# PMATH 365 Course Notes

Differential Geometry

Ruxandra Moraru • Winter 2023 • University of Waterloo

## Table of Contents

# 1 Submanifolds of $\mathbb{R}^n$

## 1.1 Preliminaries

To begin, we'll recall some facts about the topology of $\mathbb{R}^n$ and vector-valued functions.

In this course, we'll be working with the metric topology with respect to the Euclidean norm (or metric). Let $x = (x_1, \ldots, x_n)$, $y = (y_1, \ldots, y_n) \in \mathbb{R}^n$. The **Euclidean norm** is defined to be

$$\|x\| = \sqrt{x_1^2 + \cdots + x_n^2},$$

and **Euclidean distance** is given by

$$\operatorname{dist}(x, y) = \|y - x\| = \sqrt{(y_1 - x_1)^2 + \cdots + (y_n - x_n)^2}.$$

We define the **open ball** of radius $r > 0$ centered at $x \in \mathbb{R}^n$ by

$$B_r(x) := \{y \in \mathbb{R}^n : \operatorname{dist}(x, y) < r\} \subset \mathbb{R}^n.$$

A **topology** on $\mathbb{R}^n$ is a collection $\mathcal{U} = \{U_\alpha\}_{\alpha \in A}$ of subsets $U_\alpha \subset \mathbb{R}^n$ that satisfy the following properties.

(i) $\varnothing$ and $\mathbb{R}^n$ are in $\mathcal{U}$.

(ii) For any subcollection $\mathcal{V} = \{U_\beta\}_{\beta \in B}$ with $U_\beta \in \mathcal{U}$ for all $\beta \in B$, we have $\bigcup_{\beta \in B} U_\beta \in \mathcal{U}$.

(iii) For any *finite* subcollection $\{U_{\alpha_1}, \ldots, U_{\alpha_m}\} \subset \mathcal{U}$, we have $\bigcap_{i=1}^m U_{\alpha_i} \in \mathcal{U}$.

The sets $U_\alpha \in \mathcal{U}$ are called the **open sets** of the topology; their complements $F_\alpha = \mathbb{R}^n \setminus U_\alpha$ are called the **closed sets**.

Note that the sets $\varnothing$ and $\mathbb{R}^n$ are both open and closed. Moreover, the notion of a topology can be extended to more general sets $X$, not just $\mathbb{R}^n$. A topology can also be defined starting with closed sets, but we prefer to work with open sets because many nice properties, such as differentiability, are better described with them.

Under the metric topology, we say that a set $A \subset \mathbb{R}^n$ is **open** if $A = \varnothing$ or if for all $p \in A$, there exists $r > 0$ such that $B_r(p) \subset A$. Moreover, $A$ is **closed** if its complement $A^c = \mathbb{R}^n \setminus A$ is open. (We leave it as an exercise to show that this is indeed a topology.)

For example, the open balls $B_r(x)$ are open sets for all $x \in \mathbb{R}^n$ and $r > 0$. Indeed, for any point $p \in B_r(x)$, one sees that by picking $r' = (r - \|p - x\|)/2$, we have $B_{r'}(p) \subset B_r(x)$.

In general, open sets are described with strict inequalities, while closed sets are described using equality or inclusive inequalities. However, note that most sets are neither open nor closed, such as the half-open interval $U = (-1, 1]$ over $\mathbb{R}$.

The metric topology is not the only topology on $\mathbb{R}^n$; one example is the one consisting of only the sets $\mathcal{U} = \{\varnothing, \mathbb{R}^n\}$. However, we generally want more open sets to work with since we might want to know the behaviour of functions around a point $p \in \mathbb{R}^n$. If the only non-empty open set we had was $\mathbb{R}^n$, then this would apply to all points in $\mathbb{R}^n$, which does not yield a lot of information.

Let $p \in \mathbb{R}^n$. The previous paragraph leads us to the definition of an **open neighbourhood** of $p$, which is just an open set $U \subset \mathbb{R}^n$ such that $p \in U$.

We now turn our discussion to vector-valued functions. Let $U \subset \mathbb{R}^n$ and consider the vector-valued function

$$\begin{aligned} F : U \subset \mathbb{R}^n &\to B \subset \mathbb{R}^m \\ x = (x_1, \ldots, x_n) &\mapsto (F_1(x), \ldots, F_m(x)). \end{aligned}$$

Then $F$ is continuous if and only if the component functions $F_i : U \to \mathbb{R}$ are continuous for all $i = 1, \ldots, m$.

We say that $F$ is a **homeomorphism** if it is a continuous bijection whose inverse

$$F^{-1} : B \subset \mathbb{R}^m \to U \subset \mathbb{R}^n$$

is also continuous. For example, the identity map $\mathrm{Id}_{\mathbb{R}^n} : \mathbb{R}^n \to \mathbb{R}^n$ and the function $f : \mathbb{R} \to \mathbb{R}$ defined by $f(x) = x^3$ are both homeomorphisms.

It is a known fact that homeomorphisms map open sets to open sets and closed sets to closed sets. This follows from the topological characterization of continuity, which states that $F$ is continuous if and only if for every open (respectively closed) set $V \subset \mathbb{R}^m$, we have that $F^{-1}(V)$ is open (respectively closed). In fact, homeomorphisms preserve much more structure than this, as we'll see later.

## 1.2 Topological submanifolds of $\mathbb{R}^n$

We now define the main object we'll be working with in this course.

DEFINITION 1.1

A **$k$-dimensional topological submanifold** (or **topological $k$-submanifold**) of $\mathbb{R}^n$ is a subset $M \subset \mathbb{R}^n$ such that for every $p \in M$, there exists an open neighbourhood $V$ of $p$ in $\mathbb{R}^n$, an open set $U \subset \mathbb{R}^k$, and a homeomorphism

$$\alpha : U \subset \mathbb{R}^k \to V \cap M \subset \mathbb{R}^n.$$

The homeomorphism $\alpha$ is called a **coordinate chart** (or **patch**) on $M$.

Note that the open neighbourhood $V \subset \mathbb{R}^n$ of $p$, the open set $U \subset \mathbb{R}^k$, and the map $\alpha$ do not need to be unique. But we'll see later that the dimension $k$ must be unique and is completely determined by $M$.

For example, $\mathbb{R}^n$ is a topological $n$-submanifold of $\mathbb{R}^n$ by taking $U = V = \mathbb{R}^n$ and $\alpha = \mathrm{Id}_{\mathbb{R}^n}$. Any open set $W \subset \mathbb{R}^n$ is a topological $n$-submanifold of $\mathbb{R}^n$ by taking $U = V = W$ and $\alpha = \mathrm{Id}_W$.

Let's now consider some non-trivial examples. Consider

$$M = \{(x, y) \in \mathbb{R}^2 : y = x^2\} \subset \mathbb{R}^2,$$

which is the graph of the parabola $f(x) = x^2$. Then $M$ is a topological 1-submanifold of $\mathbb{R}^2$ by considering the map $\alpha : \mathbb{R}^1 \to M \subset \mathbb{R}^2$, $t \mapsto (t, t^2)$. The inverse $\alpha^{-1} : M \to \mathbb{R}^1$ is just the projection of the first coordinate, which is continuous.

More generally, let $U \subset \mathbb{R}^k$ be an open set. Consider the graph of a continuous function

$$F : U \subset \mathbb{R}^k \to \mathbb{R}^{n-k},\ x \mapsto (F_1(x), \ldots, F_{n-k}(x)).$$

In other words, we are looking at the set

$$G = \{(x, y) \in \mathbb{R}^k \times \mathbb{R}^{n-k} : y = F(x),\ x \in U\} \subset \mathbb{R}^n.$$

We claim that $G$ is a $k$-dimensional topological submanifold of $\mathbb{R}^n$. To see this, define $\alpha : U \subset \mathbb{R}^k \to G \subset \mathbb{R}^n$ by $x \mapsto (x, F(x))$. Then $\alpha$ is continuous since $F$ is continuous, and it is a bijection since we are restricted to $G$. Moreover, it has continuous inverse $\alpha^{-1} : G \subset \mathbb{R}^n \to U \subset \mathbb{R}^k$, $(x, y) \mapsto x$.

Here are two more examples of this in action.

(1) Let $M = \{(x, y, z) \in \mathbb{R}^3 : z = x^2 + y^2\} \subset \mathbb{R}^3$. Then $M$ is the graph of the continuous function $f(x, y) = x^2 + y^2$, so it is a 2-dimensional topological submanifold of $\mathbb{R}^3$.

(2) Observe that $M = \{(x, y, z) \in \mathbb{R}^3 : y = x^2,\ z = x^3\} \subset \mathbb{R}^3$ is the graph of the continuous function $F(t) = (t^2, t^3)$, so it is a 1-dimensional topological submanifold of $\mathbb{R}^3$.

In all the examples above, we only needed one coordinate chart which worked for all points. However, this is not always the case! Consider the unit circle

$$\mathbb{S}^1 := \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 = 1\} \subset \mathbb{R}^2.$$

Note that $\mathbb{S}^1$ is compact. Therefore, by Heine-Borel, it is closed and bounded. Recall that homeomorphisms preserve closed sets, so it is impossible to find a unique chart $\alpha$. Indeed, if we had such a homeomorphism $\alpha : U \subset \mathbb{R}^1 \to \mathbb{S}^1 \subset \mathbb{R}^2$ for some open set $U$, then $U = \alpha^{-1}(\mathbb{S}^1)$ would be a compact subset of $\mathbb{R}^1$. But the only open and compact subset of $\mathbb{R}^n$ is $\varnothing$, which is a contradiction!

Nonetheless, two coordinate charts are enough to cover all points on $\mathbb{S}^1$. Define

$$V_1 = \mathbb{R}^2 \setminus \{(x, 0) \in \mathbb{R}^2 : x \le 0\},$$
$$V_2 = \mathbb{R}^2 \setminus \{(x, 0) \in \mathbb{R}^2 : x \ge 0\},$$

which are both open sets. Then the homeomorphism

$$\alpha_1 : U_1 = (-\pi, \pi) \to \mathbb{S}^1 \cap V_1,\ t \mapsto (\cos t, \sin t)$$

covers all points on $\mathbb{S}^1$ except for $(-1, 0)$, while

$$\alpha_2 : U_2 = (0, 2\pi) \to \mathbb{S}^1 \cap V_2,\ t \mapsto (\cos t, \sin t)$$

covers all points on $\mathbb{S}^1$ except for $(1, 0)$.

## 1.3 More preliminaries

We now introduce another definition from topology.

**Definition 1.2**

Let $A \subset \mathbb{R}^n$. A subset $U \subset A$ is **relatively open** if it is of the form $U = A \cap U'$ for some open set $U' \subset \mathbb{R}^n$. Similarly, we say that $F \subset A$ is **relatively closed** if $F = A \cap F'$ for some closed set $F' \subset \mathbb{R}^n$.

For example, consider $\mathbb{R}$ equipped with the metric topology so that the open (respectively closed) sets are the unions of open intervals (respectively finite intersections of closed intervals). Let $A = [-1, 2) \subset \mathbb{R}$, which is neither open nor closed in $\mathbb{R}$. Take $U = [-1, 1) \subset \mathbb{R}$, which is again neither open nor closed in $\mathbb{R}$. But $U$ is relatively open in $A$ since $U = A \cap (-3, 1)$. Similarly, $F = [-1, 1] = A \cap [-1, 1]$ is relatively closed in $A$.

Using the language of relatively open and closed sets, a lot of statements can be made simpler.

(1) We define an **open neighbourhood of $p$ in $A$** to be a relatively open set $U$ containing $p$.

(2) The relatively open sets form a topology on $A$, called the **relative topology** (verify this as an exercise).

(3) We now have a more concise definition of a topological submanifold. Let $M \subseteq \mathbb{R}^n$. Then $M$ is a **$k$-dimensional topological submanifold** of $\mathbb{R}^n$ if for every $p \in M$, there exists an open neighbourhood $V$ of $p$ in $M$, an open set $U \subset \mathbb{R}^k$, and a homeomorphism $\alpha : U \subset \mathbb{R}^k \to V \subset \mathbb{R}^n$.

**Definition 1.3**

Let $A \subset \mathbb{R}^n$. Then $A$ is **connected** if it cannot be written in the form $A = U \cup V$ where $U, V \neq \varnothing$ are relatively open in $A$ and $U \cap V = \varnothing$. Otherwise, we say that $A$ is **disconnected**; we call $U$ and $V$ **disconnecting sets** for $A$.

Let's go over a few example of connected sets. It can be shown that an open set in $\mathbb{R}^n$ is connected if and only if it is path connected; that is, there is a path between any two points in the set. This result can help us build some intuition for what a connected set should look like.

(1) $\mathbb{R}^n$ is connected.

(2) Let $\alpha < \beta \in \mathbb{R}$. Then $(\alpha, \beta)$, $(\alpha, \beta]$, $[\alpha, \beta)$, and $[\alpha, \beta]$ are all connected.

(3) Observe that $A = (-1, 0] \cup [1, 2]$ is a disconnected set because $(-1, 0] = A \cap (-1.3, 0.3)$ and $[1, 2] = A \cap (0.9, 2.1)$ are both relatively open in $A$ and disjoint.

(4) The open ball $B_r(p)$ is connected for all $p \in \mathbb{R}^n$ and $r > 0$.

An important property of connected sets is that the continuous image of a connected set is connected! This can be used to prove that a subset $M \subset \mathbb{R}^n$ is not a submanifold.

(1) Consider the $\alpha$-**curve** $C := \{(x, y) \in \mathbb{R}^2 : y^2 = x^2(x + 1)\} \subset \mathbb{R}^2$. This can be parametrized by the map

$$\begin{aligned} \alpha : \mathbb{R} &\to C \subset \mathbb{R}^2 \\ t &\mapsto (t^2 - 1, t(t^2 - 1)). \end{aligned}$$

Note that $\alpha$ is not injective since $\alpha(-1) = \alpha(1) = (0, 0)$. That is, $\alpha$ is not a homeomorphism on $\mathbb{R}$, but it becomes one if we remove the points $t = \pm 1$, whose inverse is

$$\begin{aligned} \alpha^{-1} : C \setminus \{(0, 0)\} &\to \mathbb{R} \setminus \{\pm 1\} \subset \mathbb{R} \\ (x, y) &\mapsto 1/x. \end{aligned}$$

Thus, $C$ is a 1-dimensional submanifold away from the point $(0, 0)$.

Our goal now is to show that the whole of $C$ is not a topological submanifold of $\mathbb{R}^2$. By contradiction, suppose that it were. By our above discussion, it must have dimension 1 because it has dimension 1 away from $(0, 0)$. Since $(0, 0) \in C$, it follows from the definition that there exists an open neighbourhood $V$ of $(0, 0)$ in $C$, an open set $U \subset \mathbb{R}^1$, and a homeomorphism

$$\alpha : U \subset \mathbb{R}^2 \to V \subset C.$$

There must be a unique point $t_0 \in U$ such that $\alpha(t_0) = (0, 0)$ since $\alpha$ is a bijection. Since $U$ is open and $t_0 \in U$, there exists $\varepsilon > 0$ such that $B_\varepsilon(t_0) \subset U$. But $U \subset \mathbb{R}^1$, so $B_\varepsilon(t_0) = (t_0 - \varepsilon, t_0 + \varepsilon) =: U'$. Then $\alpha|_{U'}$ is also a homeomorphism. Let $V' = \alpha(U')$. Observe that $V' \setminus \{(0, 0)\}$ has three or four pieces depending on how large the open set $U'$ is: one on the top right quadrant, one on the bottom right quadrant, and one or two on the left of the $y$-axis. On the other hand, $U' \setminus \{t_0\}$ has only two components, contradicting the fact that homeomorphisms preserve the number of connected components.

(2) Consider the **double cone** $M = \{(x, y, z) \in \mathbb{R}^3 : z^2 = x^2 + y^2\} \subset \mathbb{R}^3$. Away from $(0, 0, 0)$, every point in $M$ lies on the graph of one of the continuous functions $f_1(x, y) = \sqrt{x^2 + y^2}$ or $f_2(x, y) = -\sqrt{x^2 + y^2}$. Therefore, $M \setminus \{(0, 0, 0)\}$ is a 2-dimensional topological submanifold of $\mathbb{R}^3$ since $f_1$ and $f_2$ are both functions of two variables.

However, there is a problem at the point $(0, 0, 0)$ since it lies on the graph of both $f_1$ and $f_2$. Suppose that $M$ is a topological submanifold of $\mathbb{R}^3$. Then $M$ must necessarily be of dimension 2 because $M \setminus \{(0, 0, 0)\}$ is of dimension 2. Then by definition, there exists an open neighbourhood $V$ of $(0, 0, 0)$, an open set $U \subset \mathbb{R}^2$, and a homeomorphism

$$\alpha : U \subset \mathbb{R}^2 \to V \subset M.$$

Since $\alpha$ is a bijection, there exists a unique point $(x_0, y_0) \in U$ such that $\alpha(x_0, y_0) = (0, 0, 0)$. After shrinking $U$ (by the same argument as above), we may take $U' = B_\varepsilon((x_0, y_0))$ to ensure that we have a connected set. Consider now the restriction $\alpha|_{U'} : U' \to V = \alpha(U')$. Then $U' \setminus \{(x_0, y_0)\}$ has one component, whereas $V' \setminus \{(0, 0, 0)\}$ has two components (that is, it is disconnected), which is a contradiction.

Now, we want to prove the invariance of dimension.

**Theorem 1.4: Invariance of Dimension**

$\mathbb{R}^m$ is homeomorphic to $\mathbb{R}^n$ if and only if $m = n$.

If $m = n$, then $\mathbb{R}^m = \mathbb{R}^n$, so there is nothing to prove. The other implication is much harder, and we'll need the following result.

**Theorem 1.5: Brouwer Invariance of Domain**

Let $U \subset \mathbb{R}^n$ be open, and let $f : U \subset \mathbb{R}^n \to \mathbb{R}^n$ be an injective continuous map. Then $f(U) \subset \mathbb{R}^n$ is open. In particular, $f$ is a homeomorphism onto its image.

An elementary proof can be found on Terry Tao's blog where he uses the Brouwer Fixed Point Theorem to prove it. Nowadays, the standard proof uses algebraic topology.

It is important that both the domain and codomain involve the same dimension $n$. For example, consider the injective continuous map $f : \mathbb{R} \to \mathbb{R}^2$ defined by $x \mapsto (x, 0)$. Observe that $f(U)$ is the $x$-axis, which is not open in $\mathbb{R}^2$.

**Proof of Theorem 1.4.**

We proceed by contradiction. Suppose that there is a homeomorphism $f : \mathbb{R}^m \to \mathbb{R}^n$ and that $m > n$. Consider the inclusion

$$\begin{aligned} \iota : \mathbb{R}^n &\to \mathbb{R}^m \\ (x_1, \ldots, x_n) &\mapsto (x_1, \ldots, x_m, 0, \ldots 0), \end{aligned}$$

which is an injective continuous map. Then $\iota \circ f : \mathbb{R}^m \to \mathbb{R}^m$ is also an injective continuous map since it is the composition of two injective continuous maps. By Theorem 1.5, we have that $\iota \circ f(\mathbb{R}^m)$ is an open set in $\mathbb{R}^m$. But this is impossible because if $(x_1, \ldots, x_n, 0, \ldots, 0) \in \iota \circ f(\mathbb{R}^m)$, then

$$(x_1, \ldots, x_n, \varepsilon/2, 0, \ldots, 0) \notin \iota \circ f(\mathbb{R}^m)$$

for all $\varepsilon > 0$. Then $B_\varepsilon((x_1, \ldots, x_n, 0, \ldots, 0)) \not\subset \iota \circ f(\mathbb{R}^m)$ for all $\varepsilon > 0$, implying that $\iota \circ f(\mathbb{R}^m)$ is not open in $\mathbb{R}^m$. Therefore, we must have $m \leq n$. If $n < m$, then we can repeat the same argument with $f^{-1} : \mathbb{R}^n \to \mathbb{R}^m$, which again leads to a contradiction. We conclude that $n = m$. $\square$

Note that we actually proved something stronger: if $m > n$ and $U$ is a nonempty open subset of $\mathbb{R}^m$, then there is no continuous mapping from $U$ to $\mathbb{R}^n$. As a consequence, we get the following.

