example of the plane and the cylinder shows that they are not intrinsic.

We thus see that being intrinsic invariant is a quite rare property for the quantities we have introduced to describe surfaces. This is not surprising, if we compare with the analogue for curves. The corresponding definition of 'intrinsic invariant' for a quantity related to a curve, say in $\mathbb{R}^2$, requires that the quantity can be determined only from the measurement of lengths along the curve. However, we know from Theorem 2.5 that all curves can be reparametrized to unit arc length, and hence no curves at all can be distinguished from each other by means of intrinsic invariants. Remarkably, we shall see in the following sections that the situation is less hopeless for surfaces.

6.3 Christoffel symbols

We have earlier mentioned that the coefficient functions E, F and G are the analogs for a parametrized surface of the function $t \mapsto \|\gamma'(t)\|^2$ for a parametrized curve. From the latter function one can easily determine the dot product $\gamma''(t) \cdot \gamma'(t)$, since

$$\gamma''(t) \cdot \gamma'(t) = \frac{1}{2}\frac{d}{dt}\gamma'(t) \cdot \gamma'(t) = \frac{1}{2}\frac{d}{dt}\|\gamma'(t)\|^2. \tag{1}$$

We will now derive the analog for surfaces of this observation.

In order to express coefficients in an efficient way, it is convenient to change notation and use indices. We number the coordinates u and v by 1 and 2, thus

$$\sigma'_1 = \sigma'_u, \quad \sigma'_2 = \sigma'_v$$

and

$$\sigma''_{11} = \sigma''_{uu}, \quad \sigma''_{12} = \sigma''_{uv}, \quad \text{etc.}$$

The matrices of components of the two fundamental forms are denoted

$$\begin{pmatrix} g_{11} & g_{12} \\ g_{21} & g_{22} \end{pmatrix} = \begin{pmatrix} E & F \\ F & G \end{pmatrix}, \qquad \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix} = \begin{pmatrix} L & M \\ M & N \end{pmatrix},$$

that is,

$$g_{ij} = \sigma'_i \cdot \sigma'_j, \qquad b_{ij} = \sigma''_{ij} \cdot \mathbf{N}. \tag{2}$$

The analog of (1) is

Lemma 6.3. *The expression $\sigma''_{ij} \cdot \sigma'_k$ is intrinsic. It can be determined from the coefficients of the first fundamental form by means of the following formulas*

$$\sigma''_{ij} \cdot \sigma'_k = \frac{1}{2}\left(\frac{\partial g_{ik}}{\partial u_j} + \frac{\partial g_{jk}}{\partial u_i} - \frac{\partial g_{ij}}{\partial u_k}\right), \qquad (i, j, k = 1, 2). \tag{3}$$

Proof. By differentiation of $g_{ik} = \sigma'_i \cdot \sigma'_k$ we obtain

$$\frac{\partial g_{ik}}{\partial u_j} = \sigma''_{ij} \cdot \sigma'_k + \sigma''_{kj} \cdot \sigma'_i.$$

We insert this expression in the right side of (3), with proper permutations of the symbols. The equality with the left side of (3) is obtained by simplification with the symmetry rule $\sigma''_{ij} = \sigma''_{ji}$. $\square$

In the following it will be convenient to work with some quantities which are closely related to the $\sigma''_{ij} \cdot \sigma'_k$. These are the so-called *Christoffel symbols.*

Definition 6.3. The Christoffel symbols associated with σ are the functions $\Gamma^k_{ij}\colon U \to \mathbb{R}$ defined for $i, j, k = 1, 2$ by

$$\begin{pmatrix} \Gamma^1_{ij} \\ \Gamma^2_{ij} \end{pmatrix} = \begin{pmatrix} g_{11} & g_{12} \\ g_{21} & g_{22} \end{pmatrix}^{-1} \begin{pmatrix} \sigma''_{ij} \cdot \sigma'_1 \\ \sigma''_{ij} \cdot \sigma'_2 \end{pmatrix}. \tag{4}$$

At any given point $p \in U$ the three vectors

$$\sigma'_u,\ \sigma'_v,\ \mathbf{N} \tag{5}$$

constitute a basis for $\mathbb{R}^3$, which can be seen as analogous to the moving frame $(\mathbf{t}, \mathbf{n}, \mathbf{b})$ of a curve (see Section 4.7), although in general (5) is not orthonormal. The motivation for the symbols Γ^k_{ij} is that together with the coefficients b_{ij} of the second fundamental form they appear in the representation of σ''_{ij} with respect to the basis (5).

Theorem 6.3. *Let coefficients Γ^k_{ij} for $i, j, k = 1, 2$ be defined as above. Then*

$$\sigma''_{ij} = \Gamma^1_{ij}\sigma'_1 + \Gamma^2_{ij}\sigma'_2 + b_{ij}\mathbf{N}. \tag{6}$$

Proof. It follows from definition (4) that

$$\begin{aligned} g_{11}\Gamma^1_{ij} + g_{12}\Gamma^2_{ij} &= \sigma''_{ij} \cdot \sigma'_1 \\ g_{21}\Gamma^1_{ij} + g_{22}\Gamma^2_{ij} &= \sigma''_{ij} \cdot \sigma'_2. \end{aligned}$$

Since $\sigma'_l \cdot \sigma'_k = g_{lk}$ we then obtain

$$\left(\Gamma^1_{ij}\sigma'_1 + \Gamma^2_{ij}\sigma'_2 + b_{ij}\mathbf{N}\right) \cdot \sigma'_k = g_{k1}\Gamma^1_{ij} + g_{k2}\Gamma^2_{ij} = \sigma''_{ij} \cdot \sigma'_k.$$

On the other hand since $\mathbf{N}$ is a unit vector

$$\left(\Gamma^1_{ij}\sigma'_1 + \Gamma^2_{ij}\sigma'_2 + b_{ij}\mathbf{N}\right) \cdot \mathbf{N} = b_{ij} = \sigma''_{ij} \cdot \mathbf{N}$$

Thus the vectors on each side of (6) have equal dot products with all elements of a basis. This implies that they are equal. $\square$

The following corollary expresses that the Christoffel symbols can be determined from E, F and G. However, they are not intrinsic invariants, since in general they change when the surface is reparametrized (see Example 6.3.2).

Corollary 6.3. *The Christoffel symbols Γ_{ij}^k are intrinsic. They can be expressed by a formula which involves only the coefficients of the first fundamental form and their (first order) derivatives with respect to u and v.*

Proof. Immediate from (3) and (4). □

The actual formula for Γ_{ij}^k is somewhat complicated, and the fact that it exists is more important than its detailed appearance. Let the inverse matrix of g_{ij} be denoted by g^{ij}, with superscript indices, then it follows from equations (3) and (4) that

$$\Gamma_{ij}^k = \frac{1}{2}\sum_l g^{kl}\left(\frac{\partial g_{il}}{\partial u_j} + \frac{\partial g_{jl}}{\partial u_i} - \frac{\partial g_{ij}}{\partial u_l}\right). \tag{7}$$

If we insert this formula (7) into (6), we obtain an expression for σ''_{ij} which is called the *formula of Gauss.*

Consider in particular the case where we have an orthogonal parametrization, that is, where $F = 0$. In this case the formulas (3) and (7) become considerably simpler and can be expressed in our original notation of E, F and G as follows:

$$\begin{aligned}
\sigma''_{11}\cdot\sigma'_1 &= \frac{1}{2}E'_u, & \sigma''_{12}\cdot\sigma'_1 &= \frac{1}{2}E'_v, & \sigma''_{22}\cdot\sigma'_1 &= -\frac{1}{2}G'_u,\\
\sigma''_{11}\cdot\sigma'_2 &= -\frac{1}{2}E'_v, & \sigma''_{12}\cdot\sigma'_2 &= \frac{1}{2}G'_u, & \sigma''_{22}\cdot\sigma'_2 &= \frac{1}{2}G'_v,
\end{aligned}$$

and

$$\begin{aligned}
\Gamma_{11}^1 &= \frac{1}{2E}E'_u, & \Gamma_{12}^1 = \Gamma_{21}^1 &= \frac{1}{2E}E'_v, & \Gamma_{22}^1 &= -\frac{1}{2E}G'_u,\\
\Gamma_{11}^2 &= -\frac{1}{2G}E'_v, & \Gamma_{12}^2 = \Gamma_{21}^2 &= \frac{1}{2G}G'_u, & \Gamma_{22}^2 &= \frac{1}{2G}G'_v.
\end{aligned}$$

Example 6.3.1 It follows from the definition in (4) that the Christoffel symbols for a plane $\sigma(u,v) = p + uq_1 + vq_2$ are all zero, since all the second derivatives σ''_{ij} vanish. This can be seen as well from the formulas above, since $E = G = 1$ and $F = 0$ in this case.

Example 6.3.2 Consider the xy-plane with polar coordinates $\sigma(u,v) = (u\cos v, u\sin v, 0)$. Here $\sigma'_u = (\cos v, \sin v, 0)$ and $\sigma'_v = (-u\sin v, u\cos v, 0)$, and hence $E = 1$, $F = 0$ and $G = u^2$. By insertion in the formulas above we see that the Christoffel symbols are $\Gamma_{11}^1 = \Gamma_{12}^1 = \Gamma_{11}^2 = \Gamma_{22}^2 = 0$, $\Gamma_{22}^1 = -u$ and $\Gamma_{12}^2 = \frac{1}{u}$. In particular, they differ from those of the preceding example.

6.4 The remarkable theorem of Gauss

The following theorem was found by Gauss in 1827, who described it (in latin) as ‘egregium’, most remarkable. Since then it has become customary to call it ‘teorema egregium’.

Theorem 6.4. *The Gauss curvature K is intrinsic, that is, there exists a general formula expressing K by means of the component functions E, F and G of the first fundamental form.*

Proof. More precisely, we will show that a formula can be given, which expresses the value of K in a given point by means of the values of E, F and G and their derivatives (with respect to u and v) up to order 2 in this point. Since we have already seen (below Definition 6.1) that K is invariant under reparametrization, the theorem then follows.

We use the notation from the preceding section. Since

$$K = \frac{\det(b_{ij})}{\det(g_{ij})} \tag{8}$$

it suffices to show that the determinant of the matrix (b_{ij}) can be expressed in terms of the component functions g_{ij} and their derivatives.

From the expression (see Theorem 6.3)

$$\sigma''_{ij} = \sum_{m=1}^{2} \Gamma_{ij}^{m} \sigma'_m + b_{ij}\mathbf{N}$$

we obtain by differentiation with respect to u_k

$$\sigma'''_{ijk} = \sum_{m=1}^{2} (\frac{\partial \Gamma_{ij}^{m}}{\partial u_k} \sigma'_m + \Gamma_{ij}^{m} \sigma''_{mk}) + \frac{\partial b_{ij}}{\partial u_k}\mathbf{N} + b_{ij}\mathbf{N}'_k.$$

It follows that

$$\sigma'''_{ijk} \cdot \sigma'_l = \sum_{m=1}^{2} (\frac{\partial \Gamma_{ij}^{m}}{\partial u_k} g_{ml} + \Gamma_{ij}^{m}\, \sigma''_{mk} \cdot \sigma'_l) + b_{ij}\mathbf{N}'_k \cdot \sigma'_l,$$

and since

$$\mathbf{N}'_k \cdot \sigma'_l = -\mathbf{N} \cdot \sigma''_{lk} = -b_{lk}$$

(see Section 5.3, (9)-(10)) we obtain

$$\sigma'''_{ijk} \cdot \sigma'_l = \sum_{m=1}^{2} (\frac{\partial \Gamma_{ij}^{m}}{\partial u_k} g_{ml} + \Gamma_{ij}^{m}\, \sigma''_{mk} \cdot \sigma'_l) - b_{ij} b_{lk}.$$

We introduce the abbreviation R_{jkil}, called the *Riemann symbol*, for the difference

$$R_{jkil} = \sum_{m=1}^{2}(\frac{\partial \Gamma_{ij}^m}{\partial u_k} g_{ml} + \Gamma_{ij}^m\, \sigma''_{mk} \cdot \sigma'_l) - \sum_{m=1}^{2}(\frac{\partial \Gamma_{ik}^m}{\partial u_j} g_{ml} + \Gamma_{ik}^m\, \sigma''_{mj} \cdot \sigma'_l), \quad (9)$$

where the two sums only differ by j and k being interchanged.

Then since $\sigma'''_{ijk} = \sigma'''_{ikj}$ we conclude that

$$R_{jkil} - b_{ij}b_{lk} + b_{ik}b_{lj} = \sigma'''_{ijk} \cdot \sigma'_l - \sigma'''_{ikj} \cdot \sigma'_l = 0,$$

hence

$$R_{jkil} = b_{ij}b_{lk} - b_{ik}b_{lj}.$$

In particular,

$$R_{1212} = \det(b_{ij}). \quad (10)$$

The Riemann symbol R was introduced as an abbreviation for an expression involving the quantities Γ_{ij}^m, g_{ij} and $\sigma''_{ij} \cdot \sigma'_k$ (with various indices i, j, k, m). Hence it follows from Lemma 6.3 and Corollary 6.3 that R can be expressed by means of the g_{ij}. An inspection shows that derivatives up to order 2 are involved. According to (10) this implies the statement of the theorem. □

From the equation (10) one can derive an explicit, but quite complicated, expression for the Gauss curvature in terms of the coefficients of the first fundamental form. If $F = 0$ it becomes considerably simpler, and reads

$$K = -\frac{1}{2\sqrt{EG}}\left(\left(\frac{G'_u}{\sqrt{EG}}\right)'_u + \left(\frac{E'_v}{\sqrt{EG}}\right)'_v\right). \quad (11)$$

The verification of this formula is a long but straightforward computation based on (8), (9), (10) and the formulas given in the end of Section 6.3.

6.5 Isometries

A useful interpretation of the notion of intrinsic geometry is obtained from the concept of isometries of surfaces. Basically, an isometry from one surface to another is a distance-preserving map. The definition is simplest for parametrized surfaces that have a common domain U, so we shall start by considering this situation.

Definition 6.5.1. Let $\sigma: U \to \mathbb{R}^3$ and $\rho: U \to \mathbb{R}^3$ be parametrized surfaces defined on a common open set $U \subset \mathbb{R}^2$. Then σ and ρ are said to be *isometric* if their first fundamental forms are equal, that is if

$$E_\sigma = E_\rho, \quad F_\sigma = F_\rho, \quad G_\sigma = G_\rho.$$

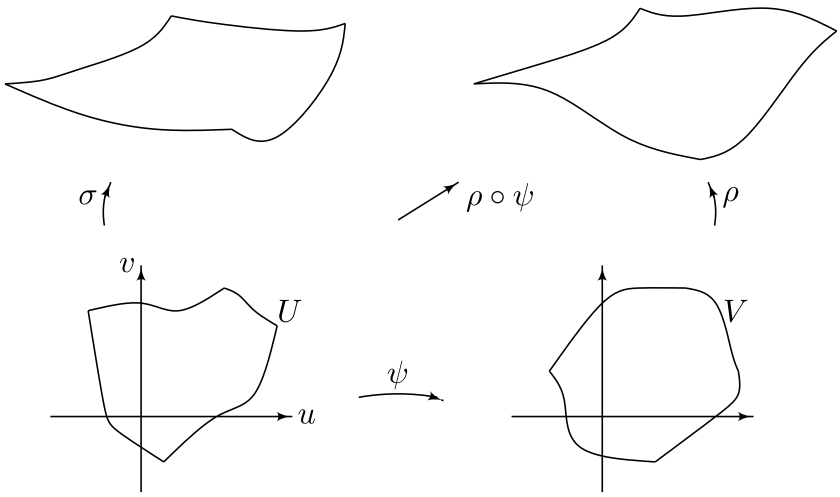
In order to provide some intuition assume temporarily that σ is injective. In this case we can define a map

$$\Psi\colon \sigma(U) \to \rho(V), \qquad \Psi(\sigma(p)) = \rho(p), \qquad (p \in U).$$

When the surfaces are isometric, this map is said to be a *bending* of one surface to the other, because the deformation (without stretching) for example of a piece of paper, provides an example.

Example 6.5.1 Let $\sigma(u,v) = (u,v,0)$ and $\rho(u,v) = (\cos v, \sin v, u)$ be the plane and the cylinder, both defined on $U = \mathbb{R}^2$. Then σ and ρ both have $E = G = 1$, $F = 0$, hence they are isometric. In this case the bending $\Psi\colon \sigma(U) \to \rho(U)$ corresponds to the folding of a cylinder from a plane piece of paper.

We now turn to the general situation of parametrized surfaces defined on different domains, say $\sigma\colon U \to \mathbb{R}^3$ and $\rho\colon V \to \mathbb{R}^3$. We assume that a diffeomorphism $\psi\colon U \to V$ is given. Then $\rho\circ\psi\colon U \to \mathbb{R}^3$ is a reparametrization of ρ (see Section 2.6) with the same domain U as σ.



Definition 6.5.2. The diffeomorphism $\psi\colon U \to V$ is said to *induce an isometry* from σ to ρ, if σ and $\rho \circ \psi$ are isometric, that is, if

$$E_\sigma = E_{\rho\circ\psi}, \quad F_\sigma = F_{\rho\circ\psi}, \quad G_\sigma = G_{\rho\circ\psi}. \tag{12}$$

It is important to stress that the condition expressed in (12) is that σ should have the same first fundamental form as ρ, but *after* the reparametrization by ψ.

Note that a reparametrization of a surface is a trivial case of an isometry. More precisely, if $\sigma = \rho \circ \psi$ is a reparametrization of ρ, then ψ induces an isometry from σ to ρ, since obviously σ is isometric to itself. The purpose of the more involved Definition 6.5.2, compared to the previous one, is exactly to get a notion of isometry that takes possible reparametrizations into account.

If in Definition 6.5.2 we assume that σ is injective, then

$$\Psi: \sigma(U) \to \rho(V), \qquad \Psi(\sigma(p)) = \rho(\psi(p)), \qquad (p \in U)$$

is well-defined. Intuitively it is this bending, called the *lift* of ψ, which is the isometry induced by ψ. It takes place between the images of the parameter sets. However, if σ is not injective, then the construction of Ψ may not be possible. Different elements p and q in U may have $\sigma(p) = \sigma(q)$ but $\rho(\psi(p)) \neq \rho(\psi(q))$, so that Ψ is not well defined. For example, the 'unfolding' from cylinder to plane, which is 'inverse' to the bending in Example 6.5.1, is really only a well-defined map on the level of the parameter sets, since the same point on the cylinder corresponds to more than one point in the plane.

Notice that by Corollary 2.11, a regular parametrized surface is injective in some neighborhood of each parameter point $p \in U$, so that the 'lifting' can be done in that neighborhood.

Example 6.5.2 Let σ and ρ both denote the sphere of radius 1, both parametrized by spherical coordinates as in Example 1.2.2, with domains

$$U = \{(u, v) \mid -\pi/2 < u < \pi/2,\ -\pi < v < \pi\}$$

for σ and

$$V = \{(s, t) \mid -\pi/2 < s < \pi/2,\ -\pi + \alpha < t < \pi + \alpha\}$$

for ρ. Here $\alpha \in \mathbb{R}$ is some constant. The map $U \to V$ defined by $\psi(u, v) = (u, v + \alpha)$ induces an isometry from σ to ρ. This follows from the fact that E, F and G are independent of v (see Example 3.4.3). The corresponding lift is the rotation of the sphere around the z-axis by the angle α.

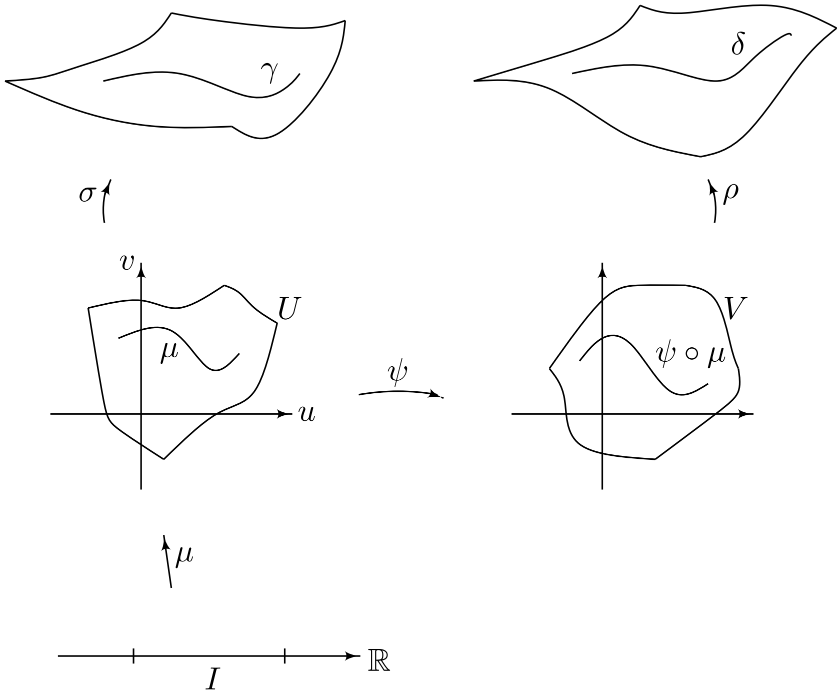
It can be shown (see Exercise 8), that if ψ induces an isometry from σ to ρ, then ψ^{-1} induces an isometry from ρ to σ. Moreover, if in addition a third parametrized surface $\tau: W \to \mathbb{R}^3$ is given, together with a diffeomorphism $\phi: V \to W$ inducing an isometry from ρ to τ, then $\phi \circ \psi$ induces an isometry from σ to τ.

The most important observation in connection with the concept of isometry is that the agreement of the first fundamental forms, as expressed by (12), ensures that all intrinsic quantities are preserved. In particular, this explains the term 'isometry', since length is intrinsic. The fact that length is preserved is expressed more precisely in the following lemma.

Let $\gamma = \sigma \circ \mu: I \to \mathbb{R}^3$ be a parametrized curve on σ, and assume ψ induces isometry from σ to ρ, as above. By $\delta = \rho \circ \psi \circ \mu: I \to \mathbb{R}^3$ we define a parametrized curve on ρ, said to be the image of γ by ψ (see the figure below).

Lemma 6.5. *When ψ induces an isometry the arc lengths of γ and δ are equal. That is, let $t_1, t_2 \in I$ then the arc length of γ from t_1 to t_2 is equal to the arc length of δ from t_1 to t_2.*

Proof. Let $\mu(t) = (u(t), v(t))$ denote the coordinates of $\gamma(t)$ in the parametrization $\gamma = \sigma \circ \mu$ by means of σ. The arc length of γ is expressed in Theorem 3.4 by means of the functions $u(t)$ and $v(t)$ together with E_σ, F_σ, G_σ.



Writing $\delta = (\rho \circ \psi) \circ \mu$ we can regard δ as a parametrized curve on $\rho \circ \psi$. When we regard δ in this fashion, its coordinates $(u(t), v(t))$ are those of $\mu(t)$, that is, they are the same as before. Applying Theorem 3.4 once more, but this time to δ on $\rho \circ \psi$, we obtain an expression for the arc length of δ by means of $u(t)$ and $v(t)$ together with the coefficients $E_{\rho\circ\psi}$, $F_{\rho\circ\psi}$, $G_{\rho\circ\psi}$ of the first fundamental form of $\rho \circ \psi$. Hence if ψ induces an isometry, the expression is exactly the same as before, and the arc lengths on γ and δ agree. $\square$

Theorem 6.5. *Assume that $\psi: U \to V$ induces an isometry from σ to ρ. Then the Gauss curvature of σ in p is equal to the Gauss curvature of ρ in $\psi(p)$, for all $p \in U$.*

We say that the Gauss curvature is *invariant* under isometries.