**Proposition 1.6**

If $M \subset \mathbb{R}^n$ is a $k$-dimensional topological submanifold of $\mathbb{R}^n$, then $k \leq n$.

**Proof of Proposition 1.6.**

If $M \subset \mathbb{R}^n$ is a $k$-dimensional topological submanifold of $\mathbb{R}^n$, then for all $p \in M$, there exists an open set $U \subset \mathbb{R}^k$, an open neighbourhood $V \subset M$ of $p$, and a homeomorphism

$$\alpha : U \subset \mathbb{R}^k \to V \subset M \subset \mathbb{R}^n.$$

Since $\alpha$ is an injective continuous map, this forces $k \leq n$ by the above discussion. $\square$

Finally, we must have the same $k$ for any chart $\alpha : U \subset \mathbb{R}^k \to V \subset M$. Indeed, let $p \in M$, and suppose that we have two different charts, say $\alpha : U \subset \mathbb{R}^k \to V \subset M$ and $\beta : U' \subset \mathbb{R}^{k'} \to V' \subset M$ where $p \in V \cap V'$. Then $V \cap V' \neq \varnothing$, so we can consider the restrictions

$$\alpha|_{\alpha^{-1}(V\cap V')} : \alpha^{-1}(V \cap V') \to V \cap V',$$
$$\beta|_{\beta^{-1}(V\cap V')} : \beta^{-1}(V \cap V') \to V \cap V'.$$

Then $\beta^{-1} \circ \alpha : \alpha^{-1}(V \cap V') \subset \mathbb{R}^k \to \beta^{-1}(V \cap V') \subset \mathbb{R}^{k'}$ is a homeomorphism. Hence, $\alpha^{-1}(V \cap V')$ is a $k$-dimensional topological submanifold of $\mathbb{R}^{k'}$. By Proposition 1.6, we have $k \leq k'$. Similarly, $\alpha^{-1} \circ \beta : \beta^{-1}(V \cap V') \subset \mathbb{R}^{k'} \to \alpha^{-1}(V \cap V') \subset \mathbb{R}^k$ is a homeomorphism, so $\beta^{-1}(V \cap V')$ is a $k'$-dimensional submanifold of $\mathbb{R}^k$. It follows that $k' \leq k$ and so $k' = k$.

## 1.4 Submanifolds of $\mathbb{R}^n$ of class $\mathcal{C}^r$

Let $U \subset \mathbb{R}^n$ be an open set and consider the vector-valued function

$$F : U \subset \mathbb{R}^n \to \mathbb{R}^m$$
$$x = (x_1, \ldots, x_n) \mapsto (F_1(x), \ldots, F_m(x)).$$

Recall that $F$ is of **class** $\mathcal{C}^r$ for $r \geq 1$ if each component function $F_i : U \subset \mathbb{R}^n \to \mathbb{R}$ is of class $\mathcal{C}^r$. That is, the partial derivatives of $F_i$ exist and are continuous up to order $r$. Also, we say that $F$ is of class $\mathcal{C}^\infty$ or **smooth** if each $F_i$ is smooth (the partial derivatives exist up to any order).

(1) All polynomials are smooth.

(2) The function $f(x) = x^{4/3}$ is of class $\mathcal{C}^1$. Its derivative $f'(x) = \frac{4}{3}x^{1/3}$ is continuous, but the second derivative $f''(x) = \frac{4}{9}x^{-2/3}$ is not defined at $x = 0$.

(3) The vector-valued function $F(x, y) = (2\cos x, xy - 1, e^{2\sin y + x})$ is smooth on $\mathbb{R}^2$ because each component function is smooth.

The **partial derivative** of $F$ with respect to the variable $x_j$ is

$$\frac{\partial F}{\partial x_j} := \left(\frac{\partial F_1}{\partial x_j}, \ldots, \frac{\partial F_m}{\partial x_j}\right).$$

If we fix a component function $F_i$, its **gradient** is

$$\nabla F_i := \left(\frac{\partial F_i}{\partial x_1}, \ldots, \frac{\partial F_i}{\partial x_n}\right).$$

The **derivative matrix** or **Jacobian matrix** of $F$ is the $m \times n$ matrix

$$DF := \begin{bmatrix} \partial F_1/\partial x_1 & \cdots & \partial F_1/\partial x_n \\ \vdots & \ddots & \vdots \\ \partial F_m/\partial x_1 & \cdots & \partial F_m/\partial x_n \end{bmatrix}.$$

That is, the rows correspond to the component functions $F_i$, and the columns correspond to the variables $x_j$. We can also think of the rows as the gradients and the columns as the partial derivatives; that is, we have

$$DF = \left[\begin{array}{c|c|c} \dfrac{\partial F}{\partial x_1} & \cdots & \dfrac{\partial F}{\partial x_n} \end{array}\right] = \left[\begin{array}{c} \nabla F_1 \\ \hline \vdots \\ \hline \nabla F_m \end{array}\right]$$

In general, we want to work with some differentiability. This leads to the following definition.

**Definition 1.7**

Let $M \subset \mathbb{R}^n$. Suppose that for every $p \in M$, there exists an open neighbourhood $V$ of $p$ in $M$, an open subset $U \subset \mathbb{R}^k$, and a homeomorphism $\alpha : U \subset \mathbb{R}^k \to V \subset M$ such that

(1) $\alpha$ is of class $\mathcal{C}^r$ for some $r \geq 1$;

(2) $D\alpha(x)$ has rank $k$ for all $x \in U$.

Then $M$ is called a **$k$-dimensional submanifold of $\mathbb{R}^n$ of class $\mathcal{C}^r$**. We call $\alpha$ a **coordinate chart** (or **coordinate patch**) about $p$.

Note that every submanifold of class $\mathcal{C}^r$ is a topological submanifold. We are only imposing the extra conditions (1) and (2) on the coordinate charts. We will see that condition (2) will allow us to define tangent spaces to the submanifolds at every point. A submanifold of class $\mathcal{C}^\infty$ is called a **smooth submanifold**.

As usual, let's go over some examples.

(1) Let $U \subset \mathbb{R}^n$ be open. Then $\alpha : U \subset \mathbb{R}^n \to V = U \subset \mathbb{R}^n$ sending $x$ to itself is smooth. Since the component functions are $F_i(x) = x_i$ for all $i = 1, \dots, n$, we have

$$D\alpha(x) = \left[\frac{\partial F_i}{\partial x_j}\right] = \left[\frac{\partial x_i}{\partial x_j}\right] = [\delta_{ij}],$$

where $\delta_{ij}$ is the Kronecker delta. In other words, $D\alpha(x)$ is the $n \times n$ identity matrix and has rank $n$ for all $x \in U$, so $U \subset \mathbb{R}^n$ is a smooth $n$-dimensional submanifold of $\mathbb{R}^n$.

(2) **Graphs of functions of class $\mathcal{C}^r$.** Let $U \subset \mathbb{R}^k$ be an open set and consider a function

$$\begin{aligned} F : U \subset \mathbb{R}^k &\to \mathbb{R}^{n-k} \\ (x_1, \dots, x_k) &\mapsto (F_1(x), \dots, F_{n-k}(x)) \end{aligned}$$

of class $\mathcal{C}^r$ (so each $F_i$ is of class $\mathcal{C}^r$). Let

$$M = \{(x, F(x)) \in \mathbb{R}^k \times \mathbb{R}^{n-k} : x \in U\} \subset \mathbb{R}^n$$

be the graph of $F$. We have already seen that $M$ is a $k$-dimensional submanifold of $\mathbb{R}^n$ by taking $V = M$ and the homeomorphism $\alpha : U \subset \mathbb{R}^k \to V \subset \mathbb{R}^n$ defined by

$$\alpha(x) = (x, F(x)) = (x_1, \dots, x_k, F_1(x), \dots, F_{n-k}(x)).$$

In particular, $\alpha$ is of class $\mathcal{C}^r$ since the identity components are smooth and the $F_i$ are of class $\mathcal{C}^r$. Let's look at the derivative matrix in terms of the columns of partial derivatives. We have

$$\frac{\partial \alpha}{\partial x_j} = \left(0, \dots, 0, 1, 0, \dots, 0, \frac{\partial F_1}{\partial x_j}, \dots, \frac{\partial F_{n-k}}{\partial x_j}\right)$$

where the 1 corresponds to the $j$-th component, and hence

$$D\alpha(x) = \left[\begin{array}{cccc} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \\ \hline \partial F_1/\partial x_1 & \partial F_1/\partial x_2 & \cdots & \partial F_1/\partial x_k \\ \vdots & \vdots & \ddots & \vdots \\ \partial F_{n-k}/\partial x_1 & \partial F_{n-k}/\partial x_2 & \cdots & \partial F_{n-k}/\partial x_k \end{array}\right] = \left[\begin{array}{c} I_{k \times k} \\ \hline DF(x) \end{array}\right].$$

This matrix has rank $k$ for all $x \in U$, so $M$ is a $k$-dimensional submanifold of class $\mathcal{C}^r$.

(3) We saw that the circle $\mathbb{S}^1 = \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 = 1\} \subset \mathbb{R}^2$ was a 1-dimensional topological submanifold using the charts $\alpha_1 : U_1 = (-\pi, \pi) \subset \mathbb{R}^1 \to \mathbb{S}^1 \setminus \{(-1, 0)\} \subset \mathbb{R}^2$ and $\alpha_2 : U_2 = (0, 2\pi) \subset \mathbb{R}^1 \to \mathbb{S}^1 \setminus \{(1, 0)\} \subset \mathbb{R}^2$, both defined by $t \mapsto (\cos t, \sin t)$. Note that both $\alpha_i$ are smooth functions with derivative matrix
$$D\alpha_i = \begin{bmatrix} \frac{\mathrm{d}\alpha_i}{\mathrm{d}t} \end{bmatrix} = \begin{bmatrix} -\sin t \\ \cos t \end{bmatrix},$$
which has rank 1 because $\sin t$ and $\cos t$ don't have the same zeroes, and hence $D\alpha_i$ is never the zero vector. Thus, $\mathbb{S}^1$ is a smooth 1-dimensional submanifold of $\mathbb{R}^2$.

Not every topological submanifold of $\mathbb{R}^n$ is of class $\mathcal{C}^r$ for some $r \geq 1$. For example, consider the graph
$$M = \{(x, |x|) : x \in \mathbb{R}\}$$
of the function $f(x) = |x|$ on $\mathbb{R}$, which fails to be differentiable at $x = 0$. Since $f$ is continuous, we know that $M$ is a 1-dimensional topological submanifold of $\mathbb{R}^2$. Note that $f$ is smooth away from $x = 0$, so $M \setminus \{(0, 0)\}$ is a smooth 1-dimensional submanifold of $\mathbb{R}^2$.

However, we claim that it cannot be a submanifold of class $\mathcal{C}^r$ on any neighbourhood of the point $(0, 0)$. Suppose otherwise, so there exists an open set $U \subset \mathbb{R}^1$, an open neighbourhood $V \subset M$ of $(0, 0)$, and a homeomorphism $\alpha : U \subset \mathbb{R}^1 \to V \subset M \subset \mathbb{R}^2$ of class $\mathcal{C}^r$ for some $r \geq 1$. Moreover, assume that $D\alpha(t)$ is of rank 1 for all $t \in U$. Since $k = 1$, we have
$$D\alpha(t) = \frac{\mathrm{d}\alpha}{\mathrm{d}t} = \alpha'(t) \neq 0$$
for all $t \in U$ using the rank 1 assumption. But $\alpha'(t)$ is tangent to $M$ at $\alpha(t)$. There are two possibilities:

- If $\alpha(t)$ is on the line $y = x$, then $\alpha'(t)$ is a direction vector of $y = x$, so for some $c : I \subset \mathbb{R} \to \mathbb{R}$, we have
$$\alpha'(t) = c(t) \begin{pmatrix} 1 \\ 1 \end{pmatrix}. \tag{1.4.1}$$
Since $\alpha(t)$ is of class $\mathcal{C}^r$ for some $r \geq 1$, we know that $\alpha'(t)$ is continuous, so $c : I \to \mathbb{R}$ is also continuous.
- If $\alpha(t)$ is on $y = -x$, then $\alpha'(t)$ is a direction vector of $y = -x$, so for some $d : I' \subset \mathbb{R} \to \mathbb{R}$, we have
$$\alpha'(t) = d(t) \begin{pmatrix} 1 \\ -1 \end{pmatrix}. \tag{1.4.2}$$
The same argument as above shows that $d : I' \to \mathbb{R}$ is continuous.

However, since $\alpha$ is a bijection, we have $(0, 0) = \alpha(t_0)$ for some $t_0 \in U$. By the continuity of $\alpha'(t)$, we obtain
$$\lim_{t \to t_0^-} \alpha'(t) = \lim_{t \to t_0^+} \alpha'(t).$$

Without loss of generality, assume that $\alpha(t)$ is moving along $M$ from left to right. (Otherwise, we can simply parametrize in the other direction.) Then if $t < t_0$, equation (1.4.2) holds, whereas if $t > t_0$, equation (1.4.1) holds. This means that
$$\lim_{t \to t_0^-} \alpha'(t) = \lim_{t \to t_0^-} d(t) \begin{pmatrix} 1 \\ -1 \end{pmatrix} = \lim_{t \to t_0^+} c(t) \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \lim_{t \to t_0^+} \alpha'(t).$$

But the above vectors are not parallel to each other, so the only way that these limits are equal is if $\lim_{t \to t_0^-} d(t) = \lim_{t \to t_0^+} c(t) = 0$. This implies that $\alpha'(t_0)$ is the zero vector, which is a contradiction to the fact that $D\alpha(t) = \alpha'(t)$ has rank 1 for all $t \in U$! This concludes the example that not every topological submanifold is a submanifold of class $\mathcal{C}^r$ for some $r \geq 1$.

In the definition of a submanifold of class $\mathcal{C}^r$, it is important that $\alpha$ is a homeomorphism and not just a function of class $\mathcal{C}^r$ with $D\alpha(x)$ of rank $k$ for all $x \in U$. These conditions alone don't even ensure that $\alpha$ is injective! For example, take the $\alpha$-curve $C = \{(x,y) \in \mathbb{R}^2 : y^2 = x^2(x+1)\} \subset \mathbb{R}^2$ which we introduced back in Section 1.3, which could be parametrized using $\alpha(t) = (t^2-1, t(t^2-1))$ for $t \in \mathbb{R}$. We saw that this map was not injective, but it is smooth with derivative matrix

$$D\alpha(t) = \frac{\mathrm{d}\alpha}{\mathrm{d}t} = \begin{pmatrix} 2t \\ 3t^2-1 \end{pmatrix} \neq \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

for all $t \in \mathbb{R}$. A map satisfying these two conditions is said to be an **immersion**, and a topological submanifold whose maps are immersions is called an **immersed manifold**. We record this in the definition below.

**Definition 1.8**

Let $U \subset \mathbb{R}^k$ be open with $k \leq n$. A map $\alpha : U \subset \mathbb{R}^k \to \mathbb{R}^n$ is an **immersion** (of class $\mathcal{C}^r$) if

(1) $\alpha$ is of class $\mathcal{C}^r$; and

(2) $D\alpha(x)$ has rank $k$ for all $x \in U$.

We give some examples of immersions.

(1) **Canonical immersion.** The inclusion map $\iota : \mathbb{R}^k \to \mathbb{R}^n$ defined by $(x_1, \ldots, x_k) \mapsto (x_1, \ldots, x_k, 0, \ldots, 0)$ is an immersion of class $\mathcal{C}^\infty$. Indeed, we see that $\iota$ is smooth and its derivative matrix is

$$D\iota = \left[\begin{array}{ccc} 1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 1 \\ \hline 0 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 0 \end{array}\right]$$

which has rank $k$ because it contains the $k \times k$ identity matrix.

(2) The parametrization $\alpha(t) = (t^2-1, t(t^2-1))$ of the $\alpha$-curve is an immersion of class $\mathcal{C}^\infty$.

(3) The charts $\alpha : U \to V$ of class $\mathcal{C}^r$ of a submanifold $M \subset \mathbb{R}^n$ of class $\mathcal{C}^r$ are immersions of class $\mathcal{C}^r$.

Recall that a **diffeomorphism** is a differentiable bijection (of class $\mathcal{C}^r$) whose inverse is also differentiable (of class $\mathcal{C}^r$). The following proposition tells us that up to a diffeomorphism (by composition), every immersion is locally the canonical immersion.

**Proposition 1.9**

Let $U \subset \mathbb{R}^k$ be open and $\alpha : \mathbb{R}^k \to \mathbb{R}^n$ be an immersion of class $\mathcal{C}^r$. Then up to a local diffeomorphism, $\alpha$ is the canonical immersion $\iota : \mathbb{R}^k \to \mathbb{R}^n$ defined by $\iota(x_1, \ldots, x_k) = (x_1, \ldots, x_k, 0, \ldots, 0)$.

In order to prove this, we require the Inverse Function Theorem.

**Theorem 1.10: Inverse Function Theorem**

Let $U \subset \mathbb{R}^\ell$ be open and let $F : U \subset \mathbb{R}^\ell \to \mathbb{R}^\ell$ be of class $\mathcal{C}^r$. Suppose that for some $x_0 \in U$, the $\ell \times \ell$ derivative matrix $DF(x_0)$ is invertible (that is, $\det(DF(x_0)) \neq 0$). Then $F$ is invertible in an open neighbourhood $U_0 \subset U$ of $x_0$ and $F^{-1} : V_0 = F(U_0) \subset \mathbb{R}^\ell \to U_0 \subset \mathbb{R}^\ell$ is also of class $\mathcal{C}^r$.

Note that when we take $\ell = 1$ in the Inverse Function Theorem and we have a function $f : U \subset \mathbb{R}^1 \to \mathbb{R}^1$, then $Df(x) = f'(x)$. If $f'(x_0) \neq 0$, then either $f'(x) > 0$ around $x_0$ or $f'(x) < 0$. In particular, $f$ is increasing or decreasing around $x_0$, which implies that it is strictly monotone and thus invertible around $x_0$.