Proof. By Theorem 6.4 the Gauss curvature in p can be expressed by means of the functions E, F and G and their derivatives in p. Hence the Gauss curvatures K_σ and $K_{\rho\circ\psi}$ for σ and $\rho\circ\psi$ are identical functions on U,

$$K_\sigma(p) = K_{\rho\circ\psi}(p), \quad p \in U.$$

It was observed in Section 6.1 that the Gauss curvature is unchanged by reparametrizations, hence

$$K_{\rho\circ\psi}(p) = K_\rho(\psi(p)). \quad \square$$

Example 6.5.3 Let $\sigma\colon U = \{(u,v) \mid u > 0\} \to \mathbb{R}^3$ be the parametrization

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

of a cone (see Example 1.2.4) and let $\rho\colon V = \{(r,\theta) \mid r > 0\} \to \mathbb{R}^3$ be the parametrization by polar coordinates

$$\rho(r,\theta) = (r\cos\theta, r\sin\theta, 0)$$

of the xy-plane (without $(0,0,0)$).

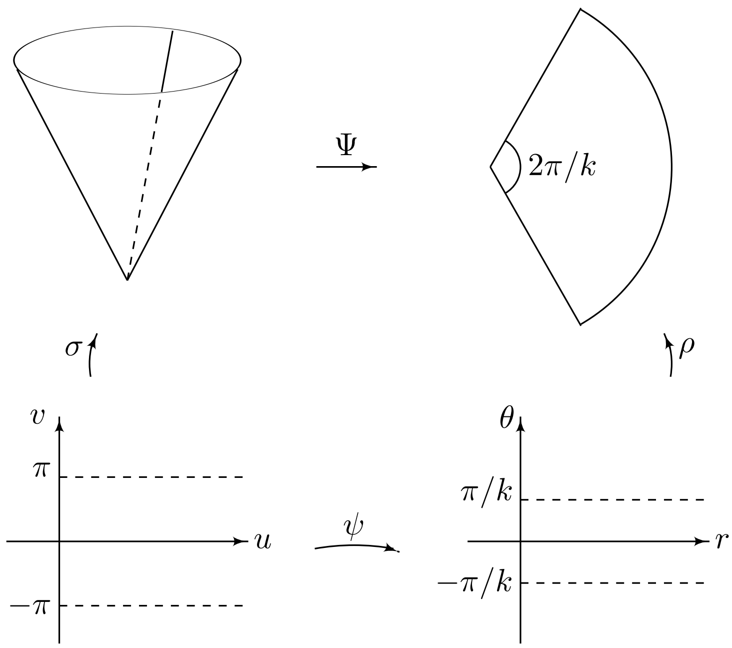
For each constant $k > 0$ the map $\psi(u,v) = (ku, v/k)$ is a diffeomorphism of U to V, since it is smooth and bijective with the smooth inverse $(r,\theta) \mapsto (u,v) = (r/k, k\theta)$.

The component functions of the first fundamental form for σ are $E = 1+\lambda^2$, $F = 0$ and $G = u^2$. The reparametrization

$$\rho\circ\psi(u,v) = (ku\cos(v/k), ku\sin(v/k), 0)$$

of ρ has components $\widetilde{E} = k^2$, $\widetilde{F} = 0$ and $\widetilde{G} = u^2$. Therefore, ψ induces an isometry of σ to ρ if and only if $k^2 = 1+\lambda^2$.

The conclusion from the theorem above is then that if $k^2 = 1+\lambda^2$ then the cone and the plane have the same Gaussian curvature in points $\sigma(u,v)$ and $\rho(ku, v/k)$ (in fact, both Gaussian curvatures are zero, as we knew already).



Notice that the comparison of component functions took place between those of σ and those of $\rho \circ \psi$, whereas those of ρ itself played no role.

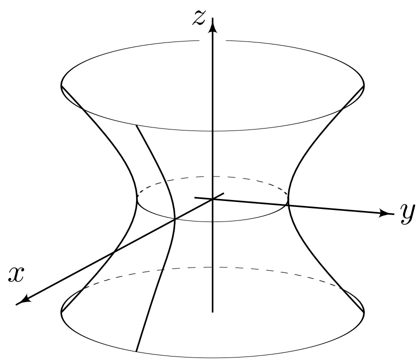
The map $\psi\colon (u,v) \mapsto (r,\theta) = (ku, v/k)$ between parameter values is 'lifted' to the map

$$\Psi\colon \sigma(u,v) \mapsto \rho(ku, v/k)$$

from cone to plane. However, Ψ is well defined only if we restrict v to an open interval of length $\leq 2\pi$, since we have $\sigma(u, v+2\pi) = \sigma(u,v)$ but in general $\rho(\psi(u, v+2\pi)) \neq \rho(\psi(u,v))$. The map Ψ can be described as the 'unfolding' of the cone.

Example 6.5.4 Since the sphere has Gauss curvature different from zero in all points, we can conclude from the Gauss theorem that *no portion of a sphere can be mapped isometrically into a plane.* In other words, it is impossible to draw a map of a portion of the globe, which preserves all lengths (in appropriate units). Such a map is called an *ideal map*, and its non-existence is a theorem originally due to Euler.

Example 6.5.5 Let $\sigma(u,v) = (a\cosh u \cos v, a\cosh u \sin v, au)$ for $(u,v) \in U = \mathbb{R}^2$, where $a > 0$ is a constant. This surface is called a *catenoid* (it is a surface of revolution, see page 35).



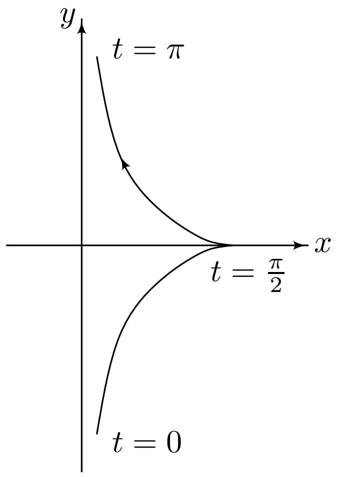
For the second surface let $\rho(s,t) = (s\cos t, s\sin t, at)$ where $(s,t) \in V = \mathbb{R}^2$. This surface is called a *helicoid* (see page 56). We shall verify that the map $\psi(u,v) = (a\sinh u, v)$ induces an isometry from σ to ρ. It is a diffeomorphism since $\sinh\colon \mathbb{R} \to \mathbb{R}$ is bijective with a smooth inverse (by Theorem 2.5). An elementary computation shows that the first fundamental forms for σ and for $\rho \circ \psi$ are given by $E = G = a^2\cosh^2 u$ and $F = 0$ in both cases. Hence ψ induces an isometry.

Notice that the catenoid is not injective. If we restrict to the subset $\{(u,v) \mid -\pi < v < \pi\}$ of U, then σ is injective. The image by ψ of this set corresponds to one winding of the helicoid.

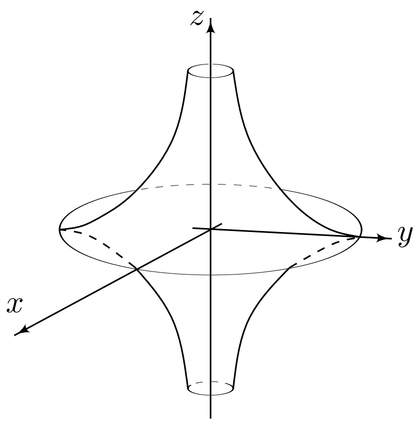
6.6 Exercises

1 Verify the following formula for the Gauss curvature of a surface of revolution
$$K = \frac{(f'g'' - f''g')g'}{(f'^2 + g'^2)^2 f}.$$
Show that if the profile curve has unit speed, then $K = -\frac{f''}{f}$.

2 The plane curve $\gamma(t) = (\sin t, \cos t + \ln \tan \frac{t}{2})$, where $0 < t < \pi$ is called the *tractrix*.



Show that the curve is regular for $t \neq \frac{\pi}{2}$. The surface of revolution



$$\sigma(u, v) = (\sin u \cos v, \sin u \sin v, \cos u + \ln \tan \frac{u}{2}), \quad 0 < u < \pi, v \in \mathbb{R}$$

is called a *pseudosphere*. Verify that $K = -1$ everywhere, except at $u = \frac{\pi}{2}$, (so that σ resembles a sphere of radius 1, which has constant $K = 1$).

3 Compute the coefficients L, M and N for the surfaces σ and ρ in Example 6.5.5 (see also Exercises 2 and 3, page 89), and use these to determine their Gauss curvatures. Verify the Teorema Egregium for these surfaces.

4 Show that the surface of revolution $\tau(s,t) = (s\cos t, s\sin t, a\ln t)$, where $(s,t) \in U = \{(s,t) \mid t > 0\}$ and $a > 0$ is constant, has the same Gauss curvature $K(s,t)$ as the helicoid ρ in Exercise 3, restricted to U. Nevertheless, the first fundamental form of τ is different. Does this contradict the Teorema Egregium?

5 **a.** Let three numbers $e, f, g \in \mathbb{R}$ with $eg > f^2$ and $e, g > 0$ be given. Prove that there exists a regular parametrized surface $\sigma: \mathbb{R}^2 \to \mathbb{R}^3$, the image of which is the xy-plane, such that $E(u,v) = e$, $F(u,v) = f$ and $G(u,v) = g$ for all $(u,v) \in \mathbb{R}^2$. Hint: Try a linear map $\sigma : \mathbb{R}^2 \to \mathbb{R}^3$.

b. Let next $\sigma: U \to \mathbb{R}^3$ be an arbitrary regular parametrized surface for which E, F and G are constant. Prove that there exists a diffeomorphism which induces an isometry from σ to a parametrized surface of which the image is contained in the xy-plane.

6 **a.** Let $0 < a < 1$ and let

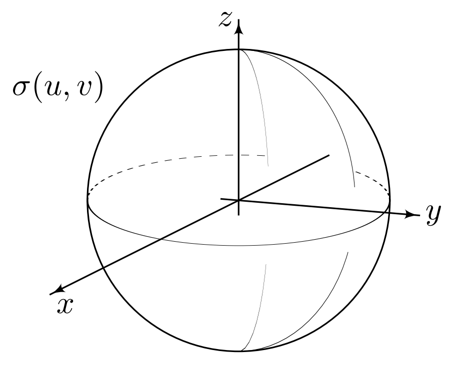
$$f(s) = a\cos s \quad \text{and} \quad g(s) = \int_0^s \sqrt{1 - a^2\sin^2 r}\, dr.$$

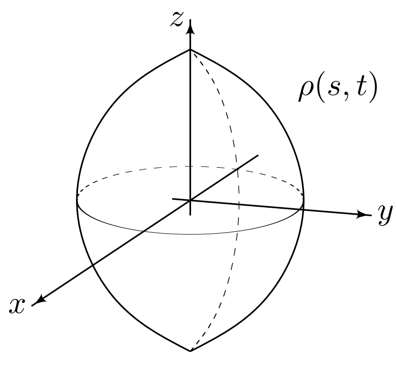
Verify that the curve $\gamma(s) = (f(s), g(s))$ has unit speed.

b. Let $\rho(s,t) = (f(s)\cos t, f(s)\sin t, g(s))$, be the surface of revolution with profile curve γ, where

$$(s,t) \in V = \{(s,t) \mid -\frac{\pi}{2} < s < \frac{\pi}{2}, -\pi < t < \pi\}.$$

Furthermore, let $\sigma(u,v)$ denote the part of a unit sphere with standard spherical coordinates (Example 1.2.2), for which the domain of definition is reduced to $U = \{(u,v) \mid -\frac{\pi}{2} < u < \frac{\pi}{2}, -a\pi < v < a\pi\}$, that is, a segment on the back has been removed.





Show that the map $\psi: U \to V$ given by $\psi(u,v) = (u, v/a)$ induces an isometry (for $a = \frac{1}{2}$ one can visualize ψ by the bending of a half sphere, for example the peel of half an orange). What can one conclude about the curvature of ρ?

7 Let $\psi(u,v) = (u, v + \frac{c}{u})$ for $(u,v) \in U = \{(u,v) \in \mathbb{R}^2 \mid u > 0\}$, where $c \in \mathbb{R}$ is a constant.

a. Let $\sigma\colon U \to \mathbb{R}^3$ be a parametrized smooth surface, and let $\tau = \sigma \circ \psi$. Verify

$$\tau'_u(u,v) = \sigma'_u(\psi(u,v)) - \frac{c}{u^2}\sigma'_v(\psi(u,v)),$$

and determine a similar expression for $\tau'_v(u,v)$.

b. Assume that the first fundamental form for σ is given by

$$E = 1 + v^2, \quad F = uv, \quad G = u^2$$

for $(u,v) \in U$. Show that ψ induces an isometry from σ to itself (that is, take $V = U$ and $\rho = \sigma$ in Definition 6.5.2).

c. Without explicitly computing the Gauss curvature $K(u,v)$ of σ, show that it does not depend on v (hint: use that c was arbitrary).

8 Prove the following statements (see page 102) by applying the chain rule and the identity (3) in Section 3.4:

a. If ψ induces an isometry from σ to ρ, then ψ^{-1} induces an isometry from ρ to σ.

b. If in addition a third parametrized surface $\tau\colon W \to \mathbb{R}^3$ is given, together with a diffeomorphism $\phi\colon V \to W$ inducing an isometry from ρ to τ, then $\phi \circ \psi$ induces an isometry from σ to τ.

Chapter 7

Geodesics

In this final chapter we investigate some properties of geodesics. Recall from Definition 4.10, that a geodesic on a surface is a curve with zero geodesic curvature. We shall see that the property of a curve, that it is geodesic, is intrinsic. Furthermore we introduce the notion of geodesic coordinates on a surface, and we use these to give a geometric interpretation of the theorem of Gauss. Some of the results presented in this chapter require more advanced analytic tools than we have presupposed in the rest of the notes, and we shall be content with stating them without proof.

7.1 The geodesic equations

We aim to show that the absolute value $|\kappa_g(t)|$ of geodesic curvature is an intrinsic property of a curve on a surface. It is invariant under reparametrizations by Theorem 4.8 (but notice the necessity of taking the absolute value).

Theorem 7.1. *Let $\gamma = \sigma \circ \mu$ be a regular parametrized curve on σ. The geodesic curvature $\kappa_g(t)$ satisfies*

$$\kappa_g = \|\gamma'\|^{-3} \det(g_{ij})^{1/2} \big((u_1)'\Lambda_2 - (u_2)'\Lambda_1\big)$$

where g_{ij} is the first fundamental form of σ at $\mu(t)$, u_1, u_2 are the coordinates of $\mu(t)$ and $(u_1)', (u_2)'$ are their derivatives with respect to t, and where Λ_i denotes the function

$$\Lambda_i(t) = (u_i)''(t) + \sum_{j,k=1}^{2} \Gamma^i_{jk}(\mu(t))\,(u_j)'(t)(u_k)'(t), \qquad i = 1, 2,$$

for $i = 1, 2$, in terms of the Christoffel symbols Γ^i_{jk}.

In view of Corollary 6.3, we see that the expressions Λ_i can be determined from E, F and G. Hence it follows from the equation above for κ_g, that it too can be determined. Hence the absolute value $|\kappa_g|$ is intrinsic. In particular, it follows that the property of being a geodesic curve is intrinsic.

The proof invokes two lemmas, which are stated and proved on the following page.

Proof. By definition $\kappa_g = \|\gamma'(t)\|^{-3} \det[\gamma' \, \gamma'' \, \mathbf{m}]$. Recall from Lemma 2.4 that

$$\gamma'(t) = u'(t)\sigma'_u + v'(t)\sigma'_v. \tag{1}$$

where the tangent vectors σ'_u and σ'_v are evaluated in $(u(t), v(t))$. The second derivative γ'' is determined in Lemma 7.1.1 below, and the determinant can then be computed by means of Lemma 7.1.2, where we take $w' = u'_1\sigma'_1 + u'_2\sigma'_2$ and $w'' = \Lambda_1\sigma'_1 + \Lambda_2\sigma'_2$ (since a multiple of $\mathbf{m}$ in γ'' does not contribute to the determinant). The equation for κ_g follows. $\square$

Lemma 7.1.1. *Let γ and Λ_1, Λ_2 be as above. Then γ'' equals $\Lambda_1\sigma'_1 + \Lambda_2\sigma'_2$ plus a multiple $c\mathbf{m}$ of $\mathbf{m}$.*

The factor is $c = \sum b_{jk}(u_j)'(u_k)'$ but we do not need this formula.

Lemma 7.1.2. *Let γ be as above and let two vectors $w', w'' \in T_{\mu(t)}$ be given. If $w' = a_1\sigma'_1 + a_2\sigma'_2$ and $w'' = b_1\sigma'_1 + b_2\sigma'_2$ then*

$$\det[w' \, w'' \, \mathbf{m}] = \det(g_{ij})^{1/2} \, (a_1b_2 - b_1a_2)$$

Proof of Lemma 7.1.1. In order to determine $\gamma''(t)$ we differentiate (1). For this we need to differentiate σ'_u and σ'_v with respect to t.

We apply the chain rule to the function $t \mapsto \sigma'_u(u(t), v(t))$. It follows that

$$\frac{d}{dt}\sigma'_u(u(t), v(t)) = u'(t)\sigma''_{uu} + v'(t)\sigma''_{uv}.$$

Similarly

$$\frac{d}{dt}\sigma'_v(u(t), v(t)) = u'(t)\sigma''_{vu} + v'(t)\sigma''_{vv}.$$

Hence

$$\begin{aligned}
\gamma''(t) &= u''(t)\sigma'_u + u'(t)\frac{d}{dt}\sigma'_u + v''(t)\sigma'_v + v'(t)\frac{d}{dt}\sigma'_v \\
&= u''(t)\sigma'_u + u'(t)(u'(t)\sigma''_{uu} + v'(t)\sigma''_{uv}) \\
&\qquad + v''(t)\sigma'_v + v'(t)(u'(t)\sigma''_{vu} + v'(t)\sigma''_{vv}) \\
&= \sum_i u''_i\sigma'_i + \sum_{jk} u'_ju'_k\sigma''_{jk}.
\end{aligned}$$

We use the expression (6) from Theorem 6.3 and insert it for σ''_{jk}. It follows that $\gamma'' = \sum_i \Lambda_i\sigma'_i + c\mathbf{m}$ for the number c mentioned below the lemma. $\square$

Proof of Lemma 7.1.2. By a straightforward computation

$$\det[w' \, w'' \, \mathbf{m}] = (w' \times w'') \cdot \mathbf{m} = (a_1b_2 - b_1a_2)\,(\sigma'_1 \times \sigma'_2) \cdot \mathbf{m}.$$

The lemma follows since by (9) page 53

$$\sigma'_1 \times \sigma'_2 = \|\sigma'_1 \times \sigma'_2\| \, \mathbf{m} = (EG - F^2)^{1/2} \, \mathbf{m}. \quad \square$$

Corollary 7.1. *Let $\gamma(s) = \sigma(u_1(s), u_2(s))$ be a regular parametrized smooth curve on σ. Then γ is a geodesic and has constant speed if and only if the coordinate functions u_1 and u_2 satisfy the following system of second order differential equations*

$$(u_i)'' + \sum_{j,k=1}^{2} \Gamma^i_{jk}\,(u_j)'(u_k)' = 0, \qquad i = 1, 2. \tag{2}$$

with coefficients Γ^i_{jk} evaluated at $\mu(t)$.

Proof. The system of equations (2) is written $\Lambda_1 = \Lambda_2 = 0$ in the notation of the preceding theorem. It follows from Lemma 7.1.1 that this condition holds exactly when $\gamma''(t)$ is proportional to $\mathbf{m}$ for all t. The corollary now follows from Theorem 4.10. $\square$

The differential equations (2) are called the *geodesic equations*. By Corollary 7.1 the determination of the geodesics on a given surface is a matter of solving these equations. However, for a general surface they are quite complicated non-linear differential equations which are not easy to solve.

As mentioned, the property of a curve of being a geodesic is intrinsic. It follows that an isometry will carry geodesic curves to geodesic curves.

Example 7.1.1 Consider again the isometry ψ in Example 6.5.3 from cone to plane. The geodesics on the plane are the straight line segments, hence we conclude that a curve on the cone is a geodesic if and only if its image by ψ is a line segment in V.

For example, the plane unit speed line $\delta(s) = (1, s, 0)$ is in polar coordinates $\rho(r, \theta) = (r\cos\theta, r\sin\theta, 0)$ given by

$$\delta(s) = \rho(r(s), \theta(s)) = \rho(\sqrt{1+s^2}, \tan^{-1} s)).$$

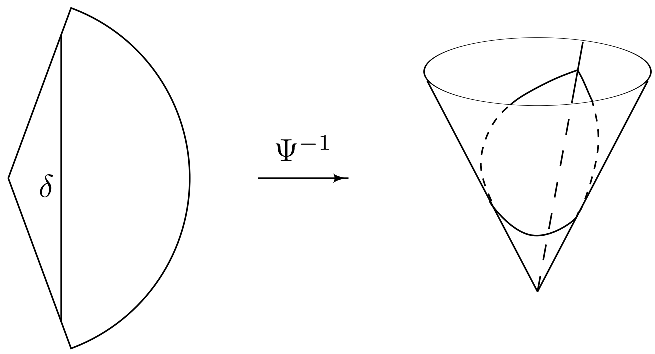
The image by ψ^{-1} is then

$$\gamma(s) = \sigma(\psi^{-1}(r(s), \theta(s))) = \sigma(k^{-1}\sqrt{1+s^2}, k\tan^{-1} s).$$

Recall that $\sigma(u, v) = (u\cos v, u\sin v, \lambda u)$. We obtain that

$$\begin{aligned}&\gamma(s)\\ &\quad = (k^{-1}\sqrt{1+s^2}\cos(k\tan^{-1} s), k^{-1}\sqrt{1+s^2}\sin(k\tan^{-1} s), \lambda k^{-1}\sqrt{1+s^2})\end{aligned}$$

is a geodesic on the cone when $k^2 = 1+\lambda^2$. An idea of the shape of the curve can be obtained by folding a cone out of a piece of paper with a straight line drawn on it (see the following figure).



7.2 Existence of geodesics

A further analysis of geodesics on a surface can be based on the differential equations (2). This requires the use of the fundamental theorem of existence and uniqueness of solutions of ordinary differential equations. Without going into details, we cite the following important consequence.

Theorem 7.2. *Through every point of a regular parametrized surface passes a unique geodesic curve in each direction.*

More precisely, let $p \in U$ and $w \in T_p\sigma \setminus \{0\}$ be given. There exists a geodesic curve $\gamma = \sigma \circ \mu : I \to \mathbb{R}^3$ on σ with

$$p = \mu(t_0) \quad \textit{and} \quad w = \gamma'(t_0) \tag{3}$$

for some $t_0 \in I$. Moreover, if two unit speed geodesics defined on intervals I, J both satisfy (3) *for some common $t_0 \in I \cap J$, then they agree on $I \cap J$.*

Proof. Omitted.

This property is of course well known for lines on a plane.

Example 7.2.1 Through every point on a sphere passes a unique great circle in each direction, namely the great circle obtained as the intersection of the sphere with the unique plane through the center of the sphere which contains the given point and the given direction vector.