**Proof of Proposition 1.9.**

Suppose that $\alpha : U \subset \mathbb{R}^k \to \mathbb{R}^n$ is an immersion given by

$$x = (x_1, \ldots, x_k) \mapsto (f_1(x), \ldots, f_n(x)).$$

The derivative matrix of $\alpha$ is the $n \times k$ matrix

$$D\alpha(x) = \begin{bmatrix} \nabla f_1(x) \\ \vdots \\ \nabla f_n(x) \end{bmatrix}$$

and $D\alpha(x)$ has rank $k$ for all $x \in U$ by definition. Then $k$ of the rows of $D\alpha(x)$ are linearly independent. Without loss of generality, we can assume that these are the first $k$ rows after possibly permuting the variables in $\mathbb{R}^n$. We now divide $\alpha$ into the parts

$$\alpha(x) = (\alpha_1(x), \alpha_2(x)),$$

where $\alpha_1 : U \subset \mathbb{R}^k \to \mathbb{R}^k$ corresponds to the first $k$ component functions of $\alpha$, and $\alpha_2 : U \subset \mathbb{R}^k \to \mathbb{R}^{n-k}$ corresponds to the remaining $n - k$ component functions. Then we can write

$$D\alpha(x) = \left[\begin{array}{c} D\alpha_1(x) \\ \hline D\alpha_2(x) \end{array}\right]$$

where $D\alpha_1(x)$ is a $k \times k$ matrix and $D\alpha_2(x)$ is an $(n-k) \times k$ matrix. Notice that $D\alpha_1(x)$ has rank $k$, which implies that $D\alpha_1(x)$ is invertible. By the Inverse Function Theorem (Theorem 1.10), there exist open neighbourhoods $U_0 \subset U$ of $x_0$ and $V_0 \subset \mathbb{R}^k$ of $\alpha_1(x_0)$ such that

$$\alpha_1|_{U_0} : U_0 \subset U \subset \mathbb{R}^k \to V_0 \subset \mathbb{R}^k$$

is invertible with inverse $(\alpha_1|_{U_0})^{-1} : V_0 \subset \mathbb{R}^k \to U_0 \subset U \subset \mathbb{R}^k$ of class $\mathcal{C}^r$. Hence, $\alpha_1|_{U_0}$ is a diffeomorphism of class $\mathcal{C}^r$. Now, consider the composition $\alpha \circ \alpha_1^{-1} : V_0 \subset \mathbb{R}^k \to \mathbb{R}^n$ which yields

$$x \mapsto \alpha(\alpha_1^{-1}(x)) = (\alpha_1(\alpha_1^{-1}(x)), \alpha_2(\alpha_1^{-1}(x))) =: (x, f(x))$$

where $f : V_0 \subset \mathbb{R}^k \to \mathbb{R}^{n-k}$. So up to the diffeomorphism, $\alpha$ is the parametrization of the graph of a function of class $\mathcal{C}^r$ (because the composition is of class $\mathcal{C}^r$).

If $f(x) = \mathbf{0}$ for all $x \in V_0$, then we are already done. Otherwise, compose the above function with $h : V_0 \times \mathbb{R}^{n-k} \subset \mathbb{R}^n \to \mathbb{R}^n$ defined by

$$(x_1, \ldots, x_k, x_{k+1}, \ldots, x_n) \mapsto (x_1, \ldots, x_k, x_{k+1} - f_{k+1} \circ \alpha^{-1}(x), \ldots, x_n - f_n \circ \alpha^{-1}(x)),$$

which is a diffeomorphism of class $\mathcal{C}^r$. For $x \in V_0$, we obtain $h \circ \alpha \circ \alpha^{-1}(x) = (x, \mathbf{0}) = \iota(x)$ as desired. $\square$

As a consequence, we have the following result.

**Corollary 1.11**

The image of an immersion of class $\mathcal{C}^r$ is locally the graph of a function of class $\mathcal{C}^r$ (up to diffeomorphism). In particular, any submanifold of $\mathbb{R}^n$ of class $\mathcal{C}^r$ is locally the graph of a function of class $\mathcal{C}^r$.

**Proof of Corollary 1.11.**

From the proof of Proposition 1.9, we had a map $\alpha \circ \alpha_1^{-1} : V_0 \subset \mathbb{R}^k \to \mathbb{R}^n$ which sent $x \mapsto (x, f(x))$ by taking $f = \alpha_2 \circ \alpha_1^{-1}$ which was of class $\mathcal{C}^r$. Then $\alpha(V_0)$ is the graph of $f$ in $\mathbb{R}^n$.

If $M \subset \mathbb{R}^n$ is a $k$-dimensional submanifold of class $\mathcal{C}^r$, then for all $p \in M$, there exists a coordinate chart

$$\alpha : U \subset \mathbb{R}^k \to V \subset M$$

with $p \in V$. Set $x_0 = \alpha^{-1}(p)$. Using the above notation, there exist open sets $U_0, V_0 \subset \mathbb{R}^k$ with $x_0 \in U_0$ and a diffeomorphism $\alpha_1 : U_0 \subset \mathbb{R}^k \to V_0 \subset \mathbb{R}^k$ such that

$$\begin{aligned} \tilde{\alpha} := \alpha \circ \alpha_1^{-1} : V_0 \subset \mathbb{R}^k &\to \alpha(U_0) \subset M \\ x &\mapsto (x, f(x)) \end{aligned}$$

is of class $\mathcal{C}^r$ since $\alpha$ and $\alpha_1^{-1}$ are of class $\mathcal{C}^r$. (Note that $\alpha(U_0)$ is open in $M$ since $\alpha$ is a homeomorphism.) This implies that $f$ is of class $\mathcal{C}^r$, so $M$ is locally the graph of the $\mathcal{C}^r$ function $f : V_0 \subset \mathbb{R}^k \to \mathbb{R}^{n-k}$. $\square$

We now introduce the notion of an embedding.

**Definition 1.12**

Let $U \subset \mathbb{R}^k$ be open and $\alpha : U \subset \mathbb{R}^k \to \mathbb{R}^n$ be a map. Then $\alpha$ is an **embedding** of class $\mathcal{C}^r$ if

(1) $\alpha$ is a homeomorphism onto its image;

(2) $\alpha$ is of class $\mathcal{C}^r$;

(3) $D\alpha(x)$ has rank $k$ for all $x \in U$.

In other words, an embedding is just an immersion that is homeomorphic onto its image. In particular, submanifolds of $\mathbb{R}^n$ of class $\mathcal{C}^r$ are subsets of $\mathbb{R}^n$ that are locally the image of an embedding of class $\mathcal{C}^r$.

The following proposition tells us that embeddings are in fact diffeomorphisms onto their images.

**Proposition 1.13**

Let $U \subset \mathbb{R}^k$ be open, and let $\alpha : U \subset \mathbb{R}^k \to \mathbb{R}^n$ be an embedding of class $\mathcal{C}^r$. Then

$$\alpha^{-1} : \alpha(U) \subset \mathbb{R}^n \to U \subset \mathbb{R}^k$$

is also of class $\mathcal{C}^r$.

**Proof of Proposition 1.13.**

Note that $\alpha$ is an immersion of class $\mathcal{C}^r$, so we can write it as $\alpha(x) = (\alpha_1(x), \alpha_2(x))$, where $\alpha_1 : U \subset \mathbb{R}^k \to \mathbb{R}^k$ is locally invertible. Letting $\pi : \mathbb{R}^n \to \mathbb{R}^k$ be the projection $(y_1, \ldots, y_n) \mapsto (y_1, \ldots, y_k)$, we have

$$\begin{aligned} \alpha^{-1} : \alpha(U) \subset \mathbb{R}^n &\to U \subset \mathbb{R}^k \\ (\alpha_1(x), \alpha_2(x)) &\mapsto x = \alpha_1^{-1} \circ \pi(x_1, \ldots, x_n). \end{aligned}$$

But $\alpha_1^{-1}$ and $\pi$ are both of class $\mathcal{C}^r$, so $\alpha^{-1}$ is also of class $\mathcal{C}^r$. $\square$

We now introduce atlases, which we could've done a while ago when we defined topological submanifolds at the beginning. However, we can now talk about atlases of class $\mathcal{C}^r$.

Definition 1.14

Let $M \subset \mathbb{R}^n$ be a $k$-dimensional topological submanifold. An **atlas** of $M$ is a collection of charts

$$\{\alpha_a : U_a \subset \mathbb{R}^k \to V_a \subset M\}_{a \in A}$$

such that $\bigcup_{a \in A} V_a = M$, where each $U_a \subset \mathbb{R}^k$ and $V_a \subset M$ is open. Moreover, if all the charts in the atlas are of class $\mathcal{C}^r$, then it is called an **atlas of class** $\mathcal{C}^r$, or a **smooth atlas** if it is of class $\mathcal{C}^\infty$.

We look at some examples of atlases.

(1) Let $U \subset \mathbb{R}^n$ be open. Then $\alpha = \mathrm{Id}_U$ is a smooth chart for $U$ such that every point in $U$ is contained in $\alpha(U) = U$. Therefore, $\{\alpha = \mathrm{Id}_U : U \to U\}$ is a smooth atlas for $U$.

(2) Consider the graph $M$ of a function $F : U \subset \mathbb{R}^k \to \mathbb{R}^{n-k}$ of class $\mathcal{C}^r$, where $U \subset \mathbb{R}^k$ is open. Then $M = \{(x, F(x)) : x \in U\}$ is a $k$-dimensional submanifold of class $\mathcal{C}^r$. It admits the chart $\alpha : U \subset \mathbb{R}^k \to M \subset \mathbb{R}^n$ defined by $x \mapsto (x, F(x))$ of class $\mathcal{C}^r$, so $\{\alpha : U \to M\}$ is an atlas of class $\mathcal{C}^r$ for $M$.

(3) Recall that the points on the circle $\mathbb{S}^1 = \{(x, y) : x^2 + y^2 = 1\} \subset \mathbb{R}^2$ can be described using the charts $\alpha_1 : (-\pi, \pi) \to \mathbb{S}^1 \setminus \{(-1, 0)\}$ and $\alpha_2 : (0, 2\pi) \to \mathbb{S}^1 \setminus \{(1, 0)\}$ via $t \mapsto (\cos t, \sin t)$. Then $\{\alpha_1, \alpha_2\}$ is a smooth atlas for $\mathbb{S}^1$.

Note that if $M$ is compact (that is, closed and bounded), then any atlas of $M$ must contain at least two charts. Moreover, atlases are not unique in general! For example, consider $\{\alpha : \mathbb{R} \to M,\, t \mapsto (t, 0)\}$ and $\{\beta : \mathbb{R} \to M,\, t \mapsto (-t, 0)\}$, which are both smooth atlases for the $x$-axis $M$ in $\mathbb{R}^2$.

We now give an alternate definition of a $k$-dimensional submanifold of $\mathbb{R}^n$ of class $\mathcal{C}^r$. We will soon prove that this is equivalent to the original definition.

Definition 1.15

Let $M \subset \mathbb{R}^n$ be such that $M$ is locally given by the zero set $\{F \equiv \mathbf{0}\}$ of a $\mathcal{C}^r$ map $F : V \subset \mathbb{R}^n \to \mathbb{R}^{n-k}$ with maximal rank. That is, $DF(x)$ has rank $n - k$ for all $x \in V \cap M$, where $V \cap M = F^{-1}(\mathbf{0})$ holds for an appropriately chosen neighbourhood $V$ of every point in $M$. Then $M$ is called a **$k$-dimensional submanifold of $\mathbb{R}^n$ of class $\mathcal{C}^r$**.

This alternate definition is useful, because it is generally easier to show that a given subset is locally the zero set of a function than to explicitly exhibit charts covering the space.

(1) Consider the circle $\mathbb{S}^1 = \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 = 1\}$, which is given by the equation $x^2 + y^2 = 1$, which we can rearrange as $x^2 + y^2 - 1 = 0$. Define $F : \mathbb{R}^2 \to \mathbb{R}$ via $F(x, y) = x^2 + y^2 - 1$. Then $\mathbb{S}^1 = F^{-1}(\mathbf{0})$, so we can take $V = \mathbb{R}^2$ in the definition. Also, we have

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

which has rank 1 for every point since $(0, 0) \notin \mathbb{S}^1$. Thus, under the alternate definition, $\mathbb{S}^1$ is a smooth 1-dimensional submanifold of $\mathbb{R}^2$.

(2) More generally, the $n$-sphere $\mathbb{S}^n = \{(x_1, \ldots, x_{n+1}) \in \mathbb{R}^{n+1} : x_1^2 + \cdots + x_{n+1}^2 = 1\} \subset \mathbb{R}^{n+1}$ can be viewed as the zero set of the smooth function $F : \mathbb{R}^{n+1} \to \mathbb{R}$ given by $F(x_1, \ldots, x_{n+1}) = x_1^2 + \cdots + x_{n+1}^2 - 1$. Here, we can take $V = \mathbb{R}^{n+1}$ in the definition. Since

$$DF(x) = \begin{bmatrix} 2x_1 & \cdots & 2x_{n+1} \end{bmatrix}$$

has rank 1 for all $x \in \mathbb{S}^n$, we see that $\mathbb{S}^n$ is a smooth $n$-dimensional submanifold of $\mathbb{R}^{n+1}$.

(3) Consider the twisted cubic $M = \{(x, y, z) : y = x^2, z = x^3\} \subset \mathbb{R}^3$. We have seen that this is a smooth 1-dimensional submanifold of $\mathbb{R}^3$. For all $(x, y, z) \in M$, we have $y = x^2$ and $z = x^3$. Rearranging gives $y - x^2 = 0$ and $z - x^3 = 0$, so we can view $M$ has the zero set of the smooth function $F(x, y, z) = (y - x^2, z - x^3)$ defined for all $(x, y, z) \in \mathbb{R}^3$. This time, we have $V = \mathbb{R}^3$, and $F : \mathbb{R}^3 \to \mathbb{R}^2$ so that $n = 3$ and $k = 1$. The derivative matrix of $F$ is

$$DF(x, y, z) = \begin{bmatrix} -2x & 1 & 0 \\ -3x^2 & 0 & 1 \end{bmatrix}$$

which is rank 2 for all $(x, y, z) \in M$ because the last two columns are the $2 \times 2$ identity matrix. Hence, $M$ is a smooth 1-dimensional submanifold of $\mathbb{R}^3$.

**Remark.** Not all zero sets of functions of class $\mathcal{C}^r$ are submanifolds of class $\mathcal{C}^r$. Take the $\alpha$-curve $C = \{(x, y) \in \mathbb{R}^2 : y^2 = x^2(x + 1)\}$, which is the zero set of the function $F : \mathbb{R}^2 \to \mathbb{R}$ defined by $(x, y) \mapsto y^2 - x^2(x + 1)$. The derivative matrix of $F$ is

$$DF(x, y) = \begin{bmatrix} -3x^2 - 2x & 2y \end{bmatrix},$$

which is equal to the zero vector if and only if $(x, y) \in \{(0, 0), (-2/3, 0)\}$. We see that $(0, 0) \in C$, and this is the problematic point.

(4) **(Graph of a function of class $\mathcal{C}^r$.)** Let $f : U \subset \mathbb{R}^k \to \mathbb{R}^{n-k}$ be function of class $\mathcal{C}^r$, where $U \subset \mathbb{R}^k$ is open. Let $M = \{(x, f(x)) \in \mathbb{R}^k \times \mathbb{R}^{n-k} : x \in U\} \subset \mathbb{R}^n$, which we have already seen is a $k$-dimensional submanifold of class $\mathcal{C}^r$. We now view this using the zero set characterization.

For all $(x, y) \in M \subset \mathbb{R}^k \times \mathbb{R}^{n-k}$, we have $y = f(x)$ if and only if $F(x, y) := f(x) - y = \mathbf{0} \in \mathbb{R}^{n-k}$. Note that $U \times \mathbb{R}^{n-k}$ is an open subset of $\mathbb{R}^k \times \mathbb{R}^{n-k}$. Then $M$ is the zero set of the $\mathcal{C}^r$ function $F : U \times \mathbb{R}^{n-k} \subset \mathbb{R}^k \times \mathbb{R}^{n-k} \to \mathbb{R}^{n-k}$ given by $(x, y) \mapsto f(x) - y$.

It remains to check the rank condition. We have that

$$DF(x, y) = \left[ \begin{array}{c|c} \dfrac{\partial F}{\partial x} & \dfrac{\partial F}{\partial y} \end{array} \right]$$

which has $n - k$ rows. Since $F(x, y) = f(x) - y$, we see that $\partial F/\partial x = Df(x)$ and $\partial F/\partial y = -I_{n-k}$, so $DF(x, y)$ has rank $n - k$ since $I_{n-k}$ does.

**Theorem 1.16**

Let $M \subset \mathbb{R}^n$. The following are equivalent:

(i) $M$ is a $k$-dimensional submanifold of class $\mathcal{C}^r$ (using Definition 1.7).

(ii) $M$ is locally the graph of a function $f : U \subset \mathbb{R}^k \to \mathbb{R}^{n-k}$ of class $\mathcal{C}^r$, where $U \subset \mathbb{R}^k$ is open.

(iii) $M$ is locally the zero set of a $\mathcal{C}^r$ function $F : V \subset \mathbb{R}^n \to \mathbb{R}^{n-k}$ of maximal rank, where $V \subset \mathbb{R}^n$ is open (using Definition 1.15).

Due to Corollary 1.11, we already know that (i) and (ii) are equivalent. Example (4) above shows that (ii) implies (iii). Therefore, it suffices to prove that (iii) implies (ii), and we will see that this is a direct consequence of the Implicit Function Theorem. Let's recall what this says.

Suppose that $\mathbb{R}^n = \mathbb{R}^{k+m} = \mathbb{R}^k \times \mathbb{R}^m$ has coordinates $(x, y) = (x_1, \ldots, x_k, y_1, \ldots, y_m)$. Let $U \subset \mathbb{R}^n$ be an open subset and $F : U \subset \mathbb{R}^n \to \mathbb{R}^m$ be a function of class $\mathcal{C}^r$. Then the derivative matrix of $F$ is

$$DF(x, y) = \left[ \begin{array}{c|c} \dfrac{\partial F}{\partial x} & \dfrac{\partial F}{\partial y} \end{array} \right].$$

where if $F_1, \ldots, F_m$ are the component functions of $F$, then we have $\partial F/\partial x = (\partial F_i/\partial x_j)_{1 \le i \le m, 1 \le j \le n}$ and $\partial F/\partial y = (\partial F_i/\partial y_j)_{1 \le i,j \le m}$. In particular, $\partial F/\partial y$ is an $m \times m$ matrix.

**Theorem 1.17: Implicit Function Theorem**

Let $(x_0, y_0) \in U$ be such that $F(x_0, y_0) = \mathbf{0}$. Suppose that

$$\det\left(\frac{\partial F}{\partial y}(x_0, y_0)\right) \neq 0.$$

Then there exists an open neighbourhood $V_0 \subset \mathbb{R}^k$ of $x_0$ and a unique function $g : V_0 \to \mathbb{R}^m$ of class $\mathcal{C}^r$ such that $g(x_0) = y_0$ and $F(x, g(x)) = \mathbf{0}$ for all $x \in V_0$.

In other words, the Implicit Function Theorem tells us that if $\det(\partial F(x_0, y_0)/\partial y) \neq 0$, then for all points $(x, y) \in \{F \equiv \mathbf{0}\}$ in an open neighbourhood of $(x_0, y_0)$, we have $y = g(x)$ for some function of class $\mathcal{C}^r$. Thus, we can express the variables $(y_1, \dots, y_m)$ as functions of $(x_1, \dots, x_k)$ of class $\mathcal{C}^r$ near $(x_0, y_0)$.