7.3 Geodesic coordinates

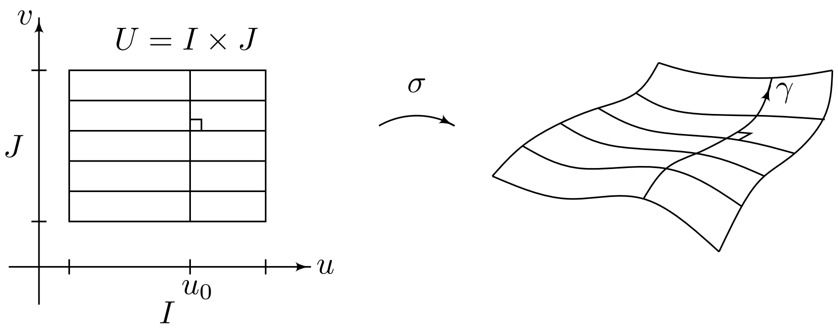
We shall now describe a particularly useful type of parametrization of a surface.

Definition 7.3. Let $\gamma : J \to \mathbb{R}^3$ be a unit speed curve. A regular parametrized surface $\sigma : U \to \mathbb{R}^3$ is called a *geodesic coordinate system* transversal to γ if $U = I \times J$ for some interval I and

(i) there exists $u_0 \in I$ such that $\gamma(v) = \sigma(u_0, v)$ for all v, and this curve is a geodesic on σ,

(ii) all the coordinate curves $I \ni u \mapsto \sigma(u,v)$ are unit speed geodesics on σ, which intersect orthogonally with γ (that is, the tangent vector $\sigma'_u(u_0, v)$ is orthogonal to $\gamma'(v) = \sigma'_v(u_0, v)$ for all $v \in J$).

Notice that while we are requiring $\sigma(u,v)$ to be geodesic as a function of u for all fixed v, we are only requiring it to be geodesic as a function of v for the fixed value u_0 of u, where it produces the original curve γ.



Example 7.3.1 The standard coordinates (x, y) on the xy-plane are geodesic coordinates. Perhaps more interestingly, the spherical coordinates $\sigma(u,v)$ on the unit sphere is a geodesic coordinate system. Indeed, the curve $\gamma(v) = \sigma(0, v)$, the 'equator', is geodesic, and the meridians $u \mapsto \sigma(u,v)$ are geodesics that intersect orthogonally with γ. Notice that in this case the curves $v \mapsto \sigma(u,v)$ are small circles if $u \neq 0$, hence not geodesics.

Theorem 7.3(Existence of geodesic coordinates). *Let $\sigma: U \to \mathbb{R}^3$ be a regular parametrized surface, and let a point $p \in U$ and a unit speed geodesic $\gamma = \sigma \circ \mu$ on σ be given with $\mu(0) = p$. There exists an open rectangle $W = I \times J$ around $(0,0)$ in $\mathbb{R}^2$ and a diffeomorphism ϕ of W onto an open neighborhood $U' \subset U$ of p such that $\phi(0,0) = p$ and such that the reparametrization*

$$\tau(s,t) = \sigma(\phi(s,t))$$

of $\sigma|_{U'}$ is a geodesic coordinate system transversal to $\gamma|_J$.

Proof. The proof which relies on Theorem 7.2 is omitted. □

7.4 The first fundamental form of a geodesic coordinate system

Let $\sigma: U \to \mathbb{R}^3$ be a regular surface, defined on a set $U \subset \mathbb{R}^2$ of the form $U = I \times J$ with open intervals $I, J \subset \mathbb{R}$. Let $u_0 \in I$ be fixed, and let $\gamma: J \to \mathbb{R}^3$ denote the curve $t \mapsto \sigma(u_0, t)$ on σ.

Theorem 7.4. *The surface σ is a geodesic coordinate system transversal to γ if and only if the following condition hold.*

The coefficients of the first fundamental form satisfy

$$E(u,v)=1, \quad F(u,v)=0$$

for all $(u,v)\in U$ and

$$G(u_0,v)=1, \quad G'_u(u_0,v)=0$$

for all $v\in J$.

Proof. The proof is based on the lemma below, from which we conclude that $u\mapsto\sigma(u,v)$ is geodesic if and only if

$$E(u,v)=1 \quad \text{and} \quad E'_v(u,v)-2F'_u(u,v)=0 \tag{4}$$

for all u, and (by interchanging u and v in the lemma) $v\mapsto\sigma(u_0,v)$ is geodesic if and only if

$$G(u_0,v)=1 \quad \text{and} \quad G'_u(u_0,v)-2F'_v(u_0,v)=0 \tag{5}$$

for all v.

Assume σ is a geodesic coordinate system. Then (4) and (5) hold for all (u,v). In particular, $E(u,v)=1$ and $G(u_0,v)=1$.

From $E=1$ we conclude that $E'_u=E'_v=0$, hence (4) implies that $F'_u(u,v)=0$, from which we infer that $u\mapsto F(u,v)$ is constant for each v. In fact this constant is 0 because the assumption that the coordinate curves intersect orthogonally with γ implies that $F(u_0,v)=0$. Finally, since $F=0$ the second condition in (5) implies $G'_u(u_0,v)=0$.

The statement 'if' is seen similarly. □

Lemma 7.4. *Let $\sigma\colon U\to\mathbb{R}^3$ be a regular parametrization. The coordinate curve $u\mapsto\sigma(u,v_0)$ is a unit speed geodesic if and only if $E=1$ and $E'_v-2F'_u=0$ in all points of the curve.*

Proof. Unit speed is equivalent with $E=1$. The second derivative of $u\mapsto\sigma(u,v_0)$ is $\sigma''_{11}=\sigma''_{uu}$, hence it follows from Theorem 4.10 that the curve is a geodesic if and only if

$$\sigma''_{11}\cdot\sigma'_k=0 \quad \text{for } k=1,2.$$

By (3) in Lemma 6.3 this condition is equivalent with

$$2\frac{\partial g_{1k}}{\partial u_1}-\frac{\partial g_{11}}{\partial u_k}=0 \quad \text{for } k=1,2.$$

For $k=1$ this equation reads $\frac{\partial g_{11}}{\partial u_1}=0$, which is already a consequence of the unit speed condition $E=1$, and for $k=2$ it reads $2\frac{\partial g_{12}}{\partial u_1}-\frac{\partial g_{11}}{\partial u_2}=0$, which is exactly the last condition of the lemma. □

7.5 Interpretation of the Gauss theorem

Let $\sigma: U \to \mathbb{R}^3$ be a geodesic coordinate system transversal to $\gamma = \sigma \circ \mu$. For simplicity we assume that $u_0 = 0$ so that $\gamma(v) = \sigma(0, v)$. It follows from Theorem 7.4 and the formula (11) in Chapter 6, that Gauss' formula for K in terms of the first fundamental form is

$$K = -\frac{1}{2\sqrt{G}}\left(\frac{G'_u}{\sqrt{G}}\right)'_u.$$

Since $(\sqrt{G})'_u = \frac{G'_u}{2\sqrt{G}}$ we can rewrite the formula as

$$K = -\frac{1}{\sqrt{G}}(\sqrt{G})''_{uu}. \tag{6}$$

We shall now give a geometric interpretation of this formula.

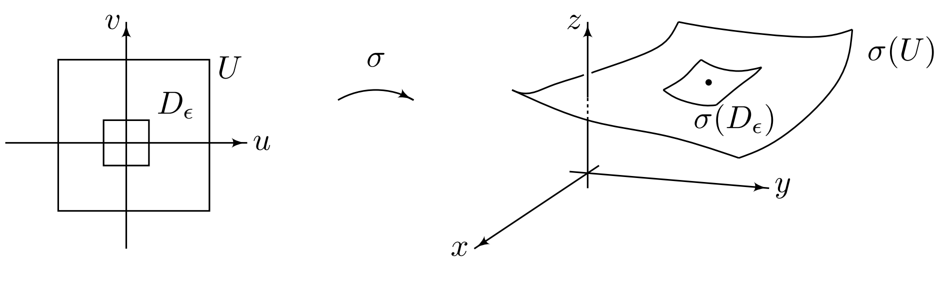
Let $p = (0, 0) = \mu(0) \in U$. For $\epsilon > 0$ let D_ϵ denote the square

$$D_\epsilon = [-\epsilon, \epsilon] \times [-\epsilon, \epsilon]$$

about $(0, 0)$ in $\mathbb{R}^2$. It has area $A(D_\epsilon) = (2\epsilon)^2$. In the following we assume that ϵ is sufficiently small so that $D_\epsilon \subset U$. The set

$$\sigma(D_\epsilon)$$

is called a *square about p on σ*. Its area is denoted $A(\sigma, D_\epsilon)$ (see Section 3.9).



Theorem 7.5. *Let $\sigma: U \to \mathbb{R}^3$ be a geodesic coordinate system around $p = (0, 0) \in U$. The Gauss curvature K of σ in p is given by*

$$K = -\frac{3}{2}\lim_{\epsilon \to 0} \epsilon^{-4}(A(\sigma, D_\epsilon) - A(D_\epsilon)). \tag{7}$$

The interpretation of $K(p)$ is thus that it is a measure for the difference between the area of a small square about p and the corresponding area of a plane square. Since areas are intrinsic properties, and since the properties that went into the definition of a square (geodesics and right angles) are also intrinsic, Gauss' Teorema Egregium is certainly a consequence of this theorem. However, this serves as a geometric explanation rather than a new proof of the theorem, since the proof given below of (7) uses Gauss' formula for K, of which the Teorema is already an immediate consequence.

In particular we notice the minus in the limit formula for K. Thus, in an elliptic point, the area of $\sigma(D_\epsilon)$ will be smaller than that of D_ϵ, for ϵ sufficiently small, and in a hyperbolic point it will be larger.

Proof. We shall use the Taylor approximation formula for the smooth function $f(u,v) = \sqrt{G(u,v)}$, see Appendix B. With $(u_0, v_0) = (0,0)$ it reads

$$\begin{aligned} f(u,v) = f(0,0) &+ f'_u(0,0)u + f'_v(0,0)v \\ &+ \frac{1}{2}(f''_{uu}(0,0)u^2 + 2f''_{uv}(0,0)uv + f''_{vv}(0,0)v^2) + R(u,v) \end{aligned}$$

where the remainder $R(u,v)$ satisfies $|R(u,v)| \le C\|(u,v)\|^3$ in a neighborhood of $(0,0)$ for a constant C.

By Theorem 7.4 we have $G(0,v) = 1$ and $G'_u(0,v) = G'_v(0,v) = 0$. Hence

$$f(0,v) = 1 \quad \text{and} \quad f'_u(0,v) = f'_v(0,v) = 0,$$

and by differentiation with respect to v,

$$f''_{uv}(0,v) = f''_{vv}(0,v) = 0.$$

Finally, by the Gauss formula (6), $f''_{uu}(0,0) = -K$. The Taylor formula is thus

$$\sqrt{G(u,v)} = 1 - \frac{1}{2}Ku^2 + R(u,v).$$

Since $EG - F^2 = G$, the area of $\sigma(D_\epsilon)$ is by definition

$$A(\sigma, D_\epsilon) = \int_{D_\epsilon} \sqrt{G}\, dA$$

and hence

$$\begin{aligned} A(\sigma, D_\epsilon) - A(D_\epsilon) &= \int_{D_\epsilon} \sqrt{G(u,v)} - 1\, dA \\ &= \int_{-\epsilon}^{\epsilon}\int_{-\epsilon}^{\epsilon} -\frac{1}{2}Ku^2 + R(u,v)\, du\, dv \\ &= -\frac{2}{3}\epsilon^4 K + \int_{-\epsilon}^{\epsilon}\int_{-\epsilon}^{\epsilon} R(u,v)\, du\, dv. \end{aligned}$$

Since $|R(u,v)|$ is bounded by a constant times ϵ^3, its integral over D_ϵ is bounded by a constant times ϵ^5. The limit formula for K follows immediately. $\square$

7.6 Exercises

1 Let $U = \{(u, v) \mid v > 1\}$ and suppose $\sigma: U \to \mathbb{R}^3$ is a regular parametrized surface with $E = G = v^{-2}$ and $F = 0$.

a. Determine the Gauss curvature K, as a function of (u, v).

b. Compute the Christoffel symbols for σ.

c. Verify that the curve $\sigma \circ \mu$, where

$$\mu(s) = (a, e^s) \quad \text{or} \quad \mu(s) = (a + r \tanh s, r\frac{1}{\cosh s}),$$

has unit speed, and show that it is a geodesic. Here $a \in \mathbb{R}$ and $r > 0$ are constants, and s is assumed to belong in an interval for which $\mu(s) \in U$. Make a sketch of each curve μ in the (u, v)-plane, say with $a = r = 1$ (Hint: Notice that $\tanh^2 s + (\frac{1}{\cosh s})^2 = 1$).

d. Suppose in addition that the mentioned surface has coefficients $M = 0$ and $N = v^{-2}(v^2 - 1)^{\frac{1}{2}}$ in the second fundamental form. Determine L and the principal curvatures κ_1, κ_2.