Before proving the theorem, we illustrate what the result tells us with a simple example. Let $F(x, y) = x^2 + y^2 - 1$ for all $(x, y) \in \mathbb{R}^2$. Then $F : \mathbb{R}^2 \to \mathbb{R}$ is smooth and has derivative matrix

$$DF(x, y) = \begin{bmatrix} 2x & 2y \end{bmatrix}.$$

In this case, we have $m = 1$ and $m + k = 2$ so that $k = 1$. We are writing $\mathbb{R}^2 = \mathbb{R}^1 \times \mathbb{R}^1$ where $x$ corresponds to the first copy of $\mathbb{R}^1$ and $y$ corresponds to the second copy. Then $\partial F/\partial y = 2y$, which is nonzero if and only if $y \neq 0$. By the Implicit Function Theorem, the points on

$$\{F \equiv 0\} = \{(x, y) : x^2 + y^2 - 1 = 0\}$$

have a $y$-coordinate that can be expressed locally as a function of $x$. This is indeed true since $x^2 + y^2 - 1 = 0$ if and only if $y = \pm\sqrt{1 - x^2}$, which is smooth for $x \notin \{\pm 1\}$ and hence for $y \neq 0$ on $\{F \equiv 0\}$. These parametrize $\mathbb{S}^1 \setminus \{(\pm 1, 0)\}$. Similarly, note that $\partial F/\partial x = 2x \neq 0$ if and only if $x \neq 0$, so this holds for the points $(x, y) \neq (0, \pm 1)$ on $\mathbb{S}^1$. We see that $\mathbb{S}^1$ can be expressed as $x = \pm\sqrt{1 - y^2}$ away from $(0, \pm 1)$.

**Proof of Implicit Function Theorem (Theorem 1.17).**

This follows from the Inverse Function Theorem (Theorem 1.10). Define $H : U \subset \mathbb{R}^n \to \mathbb{R}^n = \mathbb{R}^k \times \mathbb{R}^m$ by $(x, y) \mapsto (x, F(x, y))$. The derivative matrix is

$$DH = \left[\begin{array}{c|c} I_{k\times k} & \mathbf{0}_{m\times k} \\ \hline \partial F/\partial x & \partial F/\partial y \end{array}\right],$$

which is an $n \times n$ matrix. Since $I_{k\times k}$ has rank $k$ and $\partial F/\partial y$ has rank $m$ at $(x_0, y_0)$ using the fact that it is invertible there, it follows that $DH(x_0, y_0)$ has rank $k + m = n$. That is, $\det(DH(x_0, y_0)) \neq 0$ and $H$ is locally invertible with some inverse $G$ by the Inverse Function Theorem.

Write $G(u, v) = (G_1(u, v), G_2(u, v))$ where $u \in \mathbb{R}^k$ and $v \in \mathbb{R}^m$, and we separate $G$ into the components $G_1(u, v) \in \mathbb{R}^k$ and $G_2(u, v) \in \mathbb{R}^m$. Then we have

$$\begin{aligned}(u, v) &= H \circ G(u, v) \\ &= H(G_1(u, v), G_2(u, v)) \\ &= (G_1(u, v), F(G_1(u, v), G_2(u, v))).\end{aligned}$$

This implies that $u = G_1(u, v)$, so $G(u, v) = (u, G_2(u, v))$ for some function $G : V_0 \subset \mathbb{R}^n \to \mathbb{R}^m$ of class $\mathcal{C}^r$ (where $V_0 \subset \mathbb{R}^n$ is open). Moreover, for all $(x, y)$ with $F(x, y) = \mathbf{0}$, we have $H(x, y) = (x, F(x, y)) = (x, \mathbf{0})$, and hence $(x, y) = G \circ H(x, y) = G(x, \mathbf{0}) = (x, G_2(x, \mathbf{0}))$. Then $y = G_2(x, \mathbf{0})$ for all $(x, y) \in \{F \equiv \mathbf{0}\}$. By setting $g(x) := G_2(x, \mathbf{0})$, we have $y = g(x)$ for all $(x, y) \in \{F \equiv \mathbf{0}\}$ near $(x_0, y_0)$, and $g$ is of class $\mathcal{C}^r$. We see that $F(x, g(x)) = \mathbf{0}$. For the proof of uniqueness of $g$ and more details, we refer to *Topology* by Munkres, Theorem 9.2 on page 74. ☐

Finally, we prove that the definitions are equivalent. Recall from our earlier discussion that it suffices to show that (iii) implies (ii).

**Proof of Theorem 1.16.**

Suppose that $M$ is locally the zero set of a $\mathcal{C}^r$ function $F : V \subset \mathbb{R}^n \to \mathbb{R}^{n-k}$ of maximal rank on $M \cap V$. Then for all $(x_0, y_0) \in M \cap V \subset \mathbb{R}^k \times \mathbb{R}^{n-k} = \mathbb{R}^n$, the derivative matrix

$$DF(x_0, y_0) = \left[ \frac{\partial F}{\partial x}(x_0, y_0) \;\middle|\; \frac{\partial F}{\partial y}(x_0, y_0) \right]$$

has rank $n-k$, where $\frac{\partial F}{\partial y}(x_0, y_0)$ is an $(n-k) \times (n-k)$ matrix.

After possibly permuting the variables $(x_1, \ldots, x_k, y_1, \ldots, y_{n-k})$ and therefore the columns of $DF(x_0, y_0)$, we may assume that $\partial F(x_0, y_0)/\partial y$ has rank $n-k$. This implies that $\det(\partial F(x_0, y_0)/\partial y) \neq 0$. By the Implicit Function Theorem (Theorem 1.17), we see that $y$ is a $\mathcal{C}^r$ function of $x$ on $M \cap V$ near $(x_0, y_0)$. Write $y = g(x)$ for some $\mathcal{C}^r$ function $g : V_0 \subset \mathbb{R}^k \to \mathbb{R}^m$, where $V_0$ is an open neighbourhood of $x_0$. Then $M$ is locally of the form $\{(x, g(x)) : x \in V_0\}$, which implies that $M$ is locally the graph of $g$. $\square$

To end our discussion, we give one more example of how we can use the zero set characterization to find charts for a submanifold $M$ using the Implicit Function Theorem. Consider the special linear group

$$\mathrm{SL}(2, \mathbb{R}) = \{A \in M_{2\times 2}(\mathbb{R}) : \det A = 1\}.$$

In a natural way, we can identify the matrix

$$A = \begin{bmatrix} a_1 & a_2 \\ a_3 & a_4 \end{bmatrix} \in M_{2\times 2}(\mathbb{R})$$

with the point $(a_1, a_2, a_3, a_4) \in \mathbb{R}^4$, and we have

$$\det A - 1 = a_1 a_4 - a_2 a_3 - 1 =: F(a_1, a_2, a_3, a_4)$$

so that $\mathrm{SL}(2, \mathbb{R}) = \{F \equiv 0\}$ where $F : \mathbb{R}^4 \to \mathbb{R}$ is smooth. The derivative matrix of $F$ is

$$DF(a_1, a_2, a_3, a_4) = \begin{bmatrix} a_4 & -a_3 & -a_2 & a_1 \end{bmatrix},$$

which can have rank at most 1 since it is a $1 \times 4$ matrix. In particular, $DF(a_1, a_2, a_3, a_4)$ has rank 1 if and only if it is not the zero vector, which occurs if at least one of the $a_i$ is nonzero. But for all $(a_1, a_2, a_3, a_4) \in \mathrm{SL}(2, \mathbb{R})$, we have $a_1 a_4 - a_2 a_3 = 1$ so that at least one of the $a_i$ must be nonzero.

Suppose that $a_1 \neq 0$. Then we have $\partial F/\partial a_4 = a_1 \neq 0$, so the Implicit Function Theorem tells us that $a_4$ can be expressed as a smooth function of the remaining three variables on $\mathrm{SL}(2, \mathbb{R})$. Indeed, if $(a_1, a_2, a_3, a_4) \in \mathrm{SL}(2, \mathbb{R})$ with $a_1 \neq 0$, then $a_1 a_4 - a_2 a_3 = 1$ can be rearranged to obtain

$$a_4 = \frac{a_2 a_3 + 1}{a_1}.$$

A similar analysis can be done when the other variables are nonzero. Since $DF$ has rank 1 everywhere, it follows that $\mathrm{SL}(2, \mathbb{R})$ is a smooth submanifold of $\mathbb{R}^4$ of dimension $4 - 1 = 3$.

## 1.5 Tangent vectors and tangent vector fields

The simplest example of a tangent vector is the velocity vector of a curve.

**Definition 1.18**

Let $\gamma : (a, b) \subset \mathbb{R} \to \mathbb{R}^n$ be a map of class $\mathcal{C}^r$. We define the **velocity vector** of $\gamma$ at $\gamma(t)$ to be

$$\gamma'(t) := D\gamma(t).$$

Note that if $\gamma(t) = (\gamma_1(t), \ldots, \gamma_n(t))$, then $\gamma'(t) = (\gamma_1'(t), \ldots, \gamma_n'(t))$. Moreover, we have

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

Observe that $(\gamma(t) - \gamma(t_0))/(t - t_0)$ is the velocity vector of the secant $L$ passing through $\gamma(t)$ and $\gamma(t_0)$. Taking the limit as $t \to t_0$, it follows that $\gamma'(t_0)$ is the tangent vector to the curve in $\mathbb{R}^n$ given by $\gamma(t)$, under the assumption that $\gamma'(t_0) \neq \mathbf{0}$. Let's look at some examples.

(1) Let $x \in \mathbb{R}^n$ and $\mathbf{0} \neq v \in \mathbb{R}^n$. Set $\gamma(t) = x + tv$ for $t \in \mathbb{R}$, which parametrizes the line $L$ in $\mathbb{R}^n$ passing through $x$ with direction vector $v$. We have that $\gamma'(t) = v$ for all $t \in \mathbb{R}$, so at every point on the line, the velocity vector coincides with the direction vector.

(2) Let $\gamma : \mathbb{R} \to \mathbb{R}^2$ be defined by $t \mapsto (\cos t, \sin t)$, which parametrizes $\mathbb{S}^1$. Note that $\gamma'(t) = (-\sin t, \cos t) \neq (0, 0)$ for all $t \in \mathbb{R}$.

(3) Let $\gamma : \mathbb{R} \to \mathbb{R}^2$ be given by $t \mapsto (t^2, t^3)$. This parametrizes the cusp curve $y^2 = x^3$. We have $\gamma'(t) = (2t, 3t^2)$. Observe that $\gamma'(0) = (0, 0)$ so that the velocity vector is zero at $\gamma(0) = (0, 0)$.

(4) Recall that the $\alpha$-curve can be parametrized with $\gamma : \mathbb{R} \to \mathbb{R}^2$ defined by $t \mapsto (t^2 - 1, t(t^2 - 1))$. We have $\gamma'(t) = (2t, 3t^2 - 1) \neq (0, 0)$ for all $t \in \mathbb{R}$.

Note that if $\gamma$ is a homeomorphism onto its image

$$C = \{x \in \mathbb{R}^n : x = \gamma(t) \text{ for some } t \in (a, b)\}$$

and $D\gamma(t)$ has rank 1 for all $t \in (a, b)$, then $C$ is a 1-dimensional submanifold of $\mathbb{R}^n$ of class $\mathcal{C}^r$. In fact, we see that $D\gamma(t)$ has rank 1 if and only if $\gamma'(t) \neq \mathbf{0}$, so $\gamma'(t)$ is a direction vector for the tangent line to $\gamma$ at $\gamma(t)$. In particular, the 1-dimensional submanifold of $\mathbb{R}^n$ determined by $\gamma$ has a well-defined tangent line $L$ at every point. We call $L := \{\gamma(t_0) + s\gamma'(t_0) : s \in \mathbb{R}\}$ the **tangent line to $C$ at $\gamma(t_0)$**.

From example (1) above, we know that a line $\gamma(t) = x + tv$ coincides with its tangent line at every point.

**Definition 1.19**

Let $x \in \mathbb{R}^n$. A **tangent vector to $\mathbb{R}^n$ at $x$** is defined as a pair $(x; v)$ where $v \in \mathbb{R}^n$. We call

$$T_x(\mathbb{R}^n) := \{(x; v) : v \in \mathbb{R}^n\}$$

the **tangent space to $\mathbb{R}^n$ at $x$**. This is the set of all tangent vectors to $\mathbb{R}^n$ at $x$.

We can give $T_x(\mathbb{R}^n)$ a vector space structure with the operations $(x; v) + (x; w) = (x; v + w)$ and $c(x; v) = (x; cv)$ for all $c \in \mathbb{R}$ and $v, w \in \mathbb{R}^n$. Note that $T_x(\mathbb{R}^n) \simeq \mathbb{R}^n$ as vector spaces, where the isomorphism $T_x(\mathbb{R}^n) \to \mathbb{R}^n$ is given by $(x; v) \mapsto v$.

**Lemma 1.20**

For all $x \in \mathbb{R}^n$, we have

$$T_x(\mathbb{R}^n) = \{(x; v) : v \text{ is a velocity vector of some curve } \gamma(t) \text{ passing through } x\}.$$

**Proof of Lemma 1.20.**

Let $(x; v) \in T_x(\mathbb{R}^n)$. Set $\gamma(t) = x + tv$ for $t \in \mathbb{R}$, which satisfies $\gamma'(t) = v$ for all $t \in \mathbb{R}$. In particular, we have $\gamma(0) = x$ and $\gamma'(0) = v$. Note that if $v \neq \mathbf{0}$, then $\gamma(t)$ parametrizes the line through $x$ with velocity vector $v$. $\square$

Can we get something similar for submanifolds $M \subset \mathbb{R}^n$? Towards this direction, we make a new definition.

**Definition 1.21**

Let $U \subset \mathbb{R}^k$ be open and let $\alpha : U \to \mathbb{R}^n$ be of class $\mathcal{C}^r$. Also, let $x \in U$ and set $p = \alpha(x)$. The map

$$\begin{aligned} \alpha_* : T_x(\mathbb{R}^k) &\to T_p(\mathbb{R}^n) \\ (x; v) &\mapsto (p; D\alpha(x)v) \end{aligned}$$

is called the **pushforward of $\alpha$ at $x$.**

Note that $\alpha_* : T_x(\mathbb{R}^k) \to T_p(\mathbb{R}^n)$ is a vector space homomorphism since it is linear; in particular, it is given by multiplication by the $n \times k$ derivative matrix.

For all $(x; v) \in T_x(\mathbb{R}^k)$, consider the map $\gamma : (-\varepsilon, \varepsilon) \to \mathbb{R}^n$ defined by $t \mapsto \alpha(x + tv)$ with $\varepsilon > 0$ such that $x + tv \in U$ for all $t \in (-\varepsilon, \varepsilon)$. That is, the line segment $L = \{x + tv : t \in (-\varepsilon, \varepsilon)\}$ is included in the open set $U$. Then for all $t_0 \in (-\varepsilon, \varepsilon)$, the chain rule gives

$$\gamma'(t_0) = D\alpha(x + t_0 v)D(x + tv)(t_0) = D\alpha(x + t_0 v)v$$

since $D(x + tv)(t_0) = v$ for all $t_0 \in \mathbb{R}$. In particular, we obtain

$$\begin{aligned} (\gamma(0); \gamma'(0)) &= (\alpha(x); D\alpha(x)v) \\ &= (p, D\alpha(x)v) \\ &= \alpha_*(x; v). \end{aligned}$$

This tells us that $\alpha_*(x; v)$ is the velocity vector of $\gamma(t) := \alpha(x + tv)$ at $p = \gamma(0)$.

**Lemma 1.22**

If $\alpha : U \subset \mathbb{R}^k \to \mathbb{R}^n$ and $\beta : V \subset \mathbb{R}^\ell \to \mathbb{R}^k$ are functions of class $\mathcal{C}^r$ where $U \subset \mathbb{R}^k$ and $V \subset \mathbb{R}^\ell$ are open with $\beta(V) \subset U$, then $(\alpha \circ \beta)_* = \alpha_* \circ \beta_*$ on $V$.

**Proof of Lemma 1.22.**

Note that $\alpha \circ \beta : V \subset \mathbb{R}^\ell \to \mathbb{R}^n$ is of class $\mathcal{C}^r$. Let $x \in V$. Then for all $v \in \mathbb{R}^\ell$, we have

$$\begin{aligned} (\alpha \circ \beta)_*(x; v) &= (\alpha \circ \beta(x); D(\alpha \circ \beta)(x)v) \\ &= (\alpha(\beta(x)); D\alpha(\beta(x))D\beta(x)v) \\ &= \alpha_*(\beta(x); D\beta(x)v) \\ &= \alpha_*(\beta_*(x; v)), \end{aligned}$$

where the second equality follows from the chain rule. $\square$

From Lemma 1.22, it now makes sense to define the following.

**Definition 1.23**

Let $M \subset \mathbb{R}^n$ be a $k$-dimensional submanifold of $\mathbb{R}^n$ of class $\mathcal{C}^r$ for some $r \geq 1$. Let $p \in M$ and let $\alpha : U \subset \mathbb{R}^k \to V \subset M$ be a coordinate chart of $M$ about $p$ (i.e. $p \in V$) of class $\mathcal{C}^r$. We define the **tangent space of $M$ at $p$** to be

$$T_p(M) := \alpha_*(T_{x_0}(\mathbb{R}^k)) \subset T_p(\mathbb{R}^n),$$

where $x_0 \in U$ is the unique point in $U$ such that $\alpha(x_0) = p$.

Note that the requirement that $M$ is of class $\mathcal{C}^r$ with $r \geq 1$ is crucial to define the tangent space. We now make some observations about this definition.

(1) Since $D\alpha(x)$ has rank $k$ for all $x \in U$, we have $\dim_{\mathbb{R}}(T_pM) = k$.

(2) Let $p \in V$ and let $x_0 \in U$ be such that $p = \alpha(x_0)$. Note that we have

$$T_pM = \operatorname{span}_{\mathbb{R}} \left\{ \frac{\partial \alpha}{\partial x_1}(x_0), \dots, \frac{\partial \alpha}{\partial x_k}(x_0) \right\}.$$

Indeed, we know that we can write

$$D\alpha(x_0) = \left[ \frac{\partial \alpha}{\partial x_1}(x_0) \;\middle|\; \cdots \;\middle|\; \frac{\partial \alpha}{\partial x_k}(x_0) \right],$$

which implies that $\alpha_*(T_{x_0}\mathbb{R}^k)$ is spanned by $(p; D\alpha(x_0)e_i)$ with $i = 1, \dots, k$, where $\{e_1, \dots, e_k\}$ is the standard basis of $\mathbb{R}^k$. In fact, this spanning set is a basis of $T_pM$ since $\dim_{\mathbb{R}}(T_pM) = k$ and $D\alpha(x_0)$ has rank $k$ so that its columns are linearly independent.

(3) The definition is independent of the choice of coordinate chart. To see this, suppose that $\alpha' : U' \subset \mathbb{R}^k \to V' \subset M$ is another coordinate chart about $p$. Then we have

$$T_pM = \operatorname{span}_{\mathbb{R}} \left\{ \frac{\partial \alpha'}{\partial y_1}(y_0), \dots, \frac{\partial \alpha'}{\partial y_k}(y_0) \right\},$$

where $y_0 \in U'$ is the point such that $\alpha'(y_0) = p$ and $(y_1, \dots, y_k)$ are the coordinates on $U' \subset \mathbb{R}^k$. Given $x = (x_1, \dots, x_k) \in U$, observe that

$$y = (y_1, \dots, y_k) = (\alpha')^{-1} \circ \alpha(x_1, \dots, x_k)$$

for some $y \in U'$, where $(\alpha')^{-1} \circ \alpha$ is of class $\mathcal{C}^r$ since both $(\alpha')^{-1}$ and $\alpha$ are of class $\mathcal{C}^r$. In particular, we can think of this as a $\mathcal{C}^r$ function

$$y(x) = (\alpha')^{-1} \circ \alpha(x)$$

for all $x \in W = \alpha^{-1}(V \cap V')$ and $y \in W' = (\alpha')^{-1}(V \cap V')$. By the chain rule, we obtain

$$\begin{aligned}
\frac{\partial \alpha}{\partial x_j}(x_0) &= \frac{\partial(\alpha' \circ (\alpha')^{-1} \circ \alpha)}{\partial x_j}(x_0) \\
&= \frac{\partial(\alpha'(y(x_0)))}{\partial x_j} \\
&= \sum_{i=1}^{k} \frac{\partial \alpha'}{\partial y_i}(y(x_0)) \cdot \frac{\partial y_i}{\partial x_j}(x_0) \\
&= \sum_{i=1}^{k} \frac{\partial y_i}{\partial x_j}(x_0) \cdot \frac{\partial \alpha'}{\partial y_i}(y_0),
\end{aligned}$$

noting that $y_0 = y(x_0)$. Since $\partial y_i(x_0)/\partial x_j \in \mathbb{R}$ for all $i = 1, \dots, k$, we see that

$$\frac{\partial \alpha}{\partial x_j}(x_0) \in \operatorname{span}_{\mathbb{R}} \left\{ \frac{\partial \alpha'}{\partial y_1}(y_0), \dots, \frac{\partial \alpha'}{\partial y_k}(y_0) \right\}.$$

An analogous argument shows that

$$\frac{\partial \alpha'}{\partial y_i}(y_0) \in \operatorname{span}_{\mathbb{R}} \left\{ \frac{\partial \alpha}{\partial x_1}(x_0), \dots, \frac{\partial \alpha'}{\partial x_k}(x_0) \right\}.$$

(4) We now show that $T_pM$ is the space of velocity vectors of curves passing through $p$ on $M$, giving us a generalization of Lemma 1.20. That is, we have

$$T_pM = \{(p; v) : v = \gamma'(t_0) \text{ for some curve } \gamma(t) \text{ on } M \text{ with } p = \gamma(t_0)\} \simeq \mathbb{R}^k.$$

Indeed, let $\sigma : U \subset \mathbb{R}^k \to V \subset M$ be a chart of $M$ about $p$ and suppose that $p = \alpha(x_0)$ for some $x_0 \in U$. By definition, we know that $T_pM = \alpha_*(T_{x_0}\mathbb{R}^k)$ where

$$\alpha_*(x_0; w) = (\alpha(x_0); D\alpha(x_0)w) = (p; D\alpha(x_0)w)$$

for all $(x_0; w) \in T_{x_0}\mathbb{R}^k$. In particular, this means that for any $(p; v) \in T_pM$, we have

$$(p; v) = (p; D\alpha(x_0)w)$$

for some $w \in \mathbb{R}^k$. Let $\varepsilon > 0$ be small enough so that the line segment $x_0 + tw$ over $t \in (-\varepsilon, \varepsilon)$ is contained in $U$. Then we see that

$$\begin{aligned} \gamma : (-\varepsilon, \varepsilon) &\to V \subset M \\ t &\mapsto \alpha(x_0 + tw) \end{aligned}$$

is a curve of class $\mathcal{C}^r$ on $M$ such that $\gamma(0) = \alpha(x_0) = p$. Moreover, the chain rule gives

$$\gamma'(t) = D\alpha(x_0 + tw)\frac{\mathrm{d}}{\mathrm{d}t}(x_0 + tw) = D\alpha(x_0 + tw)w.$$

For $t = 0$, we obtain $\gamma'(0) = D\alpha(x_0)w = v$. This implies that $\gamma : (-\varepsilon, \varepsilon) \to V$ is a curve on $M$ with $\gamma(0) = p$ and $\gamma'(0) = v$. We leave the converse as an exercise.

Now that we have digested the definition, let's look at some examples of tangent spaces.

(1) Let $U \subset \mathbb{R}^n$ be an open set. Then the identity map $\alpha : U \to U$ which sends $x \in U$ to itself is a smooth chart with $D\alpha(x) = I_{n\times n}$ for all $x \in U$. By definition, we have

$$T_x(U) = \alpha_*(T_x\mathbb{R}^n).$$

Then for all $(x; v) \in T_x\mathbb{R}^n$, we get

$$\alpha_*(x; v) = (\alpha(x); D\alpha(x)v) = (x; v).$$

This implies that $T_xU = T_x\mathbb{R}^n$ for all $x \in U$.

(2) We saw two different ways of seeing if something is a submanifold. If $f : U \subset \mathbb{R}^k \to \mathbb{R}^{n-k}$ is a function of class $\mathcal{C}^r$ where $U \subset \mathbb{R}^k$ is open and $M = \{(x, f(x)) \in \mathbb{R}^n : x \in U\}$ is the graph of $f$ in $\mathbb{R}^n$, then

$$\begin{aligned} \alpha : U \subset \mathbb{R}^k &\to \mathbb{R}^n \\ x = (x_1, \dots, x_k) &\mapsto (x, f(x)) \end{aligned}$$

is a chart of class $\mathcal{C}^r$ with derivative matrix

$$D\alpha(x) = \left[\frac{I_{k\times k}}{Df(x)}\right]$$

for all $x \in U$. Let $p \in M$ so that $p = \alpha(x_0) = (x_0, f(x_0))$ for some $x_0 \in U$. Then

$$\begin{aligned} T_pM &= \alpha_*(T_{x_0}\mathbb{R}^k) \\ &= \{\alpha_*(x_0; v) : v \in T_{x_0}\mathbb{R}^k\} \\ &= \{(\alpha(x_0); D\alpha(x_0)v) : v \in \mathbb{R}^k\} \\ &= \{(p; w) : w = (v, Df(x_0)v),\, v \in \mathbb{R}^k\}, \end{aligned}$$

since $\alpha(x_0) = p$ and $D\alpha(x)$ is a block matrix with $I_{k\times k}$ upstairs and $Df(x)$ downstairs. We also have

$$T_p(M) = \operatorname{span}_{\mathbb{R}}\left\{\frac{\partial\alpha}{\partial x_1}(x_0), \dots, \frac{\partial\alpha}{\partial x_k}(x_0)\right\} = \operatorname{span}_{\mathbb{R}}\left\{\left[\frac{e_i}{\partial f(x_0)/\partial x_i}\right] : i = 1, \dots, k\right\}.$$

(3) Let $U \subset \mathbb{R}^n$ be open. If $F : U \subset \mathbb{R}^n \to \mathbb{R}^{n-k}$ is a function of class $\mathcal{C}^r$ with $DF$ having maximal rank $n-k$ on $M = \{F \equiv \mathbf{0}\}$ (so that $M$ is a $k$-dimensional submanifold of $\mathbb{R}^n$ by Theorem 1.16), then we leave it as an exercise to show that

$$T_pM = \ker(DF(p)).$$

In particular, if $M$ has dimension $k = n-1$, then $F : U \subset \mathbb{R}^n \to \mathbb{R}$ is a scalar function. Then $DF(p) = \nabla F(p)$, which implies that

$$\begin{aligned} T_pM &= \ker(DF(p)) \\ &= \{v \in \mathbb{R}^n : DF(p)v = 0\} \\ &= \{v \in \mathbb{R}^n : \nabla F(p) \cdot v = 0\}. \end{aligned}$$

We see that $T_pM$ is the $(n-1)$-dimensional subspace of $T_p\mathbb{R}^n$ which is perpendicular to $\nabla F(p)$. In other words, $T_pM$ is the hyperplane in $\mathbb{R}^n$ through $p$ with normal vector $\nabla F(p)$.

For example, consider the $n$-sphere

$$M = \mathbb{S}^n = \{x \in \mathbb{R}^{n+1} : \|x\|^2 = 1\},$$

which is the zero set of the function $F : \mathbb{R}^{n+1} \to \mathbb{R}$ defined by $F(x) = \|x\|^2 - 1 = x_1^2 + \cdots + x_{n+1}^2 - 1$. The derivative matrix is just the gradient; that is,

$$DF(x) = \nabla F(x) = \begin{bmatrix} 2x_1 & \cdots & 2x_{n+1} \end{bmatrix} = 2x.$$

Then $T_x\mathbb{S}^n$ is the hyperplane through $x$ with normal vector $\nabla F(x) = 2x$.

# 2 Curves in $\mathbb{R}^n$

## 2.1 Introduction to curves

What is a curve? Intuitively, it is a 1-dimensional subset of $\mathbb{R}^n$.

Observe that the level sets $f(x, y) = k$ of a two variable function over $\mathbb{R}^2$ are curves.

(1) For the function $f(x, y) = x^2 + y^2$, the level sets $C = \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 = k\}$ are circles centered at $(0, 0)$ of radius $\sqrt{k}$.

(2) For the function $f(x, y) = x^2 - y$, the level sets are of the form $C = \{(x, y) \in \mathbb{R}^2 : x^2 - y = k\}$. This is simply the parabola $y = x^2 - k$.

The intersection of two surfaces in $\mathbb{R}^3$ is also a curve.

(1) The intersection of $z = x^2 + y^2$ and $z = 2$ gives the circle $x^2 + y^2 = 2$ in the plane $z = 2$.

(2) **(Twisted cubic.)** Let $C = \{(x, y, z) \in \mathbb{R}^3 : y = x^2,\ z = x^3\}$, which is the intersection of the surfaces $y = x^2$ and $z = x^3$ in $\mathbb{R}^3$. This can be described using the parametrized curve $\gamma : \mathbb{R} \to C \subset \mathbb{R}^3$ defined by $t \mapsto (t, t^2, t^3)$.

We will work with parametrized curves, which are vector-valued functions $\gamma : I = (\alpha, \beta) \to \mathbb{R}^n$ of class $\mathcal{C}^r$.

**(Circular helix.)** Let $\gamma(t) = (a \cos t, a \sin t, bt)$ for $t \in \mathbb{R}$ and $a, b > 0$. Setting $x = a \cos t$ and $y = a \sin t$, we see that $x^2 + y^2 = a^2$, so $\gamma(t)$ lies on the cylinder $x^2 + y^2 = a^2$ in $\mathbb{R}^3$. We have $\gamma(0) = (a, 0, 0)$ and $\gamma(\pi/2) = (0, a, b\pi/2)$. (The circular helix looks like a spiral along the cylinder.)

**Definition 2.1**

A parameterized curve $\gamma : (\alpha, \beta) \to \mathbb{R}^n$ of class $\mathcal{C}^r$ is called **regular** if for all $t \in (\alpha, \beta)$, we have

$$\gamma'(t) = \frac{\mathrm{d}\gamma}{\mathrm{d}t}(t) \neq 0.$$

We call $\|\gamma'(t)\|$ the **speed of $\gamma$ at** $\gamma(t)$ and we say that $\gamma$ is **unit speed** if $\|\gamma'(t)\| = 1$ for all $t \in (\alpha, \beta)$.

Note that unit speed implies regular because $\|x\| = 0$ if and only if $x = 0$. Let's look at some examples.

(1) **(Circle.)** Let $\gamma(t) = (\cos(ct), \sin(ct))$ for $t \in \mathbb{R}$, where $c \in \mathbb{R}$. Then $\gamma'(t) = c(-\sin(ct), \cos(ct))$ for all $t \in \mathbb{R}$, which implies that $\|\gamma'(t)\| = c$. In particular, $\gamma$ is regular when $c \neq 0$, and is unit speed if and only if $c = 1$.

(2) **(Circular helix.)** Given $a, b > 0$, let $\gamma(t) = (a \cos t, a \sin t, bt)$ for $t \in \mathbb{R}$ as before. Then we have $\gamma'(t) = (-a \sin t, a \cos t, b)$ for all $t \in \mathbb{R}$, which implies that

$$\|\gamma'(t)\| = \sqrt{a^2 + b^2} > 0.$$

This is a constant; we see that $\gamma$ is regular for all $a, b > 0$, and unit speed if and only if $a^2 + b^2 = 1$.

(3) **(Twisted cubic.)** Consider $\gamma(t) = (t, t^2, t^3)$ for $t \in \mathbb{R}$. Then $\gamma'(t) = (1, 2t, 3t^2)$ for all $t \in \mathbb{R}$, so

$$\|\gamma'(t)\| = \sqrt{1 + 4t^2 + 9t^4} \geq 1.$$

Notice that $\gamma$ is regular, but it is not a unit speed curve since $\|\gamma'(t)\| = 1$ if and only if $t = 0$.

The following result will be useful in our study of unit speed curves.

**Proposition 2.2**

Let $u : (\alpha, \beta) \to \mathbb{R}^n$ be a vector-valued function of class $\mathcal{C}^r$ such that $\|u(t)\| = 1$ for all $t \in (\alpha, \beta)$. Then $u(t) \cdot u'(t) = 0$ for all $t \in (\alpha, \beta)$, so we have $u'(t) = \mathbf{0}$ or $u(t) \perp u'(t)$.

**Proof of Proposition 2.2.**

Suppose that $u(t) = (u_1(t), \ldots, u_n(t))$. Then $u'(t) = (u_1'(t), \ldots, u_n'(t))$. Since $\|u(t)\| = 1$, we have that

$$1 = u(t) \cdot u(t) = \sum_{i=1}^{n} (u_i(t))^2.$$

Differentiating this equation with respect to $t$ gives

$$0 = \frac{\mathrm{d}}{\mathrm{d}t}\left(\sum_{i=1}^{n}(u_i(t))^2\right) = \sum_{i=1}^{n} \frac{\mathrm{d}}{\mathrm{d}t}((u_i(t))^2) = \sum_{i=1}^{n} 2u_i(t)u_i'(t) = 2u(t) \cdot u'(t).$$

This means that $u(t) \cdot u'(t) = 0$ for all $t \in (\alpha, \beta)$, as desired. $\square$

In particular, if $\gamma$ is a unit speed curve, then $\gamma'(t) = \mathbf{0}$ or $\gamma'(t) \perp \gamma''(t)$ by setting $u(t) = \gamma'(t)$ in Proposition 2.2. Note that this result also holds if for some constant $c \in \mathbb{R}$, we have $\|\gamma'(t)\| = c$ for all $t \in (\alpha, \beta)$.

How do we measure how much a curve "curves"? We first make a few definitions.

**Definition 2.3**

Let $\gamma : (\alpha, \beta) \to \mathbb{R}^n$ be a unit speed curve.

- We define the **unit tangent vector to** $\gamma$ **at** $\gamma(s)$ to be

$$T(s) := \gamma'(s).$$

- The **curvature of** $\gamma$ **at** $\gamma(s)$ is defined to be

$$\kappa(s) := \|\gamma''(s)\|.$$

- If $\kappa(s) > 0$, the **principal unit normal to** $\gamma$ **at** $\gamma(s)$ is

$$N(s) := \frac{\gamma''(s)}{\kappa(s)},$$

  and the **radius of curvature at** $\gamma(s)$ is $\rho(s) := 1/\kappa(s)$.

We make a few remarks about these definitions.

(1) Observe that $T(s)$ is the unique unit vector tangent to $\gamma$ at $\gamma(s)$ that points in the direction in which we are traveling along the curve.

(2) Since $\gamma$ is a unit speed curve, Proposition 2.2 implies $\gamma'(s) \cdot \gamma''(s) = 0$. Assuming that the acceleration $\gamma''(s)$ is nonzero, we must have $\gamma'(s) \perp \gamma''(s)$ so that $T(s) \perp N(s)$.

(3) Note that $\kappa(s) \geq 0$. Whenever $\kappa(s) > 0$, we see that $N(s) = \gamma''(s)/\kappa(s)$ points in the same direction as $\gamma''(s)$, namely the direction that the velocity vectors are changing.

Next, let's look at some simple examples.

(1) Let $\gamma : (\alpha, \beta) \to \mathbb{R}^n$ be a unit speed curve of class $\mathcal{C}^r$ where $r \geq 2$. Then the image of $\gamma$ in $\mathbb{R}^n$ is a line if and only if $\kappa(s) = 0$ for all $s \in (\alpha, \beta)$.

**Proof.** Note that $\kappa(s) = 0$ if and only if $\|\gamma''(s)\| = 0$, which is equivalent to $\gamma''(s) = \mathbf{0}$ for all $s \in (\alpha, \beta)$. This happens if and only if $\gamma'(s) = v$ for some fixed $\mathbf{0} \neq v \in \mathbb{R}^n$ (it is nonzero since $\gamma$ is unit speed), and hence $\gamma(s) = sv + x_0$ for some $x_0 \in \mathbb{R}^n$ and $s \in (\alpha, \beta)$. $\square$

(2) **(Circle.)** For $a > 0$ and $s \in \mathbb{R}$, set $\gamma(s) = (a\cos(s/a), a\sin(s/a))$. Then $\gamma'(s) = (-\sin(s/a), \cos(s/a))$ with $\|\gamma'(s)\| = 1$, so $\gamma$ is unit speed with

$$T(s) = \gamma'(s) = (-\sin(s/a), \cos(s/a)).$$

Next, we have $\gamma''(s) = (-\frac{1}{a}\cos(s/a), -\frac{1}{a}\sin(s/a))$, so

$$\kappa(s) = \|\gamma'(s)\| = 1/a > 0$$

and $\gamma$ has constant and positive curvature. From this, we compute

$$N(s) = \frac{\gamma''(s)}{\kappa(s)} = (-\cos(s/a), -\sin(s/a)),$$

and the radius of curvature at $\gamma(s)$ is $\rho(s) = 1/\kappa(s) = a$.

(3) **(Helix.)** Let $a, b > 0$ and set $c = (a^2 + b^2)^{-1/2}$. Then $\gamma(s) = (a\cos(cs), a\sin(cs), bcs)$ for $s \in \mathbb{R}$ is a unit speed parametrization of the helix. Then

$$T(s) = \gamma'(s) = (-ac\sin(cs), ac\cos(cs), bc)$$

and $\gamma''(s) = (-ac^2\cos(cs), -ac^2\sin(cs), 0)$, which implies that

$$\kappa(s) = \|\gamma''(s)\| = ac^2 = \frac{a}{a^2 + b^2}$$

for all $s \in \mathbb{R}$, so this has constant curvature just like the circle! Therefore, we will need more than curvature to distinguish between the two. We will need torsion, which we discuss later on. The principal unit normal to $\gamma$ at $\gamma(s)$ is

$$N(s) = \frac{\gamma''(s)}{\kappa(s)} = (-\cos(cs), -\sin(cs), 0).$$

Note that in $\mathbb{R}^2$, there are only two possible unit normal vectors to the curve at any point. The principal unit normal points inwards.

On the other hand, in $\mathbb{R}^3$, there is a whole plane of normal vectors to the curve at any point. It is perpendicular to the unit tangent $T$ and called the **normal plane** to the curve at that point. The principal unit normal $N$ is a vector in that plane.

Definition 2.4

Let $\gamma : (\alpha, \beta) \to \mathbb{R}^n$ be a unit speed curve. Suppose that $\kappa(s) > 0$ where $s \in (\alpha, \beta)$.

- The **osculating plane to** $\gamma$ **at** $\gamma(s)$ is defined to be $\operatorname{span}_{\mathbb{R}}\{T(s), N(s)\}$.
- The **osculating circle of** $\gamma$ **at** $\gamma(s)$ is the circle of radius $\rho(s)$ in the osculating plane that passes through $\gamma(s)$ and whose center is on the ray of direction $N(s)$ starting at $\gamma(s)$.

If $\gamma$ is a plane curve, then the osculating plane is the plane which contains the curve. If $\gamma$ is a circle, then it coincides with the osculating circle. The osculating circle is therefore the circle that best approximates the curve around the point $\gamma(s)$.