2 Let $U = \mathbb{R}^2$, and let $\sigma: U \to \mathbb{R}^3$ be a regular parametrized surface for which $E = 1$, $F = 0$ and $G = 1 + u^2$ (see for example Exercise 3.10).

a. Determine the Christoffel symbols.

b. Show that $t \mapsto \sigma(t, v)$ is a geodesic for all v.

c. Find the geodesic curvature of the curve $t \mapsto \sigma(u, t)$ for $u \in \mathbb{R}$.

d. Verify that σ is a geodesic coordinate system, and determine the Gauss curvature by means of equation (6).

3 Let $\sigma: U \to \mathbb{R}^3$ be a geodesic coordinate system for which the Gaussian curvature is constant, $K = 0$. Show that $G = 1$ and that σ is isometric to a part of a plane (Hint: Conclude from (6) that $G = (au + b)^2$ where a and b are functions of v. Determine a and b from Theorem 7.4).

4 Let $\sigma: U = I \times J \to \mathbb{R}^3$ be a geodesic coordinate system transversal to the curve $\gamma(t) = \sigma(0, t)$. Assume that the Gaussian curvature is constant, $K = 1$. Show that $G = \cos^2 u$ and that σ is isometric to a part of the unit sphere (Hint: Conclude from (6) that $G = (a \cos u + b \sin u)^2$ where a and b are functions of v. Determine a and b from Theorem 7.4).

Appendices

Appendix A. Euclidean spaces

The set $\mathbb{R}^n$ is called *Euclidean n-space.* It is a vector space with the standard addition and scalar multiplication. In this appendix we recall some elementary notions for this space. The *dot product* of two vectors $v, w \in \mathbb{R}^n$ is the number defined by

$$v \cdot w = v_1 w_1 + \cdots + v_n w_n \in \mathbb{R}.$$

The *norm* of $v \in \mathbb{R}^n$ is given by

$$\|v\| = (v \cdot v)^{1/2} = (v_1^2 + \cdots + v_n^2)^{1/2},$$

and the *Euclidean distance* between $v, w \in \mathbb{R}^n$ is then defined as the norm $\|v - w\|$ of their difference. The *angle* between v and w is defined to be the number $\theta \in [0, \pi]$ for which

$$\cos\theta = \frac{v \cdot w}{\|v\| \, \|w\|} \tag{A.1}$$

provided the vectors are non-zero. It follows from the *Cauchy-Schwarz inequality*

$$|v \cdot w| \leq \|v\| \, \|w\|$$

that the right hand side of (A.1) belongs to $[-1, 1]$, so that the angle θ is well defined.

The vectors v and w are said to be *orthogonal* if $v \cdot w = 0$, or equivalently, if the angle between them is $\frac{\pi}{2}$, and they are said to be *orthonormal* if in addition they both have length 1. An *orthonormal basis* for $\mathbb{R}^n$ (or for a subspace) is a basis whose members are pairwise orthonormal, as for example the standard basis e_1, e_2, e_3 for $\mathbb{R}^3$.

For $r > 0$ and $p \in \mathbb{R}^n$ the set

$$B_r(p) = \{x \mid \|x - p\| < r\}$$

is called the *open ball* around p of radius r. A *neighborhood* of p is a set $U \subset \mathbb{R}^n$ which contains the open ball $B_r(p)$ for some $r > 0$. A set $U \subset \mathbb{R}^n$ is called *open* if it is a neighborhood of each of its points, that is, if *for every $p \in U$ there exists $r > 0$ such that all $x \in \mathbb{R}^n$ with $\|x - p\| < r$ belong to U.*

For instance, a set in $\mathbb{R}^2$ of the form $U =]a, b[\times]c, d[$, with open intervals $]a, b[$ and $]c, d[$, is open.

The *interior* of an arbitrary set $A \subset \mathbb{R}^n$ is the set of points $p \in A$ for which A is a neighborhood. This set is often denoted A°, and it is an open set. It is the largest open set contained in A. In particular, the interior of an open set is the set itself.

The *boundary* of $A \subset \mathbb{R}^n$ is the set of points $p \in \mathbb{R}^n$ (not necessarily in A) for which every open ball around p contains at least one point of A and at least one point of the complement $\mathbb{R}^n \setminus A$. It is often denoted ∂A. A set $A \subset \mathbb{R}^n$ is called *closed* if $\partial A \subset A$.

For example, the boundary of $U =]a, b[\times]c, d[$ consists of the four line segments that connect the corners of U.

A set $A \subset \mathbb{R}^n$ is called *bounded* if there exists $R > 0$ such that $\|x\| \leq R$ for all $x \in A$.

We recall that a function $f: A \to \mathbb{R}$, where $A \subset \mathbb{R}^n$, is called *continuous* if for each $p \in A$ and each $\epsilon > 0$ there exists $\delta > 0$ such that if $\|x - p\| < \delta$ then $|f(x) - f(p)| < \epsilon$. A function $f: A \to \mathbb{R}^m$ is continuous if the *components* $f_1, \ldots, f_m: A \to \mathbb{R}$ defined by $f(x) = (f_1(x), \ldots, f_m(x))$, are continuous.

Exercises

A.1 Determine the angle between $(1, 1, 1, 1)$ and $(1, 1, 1, 0)$ in $\mathbb{R}^4$.

A.2 Let $\gamma(t) = (3t, 3t^2, 2t^3)$. Show that the angle between the tangent vector of γ and the line given by $y = 0$, $z = x$, is a constant.

A.3 Verify that $u = (\frac{2}{3}, \frac{2}{3}, \frac{1}{3})$ and $v = (\frac{1}{3}, -\frac{2}{3}, \frac{2}{3})$ are orthonormal vectors. Find a third vector $w \in \mathbb{R}^3$, such that u, v, w is orthonormal basis. Determine the coordinates for $a = (1, 1, 1)$ with respect to this basis.

A.4 Prove that the set $\{(u, v) \in \mathbb{R}^2 \mid u, v > 0\}$ is open in $\mathbb{R}^2$, and that $\{(u, v) \in \mathbb{R}^2 \mid u, v \geq 0\}$ is not open.

A.5 Assume that $f: \mathbb{R}^n \to \mathbb{R}$ is continuous. Prove that $\{x \in \mathbb{R}^n \mid f(x) < c\}$ is an open set for every constant $c \in \mathbb{R}$.

A.6 Verify that the open ball $B_r(p)$ really is open.

Appendix B. Differentiable functions of several variables

Differentiability

Let $\Omega \subset \mathbb{R}^n$ be open, and let $f: \Omega \to \mathbb{R}$. A *partial derivative* of f is defined as the derivative of f with respect to one of the variables $x_1, \ldots, x_n$, the others being treated as constants. For example the first partial derivative $f'_{x_1} = \frac{\partial f}{\partial x_1}$ at $a \in \Omega$ is the derivative at a_1 of

$$t \mapsto f(t, a_2, \ldots, a_n).$$

The partial derivative at a is defined when this function of t is differentiable at a_1. If this is the case for all $i = 1, \dots, n$, we say that f *has partial derivatives at* a. If f has partial derivatives at all $a \in \Omega$, and if these partial derivatives are continuous functions of a, then we say that f is *continuously differentiable* or a C^1*-function* on Ω. The set of such functions on Ω is denoted $C^1(\Omega)$.

Let $F\colon \Omega \to \mathbb{R}^m$ be a vector function, and let $F_1, \dots, F_m\colon \Omega \to \mathbb{R}$ denote the components. The partial derivatives (if they exist) of these components functions are conveniently arranged in the *Jacobi* matrix

$$DF(a) = \begin{pmatrix} \frac{\partial F_1}{\partial x_1}(a) & \dots & \frac{\partial F_1}{\partial x_n}(a) \\ \vdots & & \vdots \\ \frac{\partial F_m}{\partial x_1}(a) & \dots & \frac{\partial F_m}{\partial x_n}(a) \end{pmatrix}.$$

Notice that DF is a map that associates a matrix to each point $a \in \Omega$. If $n = 1$ we identify the single column matrix $DF(a)$ with a vector in $\mathbb{R}^m$. The vector function $DF\colon \mathbb{R} \to \mathbb{R}^m$ is in this case denoted F' and called the *derivative* of F. The Jacobi matrix is the analogue for functions of several variables of this derivative.

Example B.1 A *linear* map $L\colon \mathbb{R}^n \to \mathbb{R}^m$ is represented by an $m \times n$ matrix $A = (a_{ij})$ as follows:

$$L(x) = Ax = \begin{pmatrix} a_{11} & \dots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \dots & a_{mn} \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} a_{11}x_1 + \cdots + a_{1n}x_n \\ \vdots \\ a_{m1}x_1 + \cdots + a_{mn}x_n \end{pmatrix}.$$

It is easily seen that the Jacobian of this map is exactly the matrix A, that is, $DL(x) = A$ for all $x \in \mathbb{R}^n$.

We call F continuously differentiable if each coordinate function F_j is continuously differentiable, or in other words, if each entry in the Jacobi matrix exists and depends continuously on a. Recall the following fundamental theorem, which asserts that $x \mapsto F(a) + DF(a)(x - a)$ approximates F near a.

Theorem B.1. *Let $F\colon \Omega \to \mathbb{R}^m$ be continuously differentiable and let $a \in \Omega$. Then*

$$\frac{\|F(x) - [F(a) + DF(a)(x - a)]\|}{\|x - a\|} \to 0 \quad \text{for } x \to a, \tag{B.1}$$

that is, the vector difference $F(x) - [F(a) + DF(a)(x - a)]$ tends to 0 even after division by $\|x - a\|$.

A function F which satisfies (B.1) is called *differentiable* at a, and the theorem simply asserts that 'continuously differentiable' implies 'differentiable'.

Composition

The differentiation of composed maps is governed by the *chain rule*. For functions of one variable it is the well known rule

$$(g \circ f)'(a) = g'(f(a))f'(a),$$

and for functions of several variables it takes the following form.

Theorem B.2 (Chain rule). *Let $\Omega \subset \mathbb{R}^n$ and $\Omega' \subset \mathbb{R}^m$ be open, and let*

$$F\colon \Omega \to \Omega' \quad \text{and} \quad G\colon \Omega' \to \mathbb{R}^l,$$

be continuously differentiable. Then

$$G \circ F\colon \Omega \to \mathbb{R}^l$$

is continuously differentiable and has the Jacobi matrix

$$D(G \circ F)(a) = DG(F(a))\, DF(a)$$

for all $a \in \Omega$, where the product on the right is given by ordinary matrix multiplication.