## 2.2 Binormal vectors, Frenet frames, and torsion

We first recall the cross product in $\mathbb{R}^3$. Given $u, v \in \mathbb{R}^3$, we denote the cross product of $u$ and $v$ by $u \times v$, and it is the unique vector in $\mathbb{R}^3$ such that:

- $u \times v \perp u$ and $u \times v \perp v$, with $u \times v$ pointing in the direction given by the right-hand rule;
- $\|u \times v\|$ is the area of the parallelogram spanned by $u$ and $v$.

Note that $u \times v = \|u\|\|v\| \sin\theta\, \mathbf{n}$, where $\theta$ is the angle between $u$ and $v$, and $\mathbf{n}$ is the unit normal vector perpendicular to the plane containing $u$ and $v$ (if $\theta \notin \{0, \pi\}$) with direction such that $\{u, v, \mathbf{n}\}$ is positively oriented.

(i) If $\theta \in \{0, \pi\}$ so that $u$ and $v$ are parallel, then $\sin\theta = 0$ and $u \times v = \mathbf{0}$. In fact, we have $u \times v = \mathbf{0}$ if and only if $u$ and $v$ are parallel.

(ii) If $\|u\| = \|v\| = 1$ and $\theta = \pi/2$, then $u \times v = \mathbf{n}$ so that $\|u \times v\| = 1$ and $\{u, v, u \times v\}$ is a positively oriented orthonormal basis.

Let $\mathbf{i} = (1, 0, 0)$, $\mathbf{j} = (0, 1, 0)$, and $\mathbf{k} = (0, 0, 1)$ so that $\{\mathbf{i}, \mathbf{j}, \mathbf{k}\}$ is the standard basis of $\mathbb{R}^3$. In practice, given $u = (u_1, u_2, u_3)$ and $v = (v_1, v_2, v_3)$, we compute $u \times v$ via

$$\begin{aligned} u \times v &= \det \begin{bmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ u_1 & u_2 & u_3 \\ v_1 & v_2 & v_3 \end{bmatrix} \\ &= (u_2 v_3 - v_2 u_3)\mathbf{i} - (u_1 v_3 - v_1 u_3)\mathbf{j} + (u_1 v_2 - v_1 u_2)\mathbf{k} \\ &= (u_2 v_3 - v_2 u_3, v_1 u_3 - u_1 v_3, u_1 v_2 - v_1 u_2). \end{aligned}$$

Finally, we observe that $v \times u = -u \times v$ because by definition, $v \times u$ points in the opposite direction of $u \times v$ but has the same length.

**Definition 2.5**

Let $\gamma : (\alpha, \beta) \to \mathbb{R}^3$ be a unit speed curve with $\kappa > 0$. The **unit binormal vector** is defined to be

$$B := T \times N,$$

and we call $\{T, N, B\}$ the **Frenet frame**.

Note that $B \perp T$ and $B \perp N$ by definition of the cross product, and $\|B\| = 1$ since $\|T\| = \|N\| = 1$ with $T \perp N$. This implies that $\{T, N, B\}$ is a (right-handed) orthonormal basis at every point.

(i) Observe that $B$ is perpendicular to the osculating plane $\operatorname{span}_{\mathbb{R}}\{T, N\}$.

(ii) We call $\operatorname{span}_{\mathbb{R}}\{N, B\}$ the **normal plane** because it is the plane perpendicular to $T$.

(iii) If $\gamma$ is a curve in the $xy$-plane, then $B$ must be constantly $(0, 0, 1)$ or $(0, 0, -1)$. In general, if $\gamma$ is a plane curve, then $B(s)$ is a constant vector (exercise) so that $\frac{\mathrm{d}B}{\mathrm{d}s}$ is identically zero. In particular, $\frac{\mathrm{d}B}{\mathrm{d}s}$ measures how much a curve fails to be a plane curve (i.e. how much it twists). We will see later that $\|\frac{\mathrm{d}B}{\mathrm{d}s}\|$ is equal to the absolute value of the torsion.

Let's consider an example. A circle or helix can parametrized using the unit speed curve

$$\gamma(s) = (a\cos(cs), a\sin(cs), bcs)$$

for $s \in \mathbb{R}$, where $a > 0$, $b \geq 0$, and $c = (a^2 + b^2)^{-1/2}$. We computed in an earlier example that $T(s) = \gamma'(s) = c(-a\sin(cs), a\cos(cs), b)$ and $N(s) = \gamma''(s)/\kappa(s) = -(\cos(cs), \sin(cs), 0)$. Then the unit binormal vector is

$$B = T \times N = \det \begin{bmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ -ac\sin(cs) & ac\cos(cs) & bc \\ -\cos(cs) & -\sin(cs) & 0 \end{bmatrix} = (bc\sin(cs), -bc\cos(cs), ac).$$

Note that if $b = 0$, then $\gamma$ is a circle in the $xy$-plane. In this case, we have $B = (0, 0, ac)$ with $c = 1/a$, so $B = (0, 0, 1)$ as expected. On the other hand, if $b > 0$, then $B$ is not a constant vector, so the osculating plane changes (and $\frac{\mathrm{d}B}{\mathrm{d}s} \neq 0$).

We have seen that $\frac{\mathrm{d}B}{\mathrm{d}s}$ measures to what extent the curve fails to be a plane curve. In addition, we have the following result.

**Lemma 2.6**

For all $s \in (\alpha, \beta)$, we have $\frac{\mathrm{d}B}{\mathrm{d}s}(s) = 0$ or $\frac{\mathrm{d}B}{\mathrm{d}s}(s)$ is parallel to $N(s)$.

**Proof of Lemma 2.6.**

Since $\|B\| = 1$, we know from Proposition 2.2 that

$$\frac{\mathrm{d}B}{\mathrm{d}s}(s) \cdot B(s) = 0$$

so that $\frac{\mathrm{d}B}{\mathrm{d}s}(s) = 0$ or $\frac{\mathrm{d}B}{\mathrm{d}s}(s) \perp B(s)$. By the product rule, we obtain

$$\frac{\mathrm{d}B}{\mathrm{d}s} = \frac{\mathrm{d}}{\mathrm{d}s}(T \times N) = \frac{\mathrm{d}T}{\mathrm{d}s} \times N + T \times \frac{\mathrm{d}N}{\mathrm{d}s} = \kappa N \times N + T \times \frac{\mathrm{d}N}{\mathrm{d}s} = T \times \frac{\mathrm{d}N}{\mathrm{d}s},$$

where we used the fact that $\frac{\mathrm{d}T}{\mathrm{d}s} = \kappa N$ and $N \times N = \mathbf{0}$. This implies that $\frac{\mathrm{d}B}{\mathrm{d}s} \perp T$. This combined with $\frac{\mathrm{d}B}{\mathrm{d}s} \perp B$ means that $\frac{\mathrm{d}B}{\mathrm{d}s}(s)$ is parallel to $B(s) \times T(s) = N(s)$. $\square$

In particular, we have that $\frac{\mathrm{d}B}{\mathrm{d}s}(s) = -\tau(s)N(s)$ for some scalar function $\tau : (\alpha, \beta) \to \mathbb{R}$.

**Definition 2.7**

The scalar function $\tau : (\alpha, \beta) \to \mathbb{R}$ such that $\frac{\mathrm{d}B}{\mathrm{d}s} = -\tau N$ is called the **torsion** of $\gamma$.

Note that $\tau$ can take both positive and negative values (and can even be zero). When $\tau > 0$, this reflects the fact that the slopes of lines of intersection with the osculating plane are increasing. In other words, the curve is twisting "up". Analogously, when $\tau < 0$, the curve is twisting "down".

Revisiting the circle and helix example above, we already computed that $N(s) = -(\cos(cs), \sin(cs), 0)$ and $B(s) = (bc\sin(cs), -bc\cos(cs), ac)$. This gives

$$\frac{\mathrm{d}B}{\mathrm{d}s}(s) = bc^2(\cos(cs), \sin(cs), 0) = -bc^2 N(s),$$

so for all $s \in \mathbb{R}$, the torsion of $\gamma$ is given by

$$\tau(s) = bc^2 = \frac{b}{a^2 + b^2} \geq 0.$$

If $b = 0$, then $\tau \equiv 0$, reflecting the fact that $\gamma$ is a plane curve; that is, it is a curve in the $xy$-plane. When $b > 0$, then $\tau \equiv b/(a^2 + b^2) > 0$, and we have a "right-handed" helix that "goes up".

**Definition 2.8**

Let $\gamma : (\alpha, \beta) \to \mathbb{R}^3$ be a curve. We say that $\gamma$ is a **plane curve** if $\gamma$ is contained in an (affine) plane $ax + by + cz = d$ in $\mathbb{R}^3$.

We give some examples of plane curves.

(1) The circle parametrized by $\gamma(t) = (\cos t, 3, \sin t)$ is a curve contained in the plane $y = 3$.

(2) Consider the curve $\gamma : \mathbb{R} \to \mathbb{R}^3$ defined by $\gamma(t) = (t^2 - 2t, t + 1, -t^2 - 2)$. We see that $\gamma$ is contained in the plane $x + 2y + z = 0$ since $t^2 - 2t + 2(t + 1) + (-t^2 - 2) = 0$, so $\gamma$ is a plane curve.

(3) **(Twisted cubic.)** Consider the curve $\gamma(t) = (t^3, t^2, t)$. To see if it is a plane curve, we can try to solve for $a, b, c, d \in \mathbb{R}$ such that $at^3 + bt^2 + ct = d$. Comparing coefficients, we find that $a = b = c = d = 0$, so $\gamma$ is not a plane curve.

**Proposition 2.9**

Let $\gamma : (\alpha, \beta) \to \mathbb{R}^3$ be a unit speed curve with $\kappa(s) > 0$ for all $s \in (\alpha, \beta)$. Then $\gamma$ is a plane curve if and only if $\tau(s) = 0$ for all $s \in (\alpha, \beta)$.

In particular, torsion captures how much $\gamma$ fails to be a plane curve.

**Proof of Proposition 2.9.**

($\Rightarrow$) Suppose that $\gamma$ is a plane curve. Let $s_0 \in (\alpha, \beta)$. Note that $\gamma$ is a plane curve if and only if every point $\gamma(s)$ is contained in some fixed plane $\Pi$ in $\mathbb{R}^3$. This means that $\gamma(s_0) \in \Pi$. Moreover, if $\mathbf{n}_0$ is a unit normal to $\Pi$, then for any $s \in (\alpha, \beta)$, the vector going from $\gamma(s_0)$ to $\gamma(s)$ is parallel to $\Pi$ and thus perpendicular to $\mathbf{n}_0$. This implies that $\mathbf{n}_0 \cdot (\gamma(s) - \gamma(s_0)) = 0$ for all $s \in (\alpha, \beta)$, or equivalently $\mathbf{n}_0 \cdot \gamma(s) = \mathbf{n}_0 \cdot \gamma(s_0)$. Hence, we deduce that

$$0 = \frac{\mathrm{d}}{\mathrm{d}s}(\mathbf{n}_0 \cdot \gamma(s)) = \frac{\mathrm{d}}{\mathrm{d}s}(\mathbf{n}_0) \cdot \gamma(s) + \mathbf{n}_0 \cdot \frac{\mathrm{d}}{\mathrm{d}s}(\gamma(s)) = \mathbf{n}_0 \cdot T(s)$$

since $\frac{\mathrm{d}}{\mathrm{d}s}(\mathbf{n}_0) = \mathbf{0}$ so that $\frac{\mathrm{d}}{\mathrm{d}s}(\mathbf{n}_0) \cdot \gamma(s) = 0$. Similarly, we have $\frac{\mathrm{d}^2}{\mathrm{d}s^2}(\mathbf{n}_0 \cdot \gamma(s)) = 0$, which implies that

$$\mathbf{n}_0 \cdot N(s) = \mathbf{n}_0 \cdot \frac{1}{\kappa(s)}\gamma''(s) = \frac{1}{\kappa(s)}\mathbf{n}_0 \cdot \gamma''(s) = 0$$

using the fact that $\gamma''(s) = \kappa(s)N(s)$. This gives us $\mathbf{n}_0 \perp T(s)$ and $\mathbf{n}_0 \perp N(s)$. Since $\mathbf{n}_0$ is a unit vector, this means that

$$\mathbf{n}_0 = \pm T(s) \times N(s) = \pm B(s)$$

by definition of the cross product. Then $\frac{\mathrm{d}B}{\mathrm{d}s}(s) = 0$ for all $s \in (\alpha, \beta)$, implying that $\tau(s) = 0$ for all $s \in (\alpha, \beta)$ since $N(s) \neq \mathbf{0}$.

($\Leftarrow$) Suppose that $\tau(s) = 0$ for all $s \in (\alpha, \beta)$. Then $\frac{\mathrm{d}B}{\mathrm{d}s}(s) = \mathbf{0}$ and hence $B(s) = \mathbf{n}_0$ for all $s \in (\alpha, \beta)$ for some $\mathbf{n}_0 \in \mathbb{R}^3$. Then $T(s) \perp \mathbf{n}_0$ for all $s \in (\alpha, \beta)$ so that $T(s) \cdot \mathbf{n}_0 = 0$. But $T(s) = \gamma'(s)$, so applying the product rule gives us

$$\frac{\mathrm{d}}{\mathrm{d}s}(\gamma(s) \cdot \mathbf{n}_0) = \gamma'(s) \cdot \mathbf{n}_0 + \gamma(s) \cdot \frac{\mathrm{d}}{\mathrm{d}s}(\mathbf{n}_0) = T(s) \cdot \mathbf{n}_0 = 0$$

since $\frac{\mathrm{d}}{\mathrm{d}s}(\mathbf{n}_0) = 0$ and $T(s) \cdot \mathbf{n}_0 = 0$. It follows that $\gamma(s) \cdot \mathbf{n}_0 = c$ for some $c \in \mathbb{R}$, and thus

$$(\gamma(s) - \gamma(s_0)) \cdot \mathbf{n}_0 = \gamma(s) \cdot \mathbf{n}_0 - \gamma(s_0) \cdot \mathbf{n}_0 = c - c = 0$$

for all $s \in (\alpha, \beta)$. Then $\gamma$ is contained in the plane passing through $\gamma(s_0)$ normal to $\mathbf{n}_0$. We conclude that $\gamma$ is a plane curve. $\square$

## 2.3 Frenet-Serret equations and the fundamental theorem of space curves

Let $\gamma : (\alpha, \beta) \to \mathbb{R}^3$ be a curve of class $\mathcal{C}^3$ and suppose that $\kappa(s) > 0$ for all $s \in (\alpha, \beta)$. We have already seen that $\frac{\mathrm{d}T}{\mathrm{d}s} = \kappa N$ and $\frac{\mathrm{d}B}{\mathrm{d}s} = -\tau N$. But what about $\frac{\mathrm{d}N}{\mathrm{d}s}$?

Since $\{T, N, B\}$ is a right-handed orthonormal basis, we have that

$$\begin{aligned} B &= T \times N, \\ T &= N \times B, \\ N &= B \times T. \end{aligned}$$

Applying the product rule to $N = B \times T$ gives

$$\begin{aligned} \frac{\mathrm{d}N}{\mathrm{d}s} = \frac{\mathrm{d}}{\mathrm{d}s}(B \times T) &= \frac{\mathrm{d}B}{\mathrm{d}s} \times T + B \times \frac{\mathrm{d}T}{\mathrm{d}s} \\ &= -\tau(N \times T) + \kappa(B \times N) \\ &= \tau(T \times N) - \kappa(N \times B) \\ &= \tau B - \kappa T. \end{aligned}$$

Therefore, we obtain the system of equations

$$\begin{aligned} \frac{\mathrm{d}T}{\mathrm{d}s} &= \kappa N, \\ \frac{\mathrm{d}N}{\mathrm{d}s} &= -\kappa T + \tau B, \\ \frac{\mathrm{d}B}{\mathrm{d}s} &= -\tau N, \end{aligned}$$

which we can write more compactly as

$$\begin{bmatrix} \frac{\mathrm{d}T}{\mathrm{d}s} \\ \frac{\mathrm{d}N}{\mathrm{d}s} \\ \frac{\mathrm{d}B}{\mathrm{d}s} \end{bmatrix} = \begin{bmatrix} 0 & \kappa & 0 \\ -\kappa & 0 & \tau \\ 0 & -\tau & 0 \end{bmatrix} \begin{bmatrix} T \\ N \\ B \end{bmatrix}.$$

These are called the **Frenet-Serret equations**. Note that the above matrix is skew-symmetric. By solving the Frenet-Serret equations, we obtain the fundamental theorem of space curves.

**Definition 2.10**

A **rigid motion** of $\mathbb{R}^3$ is a rotation followed by a translation. That is, an affine map $M : \mathbb{R}^3 \to \mathbb{R}^3$ of the form $M(v) = Av + b$ with $A \in \mathrm{SO}(3)$, where $\mathrm{SO}(3) = \{C \in M_{3\times 3}(\mathbb{R}) : C^T C = I, \det C = 1\}$ is the special orthogonal group.

**Theorem 2.11: Fundamental Theorem of Space Curves**

Let $\gamma_1$ and $\gamma_2$ be two unit speed curves in $\mathbb{R}^3$ of class $\mathcal{C}^3$ with the same curvature $\kappa(s) > 0$ and the same torsion $\tau(s)$ for all $s \in (\alpha, \beta)$. Then there is a rigid motion $M$ of $\mathbb{R}^3$ such that $\gamma_2(s) = M(\gamma_1(s))$ for all $s \in (\alpha, \beta)$.

Moreover, if $k$ and $t$ are functions of class $\mathcal{C}^3$ with $k > 0$ everywhere, then there is a unit speed curve in $\mathbb{R}^3$ whose curvature is $k$ and whose torsion is $t$.

We will not give the proof here as it is quite long. However, we can use it to classify curves.

(1) If $\gamma$ is a curve with $\kappa \equiv k > 0$ where $k$ is a constant and $\tau \equiv 0$, then $\gamma$ is (part of) a circle of radius $1/k$.

(2) If $\gamma$ is a unit speed curve where $\kappa \equiv k > 0$ is constant and $\tau \equiv \ell > 0$ is constant, then $\gamma$ is (part of) a circular helix. That is, there exists a rigid motion $M$ such that $M(\gamma(s)) = (a\cos(cs), a\sin(cs), bcs)$ for some $a, b > 0$ with $k = a/(a^2+b^2)$, $\ell = b/(a^2+b^2)$, and $c = (a^2+b^2)^{-1/2}$.

In other words, the fundamental theorem of space curves says that if we know the curvature and torsion, then we know the curve up to some rigid motion.

## 2.4 Arclength and arclength parametrization

We begin with the definition of arclength.

**Definition 2.12**

Let $\gamma : (\alpha, \beta) \to \mathbb{R}^n$ be a parametrized curve of class $\mathcal{C}^r$. For all $\alpha < \alpha_0 \le \beta_0 < \beta$, the **arclength** between $\gamma(\alpha_0)$ and $\gamma(\beta_0)$ is defined to be

$$\int_{\alpha_0}^{\beta_0} \|\gamma'(t)\| \,\mathrm{d}t.$$

That is, the arclength is the length $L$ of the curve $C$ given by $\gamma(t)$ between $\gamma(\alpha_0)$ and $\gamma(\beta_0)$. Let's take a look at a few examples of arclength.

(1) **(Circle.)** Consider $\gamma(t) = (a\cos t, a\sin t)$ where $t \in \mathbb{R}$ and $a > 0$, which parametrizes the circle $x^2+y^2 = a^2$ of radius $a$ centered at $(0,0)$. For $0 \le t \le 2\pi$, we get one copy of the circle, so the expected arclength should be $2\pi a$ as the radius is $a$. Indeed, we see that $\gamma'(t) = (-a\sin t, a\cos t)$ and $\|\gamma'(t)\| = a$ for all $t \in \mathbb{R}$, so the arclength is

$$\int_0^{2\pi} \|\gamma'(t)\| \,\mathrm{d}t = \int_0^{2\pi} a \,\mathrm{d}t = 2\pi a.$$

(2) **(Circular helix.)** We compute the arclength $L$ of the circular helix $\gamma(t) = (a\cos t, a\sin t, bt)$ between $\gamma(0) = (a,0,0)$ and $\gamma(2\pi) = (a, 0, 2\pi b)$. Here, we are integrating over the range $0 \le t \le 2\pi$. Since $\gamma'(t) = (-a\sin t, a\cos t, b)$ and

$$\|\gamma'(t)\| = \sqrt{a^2\sin^2 t + a^2\cos^2 t + b^2} = \sqrt{a^2+b^2}$$

for all $t \in \mathbb{R}$, it follows that

$$L = \int_0^{2\pi} \|\gamma'(t)\| \,\mathrm{d}t = \int_0^{2\pi} \sqrt{a^2+b^2} \,\mathrm{d}t = 2\pi\sqrt{a^2+b^2}.$$

Note that when $b = 0$, we obtain $2\pi a$, which makes sense as the circular helix collapses to a circle of radius $a$ in the $xy$-plane where $b = 0$.

(3) **(Twisted cubic.)** Consider the twisted cubic $\gamma(t) = (6t, 3t^2, t^3)$ for $t \in \mathbb{R}$. Here, we have $t = x/6$ so that $y = x^2/12$ and $z = x^3/216$. Let's compute the arclength for $0 \le t \le 1$. We have $\gamma'(t) = (6, 6t, 3t^2)$ for all $t \in \mathbb{R}$, which implies that

$$\|\gamma'(t)\| = \sqrt{36 + 36t^2 + 9t^4} = \sqrt{(6+3t^2)^2} = 6 + 3t^2.$$

It follows that the arclength is

$$\int_0^1 \|\gamma'(t)\| \,\mathrm{d}t = \int_0^1 (6+3t^2) \,\mathrm{d}t = 6t + t^3\Big|_0^1 = 7.$$

**Definition 2.13**

The **arclength** of a curve $\gamma$ of class $\mathcal{C}^r$ starting at a point $\gamma(t_0)$ is the function $s(t)$ given by

$$s(t) := \int_{t_0}^{t} \|\gamma'(u)\| \, \mathrm{d}u.$$

Given a curve $\gamma$, we are now viewing arclength as a function.

(1) We have $s(t_0) = 0$ and $s(t) < 0$ if $t < t_0$.

(2) By the fundamental theorem of calculus, we know that $s(t)$ is differentiable with $\frac{\mathrm{d}s}{\mathrm{d}t} = \|\gamma'(t)\|$ for all $t$. In fact, if $\gamma(t) = (\gamma_1(t), \dots, \gamma_n(t))$, then

$$\frac{\mathrm{d}s}{\mathrm{d}t} = \sqrt{(\gamma_1'(t))^2 + \cdots + (\gamma_n'(t))^2}.$$

Therefore, if $\gamma$ is smooth and $\gamma'(t) \neq 0$ for all $t$, then $\frac{\mathrm{d}^k s}{\mathrm{d}t^k}$ exists for all $k \geq 1$, so $s(t)$ is smooth.

(3) If $\|\gamma'(t)\| = 1$ for all $t$, then we have

$$s(t) = \int_{t_0}^{t} 1 \, \mathrm{d}u = t - t_0,$$

so $s = t - t_0$, or equivalently $t = s + t_0$. This means that $\gamma$ can be parametrized by arclength, where $\gamma(t)$ is the point on the curve at a distance of $t$ *on the curve* from $\gamma(t_0)$. Thus, unit speed curves can be parametrized by arclength!

What if we weaken the assumption that the curve is unit speed and consider regular curves? Given a smooth regular parametrized curve $\gamma : (\alpha, \beta) \to \mathbb{R}^n$, we have seen that the arclength function $s$ of $\gamma$ is smooth with $\frac{\mathrm{d}s}{\mathrm{d}t} = \|\gamma'(t)\| > 0$ for all $t \in (\alpha, \beta)$. Therefore, $s$ is strictly increasing on $(\alpha, \beta)$ and its inverse is smooth (by the Inverse Function Theorem). In other words, $s = s(t)$ and $t = t(s)$ (where $t = s^{-1}$) are both smooth functions on their domains, and one can parametrize $s$ in terms of arclength via

$$\gamma(t) = \gamma(t(s)) = \tilde{\gamma}(s),$$

where $\tilde{\gamma} = \gamma \circ t$. So if $\gamma$ is regular, we can assume without loss of generality that $\gamma$ is parametrized by arclength so that

- $\|\gamma'(s)\| = 1$ for all $s$ (i.e. $\gamma$ is unit speed);
- $\gamma'(s) \cdot \gamma''(s) = 0$ for all $s$ (by Proposition 2.2).

In practice, it is usually very difficult to find the arclength parametrization of a regular curve as many functions do not have elementary antiderivatives. Fortunately, we do have general formulae for $T$, $N$, $B$, $\kappa$, and $\tau$, although they are quite cumbersome to compute.

**Proposition 2.14**

Let $\gamma : (\alpha, \beta) \to \mathbb{R}^3$ be a regular curve of class $\mathcal{C}^3$. Then we have

$$T = \frac{\gamma'(t)}{\|\gamma'(t)\|}, \qquad B = \frac{\gamma'(t) \times \gamma''(t)}{\|\gamma'(t) \times \gamma''(t)\|}, \qquad N = B \times T,$$

$$\kappa = \frac{\|\gamma'(t) \times \gamma''(t)\|}{\|\gamma'(t)\|^3}, \qquad \tau = \frac{(\gamma'(t) \times \gamma''(t)) \cdot \gamma'''(t)}{\|\gamma'(t) \times \gamma''(t)\|^2}.$$

**Proof of Proposition 2.14.**

Since $T$ is the *unit* tangent vector of $\gamma$, we have $T = \gamma'(t)/\|\gamma'(t)\|$. Then $\gamma'(t) = \|\gamma'(t)\|T = \frac{\mathrm{d}s}{\mathrm{d}t}T$, which implies that

$$\gamma''(t) = \frac{\mathrm{d}^2 s}{\mathrm{d}t^2}T + \frac{\mathrm{d}s}{\mathrm{d}t}\frac{\mathrm{d}T}{\mathrm{d}t} = \frac{\mathrm{d}^2 s}{\mathrm{d}t^2}T + \left(\frac{\mathrm{d}s}{\mathrm{d}t}\right)^2\frac{\mathrm{d}T}{\mathrm{d}s} = \frac{\mathrm{d}^2 s}{\mathrm{d}t^2}T + \left(\frac{\mathrm{d}s}{\mathrm{d}t}\right)^2\kappa N.$$

Therefore, we obtain

$$\gamma'(t) \times \gamma''(t) = \left(\frac{\mathrm{d}s}{\mathrm{d}t}\right)^3\kappa(T \times N) = \|\gamma'(t)\|^3\kappa B, \tag{$\star$}$$

and thus $\|\gamma'(t) \times \gamma''(t)\| = \|\gamma'(t)\|^3|\kappa|\|B\| = \|\gamma'(t)\|^3|\kappa|$ since $\|B\| = 1$. Rearranging this gives

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

and putting this back into $(\star)$ yields

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

For the proof of the formula for $\tau$, we refer to *Elementary Differential Geometry* by Pressley. $\square$

For example, let $\gamma(t) = (\cos t, 1 - \sin t, -\cos t)$. Then $\gamma'(t) = (-\sin t, -\cos t, \sin t)$ and

$$\|\gamma'(t)\| = \sqrt{1 + \sin^2 t},$$

which we cannot find an arclength parametrization of using elementary means. Nonetheless, we can compute $T$, $N$, $B$, $\kappa$, and $\tau$ using Proposition 2.14. We have $\gamma''(t) = (-\cos t, \sin t, \cos t)$ so that $\gamma'(t) \times \gamma''(t) = (-1, 0, -1)$ and $\|\gamma'(t) \times \gamma''(t)\| = \sqrt{2}$. The third derivative is $\gamma'''(t) = (\sin t, \cos t, -\sin t)$, so

$$(\gamma'(t) \times \gamma''(t)) \cdot \gamma'''(t) = -\sin t + \sin t = 0.$$

Putting everything together, we have

$$\begin{aligned}
T &= \frac{1}{\sqrt{1 + \sin^2 t}}(-\sin t, -\cos t, \sin t), \\
B &= -\frac{1}{\sqrt{2}}(1, 0, 1), \\
\kappa &= \frac{\sqrt{2}}{(1 + \sin^2 t)^{3/2}}, \\
N &= B \times T = -\frac{1}{\sqrt{2(1 + \sin^2 t)}}(\cos t, -2\sin t, \cos t), \\
\tau &= 0.
\end{aligned}$$

In particular, the curvature $\kappa$ is not constant, and $\tau = 0$ so that $\gamma$ is a plane curve by Proposition 2.9.

We do one more example and show that $\gamma(t) = (2\cos t, \sqrt{5}\sin t, -\cos t)$ is a circle. We first observe that $\gamma'(t) = (-2\sin t, \sqrt{5}\cos t, \sin t)$, so $\|\gamma'(t)\| = \sqrt{5} \neq 1$ and $\gamma$ is not unit speed. We have

$$\begin{aligned}
\gamma''(t) &= (-2\cos t, -\sqrt{5}\sin t, \cos t), \\
\gamma'''(t) &= (2\sin t, -\sqrt{5}\cos t, -\sin t),
\end{aligned}$$

so $\gamma'(t)\times\gamma''(t) = (\sqrt{5}, 0, 2\sqrt{5})$ and $\|\gamma'(t)\times\gamma''(t)\| = 5$. This gives

$$\begin{aligned}\kappa &= \frac{\|\gamma'(t)\times\gamma''(t)\|}{\|\gamma'(t)\|^3} = \frac{5}{(\sqrt{5})^3} = \frac{1}{\sqrt{5}} > 0,\\ \tau &= \frac{(\gamma'(t)\times\gamma''(t))\cdot\gamma'''(t)}{\|\gamma'(t)\times\gamma''(t)\|^2} = \frac{0}{25} = 0.\end{aligned}$$

We see that $\gamma$ has constant positive curvature and zero torsion. By the fundamental theorem of space curves (Theorem 2.11), we see that $\gamma$ is a circle. It is contained in the plane with normal vector

$$B = \frac{\gamma'(t)\times\gamma''(t)}{\|\gamma'(t)\times\gamma''(t)\|} = \frac{1}{5}(\sqrt{5}, 0, 2\sqrt{5}) = \left(\frac{1}{\sqrt{5}}, 0, \frac{2}{\sqrt{5}}\right),$$

and this plane is $x + 2z = 0$.

To end this section, we briefly discuss how to generalize this to higher dimensions. Let $\gamma : (\alpha, \beta) \to \mathbb{R}^n$ be a regular curve of class $\mathcal{C}^n$. Suppose that $\{\gamma'(t), \dots, \gamma^{(n-1)}(t)\}$ is a linearly independent set for all $t \in (\alpha, \beta)$. Then $\gamma$ is called a **Frenet curve**.

By applying Gram-Schmidt to this linearly independent set, we obtain an orthonormal basis $\{e_1(t), \dots, e_n(t)\}$ of $\mathbb{R}^n$ for all $t \in (\alpha, \beta)$ as follows. After reparametrizing, we may assume that $\gamma$ is unit speed. We already know that $\|\gamma'(t)\| = 1$, so we set $e_1(t) = \gamma'(t)$ and $e_2(t) = \gamma''(t)/\|\gamma(t)\| = N$. For $j = 3, \dots, n-1$, we set

$$e_j = \frac{\gamma^{(j)} - \sum_{i=1}^{j-1}(\gamma^{(j)}\cdot e_i)e_i}{\|\gamma^{(j)} - \sum_{i=1}^{j-1}(\gamma^{(j)}\cdot e_i)e_i\|}.$$

Then $\{e_1, \dots, e_{n-1}\}$ is an orthonormal set that spans an $(n-1)$-dimensional plane in $\mathbb{R}^n$. Up to a sign, there exists a unique unit vector $e_n$ such that $\{e_1, \dots, e_n\}$ is an orthonormal basis of $\mathbb{R}^n$. We pick $e_n$ such that $\det(e_1, \dots, e_n) = 1$ (such a vector is unique).

**Theorem 2.15**

We have that

$$\begin{bmatrix} e_1' \\ e_2' \\ e_3' \\ \vdots \\ e_n' \end{bmatrix} = \begin{bmatrix} 0 & \kappa_1 & 0 & \cdots & 0 \\ -\kappa_1 & 0 & \kappa_2 & \ddots & 0 \\ 0 & -\kappa_2 & 0 & \ddots & \vdots \\ 0 & 0 & \ddots & \ddots & \kappa_{n-1} \\ 0 & 0 & \cdots & -\kappa_{n-1} & 0 \end{bmatrix} \begin{bmatrix} e_1 \\ e_2 \\ e_3 \\ \vdots \\ e_n \end{bmatrix},$$

where $\kappa_i$ is the ***i*-th Frenet curvature**.

When $n = 3$, we are in the familiar situation where $T = e_1$, $N = e_2$, and $B = e_3$ with $\kappa = \kappa_1$ and $\tau = \kappa_2$.

# 3 Surfaces in $\mathbb{R}^3$

By a surface, we always mean a smooth 2-dimensional submanifold of $\mathbb{R}^3$. In particular, all the charts we use will be smooth.

Let $S$ be a surface, and let $\sigma : U \subset \mathbb{R}^2 \to V = \sigma(U) \subset S \subset \mathbb{R}^3$ be a smooth coordinate charts with components

$$(u, v) \mapsto (x(u, v), y(u, v), z(u, v)).$$

Then we know that $\sigma$ is a homeomorphism onto its image $V$, $\sigma$ is smooth, and $D\sigma$ has rank 2 on $U$. This implies that $\sigma^{-1}$ is smooth as well. If we denote $\sigma_u := \frac{\partial \sigma}{\partial u}$ and $\sigma_v := \frac{\partial \sigma}{\partial v}$, then the derivative matrix of $\sigma$ is

$$D\sigma = \left[\sigma_u \mid \sigma_v\right].$$

Since $D\sigma$ has rank 2, we see that $\{\sigma_u, \sigma_v\}$ is linearly independent on $U$. Equivalently, we have that $\sigma_u \times \sigma_v \neq \mathbf{0}$ on $U$. Therefore, if $p_0 = \sigma(u_0, v_0)$ for some $(u_0, v_0) \in U$, then

$$T_{p_0}S = \operatorname{span}_{\mathbb{R}}\{\sigma_u(u_0, v_0), \sigma_v(u_0, v_0)\}.$$

Moreover, we have $\sigma_u(u_0, v_0) \times \sigma_v(u_0, v_0) \perp T_{p_0}S$.

**Definition 3.1**

For all $p_0 = \sigma(u_0, v_0) \in V = \sigma(U)$, we define the **standard unit normal** to be

$$N_\sigma(u_0, v_0) := \frac{\sigma_u(u_0, v_0) \times \sigma_v(u_0, v_0)}{\|\sigma_u(u_0, v_0) \times \sigma_v(u_0, v_0)\|}.$$

Since $S \subset \mathbb{R}^3$ and $T_{p_0}S$ is a 2-dimensional affine subspace of $T_{p_0}\mathbb{R}^3 \simeq \mathbb{R}^3$, there are only two possible *unit* normal directions to $T_{p_0}S$ at $p_0$. Therefore, $N_\sigma(u_0, v_0)$ is uniquely determined by $S$ up to a sign.

For example, consider the paraboloid $S = \{(x, y, z) \in \mathbb{R}^3 : z = x^2 + y^2\} \subset \mathbb{R}^3$. This is the graph of the smooth function $f(x, y) = x^2 + y^2$, so we have a smooth chart

$$\begin{aligned} \sigma : \mathbb{R}^2 &\to S \subset \mathbb{R}^3 \\ (u, v) &\mapsto (u, v, u^2 + v^2). \end{aligned}$$

Then $\sigma_u = (1, 0, 2u)$ and $\sigma_v = (0, 1, 2v)$, which gives

$$D\sigma = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 2u & 2v \end{bmatrix}$$

and $\sigma_u \times \sigma_v = (-2u, -2v, 1) \neq \mathbf{0}$. At the point $p_0 = \sigma(u_0, v_0) = (u_0, v_0, u_0^2 + v_0^2)$, we have

$$T_{p_0}S = \operatorname{span}_{\mathbb{R}}\{(1, 0, 2u_0), (0, 1, 2v_0)\}.$$

For example, with $p_0 = \mathbf{0} = (0, 0, 0) = \sigma(0, 0)$, we have

$$T_{\mathbf{0}}S = \operatorname{span}_{\mathbb{R}}\{(1, 0, 0), (0, 1, 0)\},$$

which is the $xy$-plane. Moreover, since

$$\|\sigma_u \times \sigma_v\| = \sqrt{4u^2 + 4v^2 + 1} \neq 0,$$

we see that the standard unit normal is

$$N_\sigma = \frac{1}{\sqrt{4u^2 + 4v^2 + 1}}(-2u, -2v, 1).$$

We have that $N_\sigma(0, 0) = (0, 0, 1)$, which is perpendicular to $(1, 0, 0)$ and $(0, 1, 0)$ as expected.