In particular, if $n = 1$ we can write the chain rule in the following form

$$(G \circ F)'(a) = DG(F(a))\, F'(a). \tag{B.2}$$

Example B.2 Let $F\colon \mathbb{R} \to \mathbb{R}^2$ be given by $F(t) = (t^2, t+1)$ and let $G\colon \mathbb{R}^2 \to \mathbb{R}$ be given by $G(y_1, y_2) = y_1 y_2^2 - y_1^2$. Then

$$F'(t) = \begin{pmatrix} 2t \\ 1 \end{pmatrix}, \quad \text{and} \quad DG(y) = \begin{pmatrix} y_2^2 - 2y_1 & 2y_1y_2 \end{pmatrix}.$$

Hence $G \circ F\colon \mathbb{R} \to \mathbb{R}$ has the derivative

$$\begin{aligned}(G \circ F)'(t) &= DG(F(t))F'(t) \\ &= \begin{pmatrix} (t+1)^2 - 2t^2 & 2t^2(t+1) \end{pmatrix} \begin{pmatrix} 2t \\ 1 \end{pmatrix} \\ &= (-t^2 + 2t + 1)2t + (2t^3 + 2t^2)1 = 6t^2 + 2t.\end{aligned}$$

Notice that we could also first have determined the expression $G \circ F(t) = t^2(t+1)^2 - t^4 = 2t^3 + t^2$ and then differentiated $(G \circ F)' = 6t^2 + 2t$. For the purpose of computing $(G \circ F)'$, this would clearly be much faster. The importance of the chain rule is more theoretical, it gives a general expression for the derivative.

Symmetry of mixed partials

A function $f\colon \Omega \subset \mathbb{R}^n \to \mathbb{R}$ is called a C^2-function if it is C^1 and all the first order partial derivatives are also C^1-functions. The partial derivatives of the partial derivatives, that is, the functions

$$\frac{\partial^2 f}{\partial x_i \partial x_j} = \frac{\partial \frac{\partial f}{\partial x_j}}{\partial x_i}$$

are called *higher* or *mixed* partial derivatives.

Theorem B.3. *Let $f\colon \Omega \to \mathbb{R}$ be C^2. Then*

$$\frac{\partial^2 f}{\partial x_i \partial x_j} = \frac{\partial^2 f}{\partial x_j \partial x_i}$$

for all i and j.

Similarly, we can consider derivatives of order higher than 2. If f has partial derivatives up to order k, and if these are continuous, then f is called a C^k-function. From the theorem above we can derive similar statements about symmetry of these higher derivatives, for example

$$\frac{\partial^3 f}{\partial x_1^2 \partial x_2} = \frac{\partial^3 f}{\partial x_1 \partial x_2 \partial x_1} = \frac{\partial^3 f}{\partial x_2 \partial x_1^2}$$

when f is a C^3-function. In short, the conclusion is that differentiations with respect to $x_1, \dots, x_n$ commute with each other (when applied to functions which are continuously differentiable up to sufficient order).

A function which is C^k for all k is called C^∞ or *smooth*. The set of such functions on Ω is denoted $C^\infty(\Omega)$. This is the class of functions that is mainly used in differential geometry.

Taylor's theorem

Taylor's theorem allows us to approximate a smooth function by a polynomial of any given order, in the vicinity of a given point.

For a smooth function $f\colon \Omega \to \mathbb{R}$, where $\Omega \subset \mathbb{R}^2$, it reads to the order one

$$f(u_0 + h, v_0 + k) \simeq f(u_0, v_0) + f'_u(u_0, v_0)h + f'_v(u_0, v_0)k$$

and to the order two

$$\begin{aligned} f(u_0 + h, v_0 + k) \simeq f(u_0, v_0) &+ f'_u(u_0, v_0)h + f'_v(u_0, v_0)k \\ &+ \tfrac{1}{2} f''_{uu}(u_0, v_0)h^2 + f''_{uv}(u_0, v_0)hk + \tfrac{1}{2} f''_{vv}(u_0, v_0)k^2. \end{aligned}$$

These statements are qualitative, because the 'approximation' $\simeq$ is not a well defined relation.

There are more precise versions, where the remainder, which by definition is the difference between the two sides of $\simeq$, is estimated. To the order one

$$f(u_0 + h, v_0 + k) = f(u_0, v_0) + f'_u(u_0, v_0)h + f'_v(u_0, v_0)k + R_1(h, k),$$

and the estimate, which is valid for a C^2-function $f\colon \Omega \to \mathbb{R}$, is as follows. For a given point $(u_0, v_0) \in \Omega$ there exist constants $\epsilon > 0$ and $C > 0$ such that

$$|R_1(h, k)| \leq C\|(h, k)\|^2$$

for all $(h,k) \in \mathbb{R}^2$ with $\|(h,k)\| < \epsilon$.

Likewise, to the order two,

$$\begin{aligned} f(u_0+h, v_0+k) &= f(u_0,v_0) + f'_u(u_0,v_0)h + f'_v(u_0,v_0)k \\ &\quad + \tfrac{1}{2}f''_{uu}(u_0,v_0)h^2 + f''_{uv}(u_0,v_0)hk + \tfrac{1}{2}f''_{vv}(u_0,v_0)k^2 + R_2(h,k). \end{aligned}$$

with the following estimate valid for a C^3-function f. For a given point $(u_0,v_0) \in \Omega$ there exist constants $\epsilon > 0$ and $C > 0$ such that

$$|R_2(h,k)| \le C\|(h,k)\|^3$$

for all $(h,k) \in \mathbb{R}^2$ with $\|(h,k)\| < \epsilon$.

Exercises

B.1 Find f'_u and f'_v for each of the functions $f\colon \mathbb{R}^2 \to \mathbb{R}$:
1) $f(u,v) = u^2 + v^2 + 3uv - u - 4v$,
2) $f(u,v) = e^{2u-v+1}$

B.2 Determine the Jacobi matrix at $(1,1)$ for $f\colon \{(u,v) \mid u,v > 0\} \to \mathbb{R}^2$, given by $f(u,v) = (u^2v, 2\sqrt{uv})$.

B.3 Let f be a differentiable map $\mathbb{R}^3 \to \mathbb{R}$. Determine the derivative of $t \mapsto f(t,t^2,e^t)$.

B.4 Let $g\colon \mathbb{R}^2 \to \mathbb{R}$ be given by $g(x,y) = xy$, and let $\varphi\colon \mathbb{R}^2 \to \mathbb{R}^2$ be given by $\varphi(x,y) = (x+y, x-y)$. Determine the Jacobi matrices for g, φ, and for the inverse map $f(u,v) = \varphi^{-1}(u,v)$. Determine the Jacobi matrix for $g \circ f$ in each of the following two ways:
1) By using the chain rule.
2) Through explicit computation of $g \circ f(u,v)$.

B.5 Let φ be an arbitrary differentiable function $\mathbb{R} \to \mathbb{R}$, and let $F(x,y) = xy - \varphi(y/x)$ for $(x,y) \in \mathbb{R}^2$ with $x \neq 0$. Show that $x\frac{\partial F}{\partial x} + y\frac{\partial F}{\partial y} = 2xy$.

B.6 Prove the rule
$$(f \cdot g)' = f' \cdot g + f \cdot g'$$
for $f, g\colon \mathbb{R} \to \mathbb{R}^n$ (see page 62).

B.7 Let $\gamma\colon I \to \mathbb{R}^n$ be smooth with $\gamma(t) \neq 0$ for all t. Show that $t \mapsto \|\gamma(t)\|$ is differentiable and has the derivative
$$\frac{\gamma'(t) \cdot \gamma(t)}{\|\gamma(t)\|}.$$

Appendix C. Normal vectors and cross products

In this appendix the construction of normal vectors in $\mathbb{R}^2$ and cross products in $\mathbb{R}^3$ is briefly presented. These notions appear naturally in many geometrical constructions. For example the geometry of planes in $\mathbb{R}^3$ is often expressed by means of cross products. Cross products also play a prominent role in mechanics and electromagnetic theory.

The common background for the definitions in this appendix for $\mathbb{R}^2$ and $\mathbb{R}^3$ is a choice of *orientation*, which we will first explain generally for $\mathbb{R}^n$. For an ordered set of n vectors $a_1, a_2, \dots, a_n$ in $\mathbb{R}^n$ we denote by $[a_1 a_2 \dots a_n]$ the $n \times n$ matrix which has a_1, a_2 etc as its columns (in that order). We divide the bases for $\mathbb{R}^n$ in two classes, depending on whether the determinant of $[a_1 a_2 \dots a_n]$ is positive or negative (the determinant is non-zero since the vectors are linearly independent). An orientation of $\mathbb{R}^n$ is a choice of one of the two classes. The standard choice is the class of bases which have positive determinant. Such a basis is then called positively ordered.

Having made this standard choice we thus say that two basis vectors a and b in $\mathbb{R}^2$ are *positively ordered* if $\det[ab] > 0$ and we say that three basis vectors a, b and c in $\mathbb{R}^3$ are positively ordered if $\det[abc] > 0$. For example, the standard basis vectors $e_1 = (1, 0)$ and $e_2 = (0, 1)$ for $\mathbb{R}^2$ and $e_1 = (1, 0, 0)$, $e_2 = (0, 1, 0)$ and $e_3 = (0, 0, 1)$ for $\mathbb{R}^3$ are positively ordered.

In $\mathbb{R}^2$ this choice of orientation means that a, b is a positively ordered basis if and only if the direction of b can be reached from the direction of a by a counter clockwise rotation of an angle between 0 and π, and in $\mathbb{R}^3$ it means that a, b, c is positively ordered if and only if the vectors form a right-handed triple.

Let $a = (a_1, a_2) \in \mathbb{R}^2$. We define the *normal vector* by $\hat{a} = (-a_2, a_1)$. It is the vector obtained by rotating a the angle $\frac{\pi}{2}$ in counter clockwise direction (which is the positive direction according to our chosen orientation). Notice that

$$\det[ab] = a_1 b_2 - a_2 b_1 = \hat{a} \cdot b \tag{C.1}$$

for all $b \in \mathbb{R}^2$, where the dot denotes the standard dot product (see Appendix A). The map $a \mapsto \hat{a}$ is linear.

The construction of $\hat{a}$ is particular for $\mathbb{R}^2$. In $\mathbb{R}^3$ there is no way to distinguish a normal vector to a given vector, since there are infinitely many normal vectors. Instead, the analog of the construction points out a normal vector to *two* given vectors.

For two vectors $a = (a_1, a_2, a_3)$ and $b = (b_1, b_2, b_3)$ in $\mathbb{R}^3$ we define the *cross product*, which is again a vector in $\mathbb{R}^3$, by

$$a \times b = \left(\begin{vmatrix} a_2 & b_2 \\ a_3 & b_3 \end{vmatrix}, - \begin{vmatrix} a_1 & b_1 \\ a_3 & b_3 \end{vmatrix}, \begin{vmatrix} a_1 & b_1 \\ a_2 & b_2 \end{vmatrix} \right).$$

It follows from the determinant identity

$$\begin{vmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{vmatrix} = c_1 \begin{vmatrix} a_2 & b_2 \\ a_3 & b_3 \end{vmatrix} - c_2 \begin{vmatrix} a_1 & b_1 \\ a_3 & b_3 \end{vmatrix} + c_3 \begin{vmatrix} a_1 & b_1 \\ a_2 & b_2 \end{vmatrix}$$

that $a \times b$ is uniquely characterized by the property that

$$\det[abc] = (a \times b) \cdot c \tag{C.2}$$

for all $c \in \mathbb{R}^3$. Equation (C.2) is the 3-dimensional analogue of (C.1), the expression on the right is called the triple product of a, b and c. It follows from this characterization that $a \times b$ depends linearly on both a and b (since the determinant depends linearly on each of its columns).

The following properties of $a \times b$ are easily verified

(i) $a \times b$ is perpendicular to both a and b
(ii) $b \times a = -a \times b$
(iii) $\|a \times b\|^2 = \|a\|^2 \|b\|^2 - (a \cdot b)^2$
(iv) $\|a \times b\|$ is the area of the parallelogram spanned by a and b
(v) a and b are linearly independent if and only if $a \times b \neq 0$
(vi) if $a \times b \neq 0$ then a, b, and $a \times b$ is a right handed triple.

The last property in this list reflects our chosen orientation of $\mathbb{R}^3$. Notice that properties (i), (iv) and (vi) together determine $a \times b$ uniquely in geometric terms.

Exercises

C.1 Prove that if three vectors $u, v, w \in \mathbb{R}^3$ satisfy $u + v + w = 0$, then $u \times v = v \times w = w \times u$.

C.2 Let A be an orthogonal 3×3-matrix with $\det A = 1$. Prove

$$A(w_1 \times w_2) = Aw_1 \times Aw_2$$

for all $w_1, w_2 \in \mathbb{R}^3$ (Hint: Use the characterization by (C.2) of the cross product, together with the equation (D.1)).

Appendix D. Diagonalization of symmetric matrices

Let A be an $n \times n$ matrix. By definition, a *diagonalization* of A is accomplished by an invertible matrix C, if the matrix $D = C^{-1}AC$ is diagonal (that is, all entries outside the diagonal are 0). Diagonalization plays a very important role in linear algebra, basically because it is a means to simplify expressions involving A. It is not possible to diagonalize all matrices A, certain conditions have to be imposed. The main result of this appendix, which is stated in the theorem below, gives one such condition (but not the most general one).

The theory of diagonalization is closely connected with the theory of eigenvectors and eigenvalues. Recall that an *eigenvalue* for the matrix A is a number λ for which there exists a non-zero vector $w \in \mathbb{R}^n$ such that

$$Aw = \lambda w.$$

The vector w is called a corresponding *eigenvector*.

It is a fact known from linear algebra that a number λ is an eigenvalue if and only if it is a root in the *characteristic polynomial*

$$\lambda \mapsto \det(A - \lambda I).$$

The corresponding eigenvectors are the nonzero solutions w to $(A - \lambda I)w = 0$.

Recall that the matrix A is called *symmetric* if it equals its transposed matrix A^t, that is, if its elements satisfy $a_{ij} = a_{ji}$ for all i, j. Recall also that an *orthogonal matrix* is a square matrix C with real entries for which the transposed matrix C^t is inverse to C, that is, $C^t C = I$. Equivalently, it is a matrix whose columns form an orthonormal set.

Theorem D.1. *Let A be a symmetric $n \times n$ matrix. There exists an orthogonal $n \times n$ matrix C such that $D = C^{-1}AC$ is a diagonal matrix with these roots as its entries. The columns of C are eigenvectors for A, and the eigenvalues are the diagonal elements of D (in the corresponding order).*

Proof. For simplicity we shall assume $n = 2$ in the proof. This is not a serious restriction for the present notes, where all the applications have $n = 2$. In order to pinpoint where the assumption is critically used, we will keep n arbitrary in the beginning of the proof.

We regard vectors in $\mathbb{R}^n$ as matrices with a single column, and note that the dot product $v \cdot w = v_1 w_1 + \dots v_n w_n$ can be written by means of matrix multiplication as $v \cdot w = v^t w$. For any $n \times n$ matrix A we have

$$Av \cdot w = (Av)^t w = v^t A^t w = v \cdot A^t w, \tag{D.1}$$

and if A is symmetric we thus obtain

$$Av \cdot w = v \cdot Aw \tag{D.2}$$

for all $v, w \in \mathbb{R}^n$.

We first prove that eigenvectors corresponding to different eigenvalues are orthogonal, that is, if $v, w \in \mathbb{R}^n \setminus \{0\}$ and $Av = \lambda v, Aw = \mu w$ with $\lambda \neq \mu$, then $v \cdot w = 0$. By (D.2) we have

$$\lambda v \cdot w = (Av) \cdot w = v \cdot (Aw) = \mu v \cdot w,$$

and since $\lambda \neq \mu$ we must indeed have $v \cdot w = 0$.

We next observe that if $w_1, \dots, w_n \in \mathbb{R}^n$ is an orthonormal set of eigenvectors for A, then the matrix $C = [w_1 \dots w_n]$ with these columns is orthogonal. Since the columns of the matrix product AC are obtained by multiplication of A with the columns of C, we see that

$$AC = [Aw_1 \dots Aw_n] = [\lambda_1 w_1 \dots \lambda_n w_n] = CD,$$

where D is the diagonal matrix with the eigenvalues $\lambda_1, \dots, \lambda_n$ as entries. It follows that $C^{-1}AC = D$, so that C diagonalizes A.

It remains to be shown that there exist such an orthonormal set of eigenvectors. For this last step we assume $n = 2$. Let

$$A = \begin{pmatrix} a & b \\ b & d \end{pmatrix}$$

We may assume $b \neq 0$, since otherwise A is already diagonal and we can take $w_1 = e_1, w_2 = e_2$. The characteristic polynomial is $\lambda^2 - (a+d)\lambda + ad - b^2$, and its roots are given by

$$\lambda = \frac{a + d \pm \sqrt{(a+d)^2 - 4(ad - b^2)}}{2} = \frac{a + d \pm \sqrt{(a-d)^2 + 4b^2}}{2}.$$

Since $b \neq 0$ the expression inside the square root is positive, and thus there are two different real roots λ_1, λ_2. As mentioned above each root in the characteristic polynomial is an eigenvalue, let $w_1, w_2 \in \mathbb{R}^2$ be normalised eigenvectors. Then $w_1 \perp w_2$ by what was shown above, and hence they form an orthonormal set. $\square$

Example D.1 The eigenvalues of the symmetric matrix

$$\begin{pmatrix} 1 & \frac{1}{2} \\ \frac{1}{2} & 1 \end{pmatrix}$$

are determined as the roots $\frac{1}{2}$ and $\frac{3}{2}$ of

$$\det \begin{pmatrix} 1-\lambda & \frac{1}{2} \\ \frac{1}{2} & 1-\lambda \end{pmatrix} = \lambda^2 - 2\lambda + \tfrac{3}{4}.$$

An eigenvector corresponding to $\lambda = \frac{1}{2}$ is determined by solution of

$$\begin{pmatrix} 1 & \frac{1}{2} \\ \frac{1}{2} & 1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \tfrac{1}{2} \begin{pmatrix} x \\ y \end{pmatrix}$$

which leads to, for example

$$\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 1 \\ -1 \end{pmatrix}.$$

An eigenvector for $\lambda = \frac{3}{2}$ is then found as

$$\begin{pmatrix} x \\ y \end{pmatrix} = \widehat{\begin{pmatrix} 1 \\ -1 \end{pmatrix}} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$

The corresponding matrix C has normalisations of these eigenvectors as its columns

$$C = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix},$$

and the diagonalized matrix is

$$D = \begin{pmatrix} \frac{1}{2} & 0 \\ 0 & \frac{3}{2} \end{pmatrix}.$$

It is useful to express the preceding theorem as a result about linear maps, rather than matrices. This is done in the following corollary.

Corollary D.1. *Let $U \subset \mathbb{R}^m$ be a linear subspace, and let $L: U \to U$ be a linear map which is symmetric, that is*

$$L(u_1) \cdot u_2 = u_1 \cdot L(u_2), \qquad u_1, u_2 \in U. \tag{D.3}$$

Then there exists an orthonormal basis for U consisting of eigenvectors for L.

Proof. Let $\eta_1, \ldots, \eta_n$ be an arbitrary orthonormal basis for U, and let A denote the $n \times n$ matrix which represents L with respect to this basis, that is,

$$L\eta_j = \sum_{i=1}^{n} a_{ij}\eta_i, \qquad (j = 1, \ldots, n). \tag{D.4}$$

The proof is based on the following two observations, which are well known from linear algebra. A vector $u = x_1\eta_1 + \cdots + x_n\eta_n \in U$ is an eigenvector for L with eigenvalue λ,

$$Lu = \lambda u,$$

if and only if the column of its coordinates

$$x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \in \mathbb{R}^n$$

is an eigenvector for A,

$$Ax = \lambda x$$

with the same eigenvalue. The second observation is that the dot product of two vectors $u, v \in U$ is computed as the dot product of their coordinates:

$$(x_1\eta_1 + \cdots + x_n\eta_n) \cdot (y_1\eta_1 + \cdots + y_n\eta_n) = x_1y_1 + \cdots + x_ny_n.$$

Since the basis is orthonormal, it follows from (D.4) that

$$a_{ij} = L(\eta_j) \cdot \eta_i,$$

and hence it follows from the symmetry of L, that A is a symmetric matrix. By Theorem D.1 there exists an orthogonal $n \times n$ matrix C whose columns are eigenvectors for A. Let $v_1, \dots, v_n$ be the vectors in U whose coordinates with respect to $\eta_1, \dots, \eta_n$ are the columns of C, that is

$$v_i = c_{1i}\eta_1 + \cdots + c_{ni}\eta_n \in U, \qquad (i = 1, \dots, n).$$

The first observation made above implies that $v_1, \dots, v_n$ are eigenvectors for L, and the second observation implies that they form an orthonormal set. Since the dimension of U is n, they form a basis for U. $\square$

Exercises

D.1 Let $w = (3, 4) \in \mathbb{R}^2$, and let

$$A = \begin{pmatrix} a & 12 \\ 12 & a+7 \end{pmatrix}.$$

Show that w and $\hat{w}$ are eigenvectors for A, and determine their eigenvalues.

D.2 Let

$$A = \begin{pmatrix} 2 & 2 \\ 2 & 5 \end{pmatrix}$$

Determine an orthogonal matrix C and a diagonal matrix D, such that $D = C^{-1}AC$. How many pairs of 2×2-matrices (C, D) satisfy these requirements?

D.3 Let A be an $n \times n$-matrix. Show that if there exists an orthogonal matrix diagonalizing A, then A is symmetric.

Appendix E. Hyperbolic functions

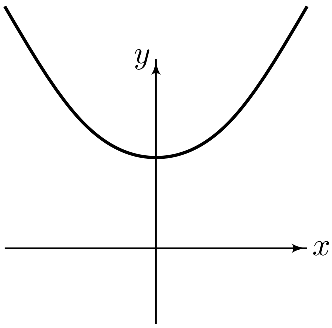
The *hyperbolic cosine* and *hyperbolic sine* functions are defined by

$$\cosh t = \frac{e^t + e^{-t}}{2}, \quad \sinh t = \frac{e^t - e^{-t}}{2}, \tag{E.1}$$

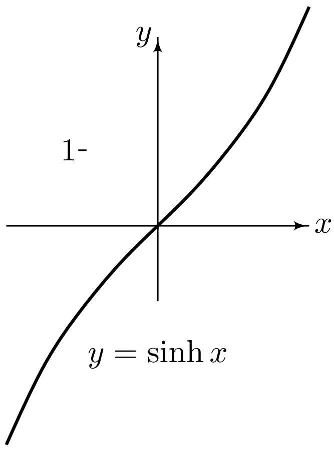
for $t \in \mathbb{R}$. The equation

$$\cosh^2 t - \sinh^2 t = 1,$$

which resembles the well-known $\cos^2 t + \sin^2 t = 1$, is easily derived from (E.1). It follows that the point $(\cosh t, \sinh t)$ lies on a hyperbola (see Example 1.1.3), and because of this the functions are viewed as hyperbolic counterparts to the trigonometric functions cos and sin. The graphs of the two functions are shown below.



$y = \cosh x$



$y = \sinh x$

Notice that cosh is even and sinh is odd. In analogy with the definitions of the tangent and cotangent one defines

$$\tanh x = \frac{\sinh x}{\cosh x}, \qquad \coth x = \frac{\cosh x}{\sinh x}.$$

It is easily seen that the derivatives of these functions are

$$\begin{aligned} &\frac{d}{dx}\cosh x = \sinh x, && \frac{d}{dx}\sinh x = \cosh x, \\ &\frac{d}{dx}\tanh x = \frac{1}{\cosh^2 x}, && \frac{d}{dx}\coth x = -\frac{1}{\sinh^2 x}. \end{aligned}$$

Notational Index

Index