What happens when we change the coordinate chart? Suppose that $\tilde{\sigma} : \tilde{U} \to \tilde{V}$ is another coordinate chart with $V \cap \tilde{V} \neq \varnothing$. Then we can write

$$\tilde{\sigma} = \sigma \circ (\sigma^{-1} \circ \tilde{\sigma}) = \sigma \circ \Phi,$$

where $\Phi := \sigma^{-1} \circ \tilde{\sigma} : \sigma^{-1}(V \cap \tilde{V}) \subset \mathbb{R}^2 \to \sigma^{-1}(V \cap \tilde{V}) \subset \mathbb{R}^2$ is a smooth diffeomorphism with components

$$(\tilde{u}, \tilde{v}) \mapsto (u(\tilde{u}, \tilde{v}), v(\tilde{u}, \tilde{v})).$$

By the chain rule, we obtain $D\tilde{\sigma} = D\sigma D\Phi$, or more explicitly

$$\left[\tilde{\sigma}_{\tilde{u}} \mid \tilde{\sigma}_{\tilde{v}}\right] = \left[\sigma_u \mid \sigma_v\right] D\Phi.$$

Here, $D\Phi$ is the change of basis matrix of $T_pS$. Note that

$$D\Phi = \begin{bmatrix} \partial u/\partial \tilde{u} & \partial u/\partial \tilde{v} \\ \partial v/\partial \tilde{u} & \partial v/\partial \tilde{v} \end{bmatrix}.$$

Moreover, we have

$$\begin{aligned} \tilde{\sigma}_{\tilde{u}} &= \sigma_u \cdot \frac{\partial u}{\partial \tilde{u}} + \sigma_v \cdot \frac{\partial v}{\partial \tilde{u}}, \\ \tilde{\sigma}_{\tilde{v}} &= \sigma_u \cdot \frac{\partial u}{\partial \tilde{v}} + \sigma_v \cdot \frac{\partial v}{\partial \tilde{v}}, \end{aligned}$$

which implies that

$$\begin{aligned} \tilde{\sigma}_{\tilde{u}} \times \tilde{\sigma}_{\tilde{v}} &= \left(\sigma_u \cdot \frac{\partial u}{\partial \tilde{u}} + \sigma_v \cdot \frac{\partial v}{\partial \tilde{u}}\right) \times \left(\sigma_u \cdot \frac{\partial u}{\partial \tilde{v}} + \sigma_v \cdot \frac{\partial v}{\partial \tilde{v}}\right) \\ &= (\sigma_u \times \sigma_u)\left(\frac{\partial u}{\partial \tilde{u}}\frac{\partial u}{\partial \tilde{v}}\right) + (\sigma_u \times \sigma_v)\left(\frac{\partial u}{\partial \tilde{u}}\frac{\partial v}{\partial \tilde{v}} - \frac{\partial v}{\partial \tilde{u}}\frac{\partial u}{\partial \tilde{v}}\right) + (\sigma_v \times \sigma_v)\left(\frac{\partial v}{\partial \tilde{u}}\frac{\partial v}{\partial \tilde{v}}\right) \\ &= \left(\frac{\partial u}{\partial \tilde{u}}\frac{\partial v}{\partial \tilde{v}} - \frac{\partial v}{\partial \tilde{u}}\frac{\partial u}{\partial \tilde{v}}\right)\sigma_u \times \sigma_v = (\det D\Phi)\sigma_u \times \sigma_v, \end{aligned}$$

where $\sigma_u \times \sigma_u = \sigma_v \times \sigma_v = \mathbf{0}$. It follows that

$$N_{\tilde{\sigma}} = \frac{\tilde{\sigma}_{\tilde{u}} \times \tilde{\sigma}_{\tilde{v}}}{\|\tilde{\sigma}_{\tilde{u}} \times \tilde{\sigma}_{\tilde{v}}\|} = \frac{\det D\Phi}{|\det D\Phi|} \cdot \frac{\sigma_u \times \sigma_v}{\|\sigma_u \times \sigma_v\|} = \pm N_\sigma,$$

where $\pm$ is given by the sign of $\det D\Phi$. Therefore, changing the coordinate chart only changes the standard unit normal up to a sign.

## 3.1 First fundamental form

Let $p_0 \in \sigma(U) \subset S$. If $p_0 = \sigma(u_0, v_0)$, then

$$T_{p_0}S = \operatorname{span}_{\mathbb{R}}\{\sigma_u(u_0, v_0), \sigma_v(u_0, v_0)\}$$

is a 2-dimensional subspace of $T_{p_0}\mathbb{R}^3 \simeq \mathbb{R}^3$. Suppose that the inner product on $T_{p_0}\mathbb{R}^3$ is the usual dot product. That is, for $X, Y \in T_{p_0}\mathbb{R}^3 \simeq \mathbb{R}^3$, we have

$$X \cdot Y := X^T Y = X^T I Y,$$

where the latter is the matrix representation with respect to the standard basis.

What if we restrict the dot product to $T_{p_0}S$? Since $T_{p_0}S$ is 2-dimensional, the restriction of the dot product can be represented by a $2 \times 2$ matrix with respect to a basis of $T_{p_0}S$.

Using the basis $\mathcal{B} = \{\sigma_u(u_0, v_0), \sigma_v(u_0, v_0)\}$, then for all $X, Y \in T_{p_0}S$, we have

$$\begin{aligned} X &= a\sigma_u(u_0, v_0) + b\sigma_v(u_0, v_0), \\ Y &= c\sigma_u(u_0, v_0) + d\sigma_v(u_0, v_0) \end{aligned}$$

for some $a, b, c, d \in \mathbb{R}$. More concisely, we have

$$\begin{aligned} X &= D\sigma(u_0, v_0) \begin{bmatrix} a \\ b \end{bmatrix}, \\ Y &= D\sigma(u_0, v_0) \begin{bmatrix} c \\ d \end{bmatrix}, \end{aligned}$$

which gives us

$$\begin{aligned} X \cdot Y = X^T Y &= \left( D\sigma(u_0, v_0) \begin{bmatrix} a \\ b \end{bmatrix} \right)^T \left( D\sigma(u_0, v_0) \begin{bmatrix} c \\ d \end{bmatrix} \right) \\ &= \begin{bmatrix} a & b \end{bmatrix} D\sigma(u_0, v_0)^T D\sigma(u_0, v_0) \begin{bmatrix} c \\ d \end{bmatrix} \\ &= \begin{bmatrix} a & b \end{bmatrix} \begin{bmatrix} \sigma_u \cdot \sigma_u & \sigma_u \cdot \sigma_v \\ \sigma_v \cdot \sigma_u & \sigma_v \cdot \sigma_v \end{bmatrix} \begin{bmatrix} c \\ d \end{bmatrix}. \end{aligned}$$

The matrix $D\sigma(u_0, v_0)^T D\sigma(u_0, v_0)$ above is our matrix representation of the dot product with respect to the basis $\mathcal{B} = \{\sigma_u(u_0, v_0), \sigma_v(u_0, v_0)\}$.

**Definition 3.2**

The **first fundamental form of** $\sigma$ is defined to be

$$\mathcal{F}_{\mathrm{I}} = D\sigma(u_0, v_0)^T D\sigma(u_0, v_0) = \begin{bmatrix} \sigma_u \cdot \sigma_u & \sigma_u \cdot \sigma_v \\ \sigma_v \cdot \sigma_u & \sigma_v \cdot \sigma_v \end{bmatrix}.$$

Note that $\mathcal{F}_{\mathrm{I}}$ is symmetric since $\sigma_u \cdot \sigma_v = \sigma_v \cdot \sigma_u$. Moreover, if $\tilde{\sigma} = \sigma \circ \Phi$ is another chart, then

$$D\tilde{\sigma} = D\sigma D\Phi$$

where $D\Phi$ is a change of basis matrix, with

$$\begin{aligned} \tilde{\mathcal{F}_{\mathrm{I}}} &= (D\tilde{\sigma})^T D\tilde{\sigma} \\ &= (D\sigma D\Phi)^T (D\sigma D\Phi) \\ &= D\Phi^T (D\sigma^T D\sigma) D\Phi \\ &= D\Phi^T \, \mathcal{F}_{\mathrm{I}} \, D\Phi. \end{aligned}$$

Therefore, $\tilde{\mathcal{F}_{\mathrm{I}}}$ is similar to $\mathcal{F}_{\mathrm{I}}$, and the first fundamental form depends on the coordinate chart $\sigma$.

Let's compute some examples.

(1) **(Plane.)** Let $\sigma(u, v) = p_0 + uw_1 + vw_2$ for $u, v \in \mathbb{R}$, where $p_0$ is a point and $w_1$ and $w_2$ are direction vectors. We can choose $w_1$ and $w_2$ such that $\{w_1, w_2\}$ is orthonormal. Then $\sigma_u = w_1$ and $\sigma_v = w_2$, so

$$\mathcal{F}_{\mathrm{I}} = \begin{bmatrix} w_1 \cdot w_1 & w_1 \cdot w_2 \\ w_2 \cdot w_1 & w_2 \cdot w_2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

(2) **(Cylinder.)** Consider the cylinder $x^2 + y^2 = 1$ with coordinate chart $\tilde{\sigma}(u, v) = (\cos u, \sin u, v)$ where $(u, v) \in (0, 2\pi) \times \mathbb{R}$. We have $\tilde{\sigma}_u = (-\sin u, \cos u, 0)$ and $\tilde{\sigma}_v = (0, 0, 1)$, which yields $\sigma_u \cdot \sigma_u = \sigma_v \cdot \sigma_v = 1$ and $\sigma_u \cdot \sigma_v = 0$. In particular, we have that

$$\tilde{\mathcal{F}_{\mathrm{I}}} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

This is the same first fundamental form as the plane! Here, we see that the first fundamental form alone cannot distinguish between different shapes.

**Remark.** This is not too surprising since the cylinder can be obtained from the plane by folding it smoothly (without stretching or shrinking) so that length and distances are preserved. The first fundamental form is an inner product that is the restriction of the dot product, which is used to measure lengths and distances.

(3) **(Sphere.)** Consider the sphere $x^2 + y^2 + z^2 = a^2$ for some $a > 0$ and the coordinate chart

$$\sigma(\theta, \varphi) = (a\cos\theta\sin\varphi, a\sin\theta\sin\varphi, a\cos\varphi)$$

for $(\theta, \varphi) \in (0, 2\pi) \times (0, \pi)$. We have

$$\begin{aligned}\sigma_\theta &= (-a\sin\theta\sin\varphi, a\cos\theta\sin\varphi, 0),\\ \sigma_\varphi &= (a\cos\theta\cos\varphi, a\sin\theta\cos\varphi, -a\sin\varphi),\end{aligned}$$

which gives us

$$\begin{aligned}\sigma_\theta \cdot \sigma_\theta &= a^2\sin^2\varphi,\\ \sigma_\theta \cdot \sigma_\varphi &= 0,\\ \sigma_\varphi \cdot \sigma_\varphi &= a^2.\end{aligned}$$

Therefore, the first fundamental form of $\sigma$ is

$$\mathcal{F}_{\mathrm{I}} = \begin{bmatrix} a^2\sin^2\varphi & 0 \\ 0 & a^2 \end{bmatrix},$$

which is different from the first fundamental form of the plane.

**Remark.** The sphere can be obtained from the plane and adding a point at $\infty$.

## 3.2 Second fundamental form

In the above examples, we saw that the first fundamental form $\mathcal{F}_{\mathrm{I}}$ is not enough to distinguish surfaces and capture curvature since seemingly different surfaces such as planes and cylinders can have the same first fundamental form. Therefore, we will introduce a "second fundamental form".

Let $w_1, w_2 \in \mathbb{R}^3$. Recall that the **scalar projection of** $w_1$ **onto** $w_2$ is

$$\ell = \frac{w_1 \cdot w_2}{\|w_2\|}$$

because if $\theta$ is the angle between $w_1$ and $w_2$, then $\cos\theta = \ell/\|w_1\|$ so that

$$\ell = \cos\theta\|w_1\| = \frac{w_1 \cdot w_2}{\|w_1\|\|w_2\|}\|w_1\| = \frac{w_1 \cdot w_2}{\|w_2\|}.$$

In particular, if $\|w_2\| = 1$, then $\ell = w_1 \cdot w_2$.

Define the vector $\Delta\sigma = \sigma(u_0 + \Delta u, v_0 + \Delta v) - \sigma(u_0, v_0) \in \mathbb{R}^3$. Then $\Delta\sigma \cdot N_\sigma$ is the scalar projection of $\Delta\sigma$ onto $N_\sigma$ (since $N_\sigma$ is unit), and we call it the **deviation**. This measures how much $S$ moves away from the tangent plane $T_{p_0}S$ to $S$ at $p_0 = \sigma(u_0, v_0)$.

By Taylor's Theorem, we have

$$\begin{aligned}\Delta\sigma &= \sigma(u_0 + \Delta u, v_0 + \Delta v)\\ &= \sigma_u\Delta u + \sigma_v\Delta v + \frac{1}{2}(\sigma_{uu}\Delta u^2 + 2\sigma_{uv}\Delta u\Delta v + \sigma_{vv}\Delta v^2) + \text{remainder},\end{aligned}$$

where we used the fact that $\sigma_{uv} = \sigma_{vu}$ since $\sigma$ is smooth. Since $T_pS \perp N_\sigma$ and $T_pS = \operatorname{span}_{\mathbb{R}}\{\sigma_u, \sigma_v\}$, we have $\sigma_u \cdot N_\sigma = \sigma_v \cdot N_\sigma = 0$. This yields

$$\Delta\sigma \cdot N_\sigma = (\sigma_u \cdot N_\sigma)\Delta u + (\sigma_v \cdot N_\sigma)\Delta v + \frac{1}{2}[(\sigma_{uu} \cdot N_\sigma)\Delta u^2 + 2(\sigma_{uv} \cdot N_\sigma)\Delta u\Delta v + (\sigma_{vv} \cdot N_\sigma)\Delta v^2] + \cdots$$

and so the deviation is well approximated with

$$\begin{aligned}\Delta\sigma \cdot N_\sigma &\approx \frac{1}{2}[(\sigma_{uu} \cdot N_\sigma)\Delta u^2 + 2(\sigma_{uv} \cdot N_\sigma)\Delta u\Delta v + (\sigma_{vv} \cdot N_\sigma)\Delta v^2] \\ &= \frac{1}{2}\begin{bmatrix}\Delta u & \Delta v\end{bmatrix}\begin{bmatrix}\sigma_{uu} \cdot N_\sigma & \sigma_{uv} \cdot N_\sigma \\ \sigma_{vu} \cdot N_\sigma & \sigma_{vv} \cdot N_\sigma\end{bmatrix}\begin{bmatrix}\Delta u \\ \Delta v\end{bmatrix}.\end{aligned}$$

This matrix is reminiscent of the Hessian from multivariable calculus!

**Definition 3.3**

We define the **second fundamental form of** $\sigma$ to be

$$\mathcal{F}_{\mathrm{II}} := \begin{bmatrix}\sigma_{uu} \cdot N_\sigma & \sigma_{uv} \cdot N_\sigma \\ \sigma_{vu} \cdot N_\sigma & \sigma_{vv} \cdot N_\sigma\end{bmatrix}.$$

Note that $\sigma$ is smooth, so $\sigma_{uv} = \sigma_{vu}$ and $\mathcal{F}_{\mathrm{II}}$ is symmetric. We look at some examples.

(1) **(Plane.)** As before, let $\sigma(u, v) = p_0 + uw_1 + vw_2$ for $u, v \in \mathbb{R}$ where $p_0$ is a point and $\{w_1, w_2\}$ is orthonormal. We have $\sigma_u = w_1$ and $\sigma_v = w_2$. Then $\sigma_{uu} = \sigma_{uv} = \sigma_{vv} = 0$ so that

$$\mathcal{F}_{\mathrm{II}} = \begin{bmatrix}0 & 0 \\ 0 & 0\end{bmatrix}.$$

(2) **(Cylinder.)** Let $\tilde{\sigma}(u, v) = (\cos u, \sin u, v)$ for $u, v \in \mathbb{R}$, and note that $\tilde{\sigma}_u = (-\sin u, \cos u, 0)$ and $\tilde{\sigma}_v = (0, 0, 1)$. Then $\tilde{\sigma}_u \times \tilde{\sigma}_v = (\cos u, \sin u, 0)$ with $\|\tilde{\sigma}_u \times \tilde{\sigma}_v\| = 1$, so we have $N_{\tilde{\sigma}} = (\cos u, \sin u, 0)$. Moreover, we have $\tilde{\sigma}_{uu} = (-\cos u, -\sin u, 0)$ and $\tilde{\sigma}_{uv} = \tilde{\sigma}_{vv} = 0$, which implies that $\sigma_{uu} \cdot N_\sigma = -\cos^2 u - \sin^2 u = -1$ and the second fundamental form of $\tilde{\sigma}$ is

$$\tilde{\mathcal{F}}_{\mathrm{II}} = \begin{bmatrix}-1 & 0 \\ 0 & 0\end{bmatrix}.$$

We see that this is different from the second fundamental form of the plane.

(3) **(Sphere.)** Consider the sphere $x^2 + y^2 + z^2 = a^2$ where $a > 0$ with coordinate chart

$$\sigma(\theta, \varphi) = (a\cos\theta\sin\varphi, a\sin\theta\sin\varphi, a\cos\varphi)$$

for $(\theta, \varphi) \in (0, 2\pi) \times (0, \pi)$. A direct computation shows that

$$\mathcal{F}_{\mathrm{II}} = \begin{bmatrix}a\sin^2\varphi & 0 \\ 0 & a\end{bmatrix}.$$

(4) **(Graph of a smooth function.)** Let $f : U \subset \mathbb{R}^2 \to \mathbb{R}$ be smooth where $U \subset \mathbb{R}^2$ is open. The graph

$$S = \{(x, y, z) \in \mathbb{R}^3 : z = f(x, y)\} \subset \mathbb{R}^3$$

of $f$ is a surface with smooth coordinate chart

$$\begin{aligned}\sigma : U \subset \mathbb{R}^2 &\to S \subset \mathbb{R}^3 \\ (u, v) &\mapsto (u, v, f(u, v)).\end{aligned}$$

We have $\sigma_u = (1, 0, f_u)$ and $\sigma_v = (0, 1, f_v)$ where $f_u = \frac{\partial f}{\partial u}$ and $f_v = \frac{\partial f}{\partial v}$, which implies that

$$\mathcal{F}_{\mathrm{I}} = \begin{bmatrix} 1 + (f_u)^2 & f_u f_v \\ f_v f_u & 1 + (f_v)^2 \end{bmatrix}.$$

Next, we have $\sigma_u \times \sigma_v = (-f_u, -f_v, 1)$, so

$$N_\sigma = \frac{1}{\sqrt{1 + (f_u)^2 + (f_v)^2}}(-f_u, -f_v, 1).$$

Moreover, we know that $\sigma_{uu} = (0, 0, f_{uu})$, $\sigma_{uv} = (0, 0, f_{uv})$, and $\sigma_{vv} = (0, 0, f_{vv})$, so

$$\mathcal{F}_{\mathrm{II}} = \frac{1}{\sqrt{1 + (f_u)^2 + (f_v)^2}} \begin{bmatrix} f_{uu} & f_{uv} \\ f_{vu} & f_{vv} \end{bmatrix}.$$

In particular, the above matrix is exactly the Hessian $H(f)$ of $f$, which appears in the second derivative test. At a critical point of $f(x, y)$, we have $f_u = f_v = 0$ so that $\mathcal{F}_{\mathrm{I}} = I_{2\times 2}$ and $\mathcal{F}_{\mathrm{II}} = H(f)$.

## 3.3 The shape operator

We begin with the following observation.

**Lemma 3.4**

We have $\det \mathcal{F}_{\mathrm{I}} = \|\sigma_u \times \sigma_v\|^2$.

**Proof of Lemma 3.4.**

This follows directly from the cross product identity

$$(a \times b) \cdot (c \times d) = (a \cdot c)(b \cdot d) - (a \cdot d)(b \cdot c)$$

for all $a, b, c, d \in \mathbb{R}^3$. We have that

$$\begin{aligned} \|\sigma_u \cdot \sigma_v\|^2 &= (\sigma_u \times \sigma_v) \cdot (\sigma_u \times \sigma_v) \\ &= (\sigma_u \cdot \sigma_u)(\sigma_v \cdot \sigma_v) - (\sigma_u \cdot \sigma_v)(\sigma_v \cdot \sigma_u) \\ &= \det \mathcal{F}_{\mathrm{I}}\,. \end{aligned}$$ $\square$

Since $\sigma_u \times \sigma_v \neq \mathbf{0}$ everywhere for any coordinate chart $\sigma$, this means that $\mathcal{F}_{\mathrm{I}}$ is always invertible.

**Definition 3.5**

The **shape operator** (or **Weingarten matrix**) is defined to be

$$\mathcal{W} := \mathcal{F}_{\mathrm{I}}^{-1}\,\mathcal{F}_{\mathrm{II}}\,.$$

The **Gaussian curvature** is $K := \det \mathcal{W}$, and the **mean curvature** is $H := \frac{1}{2}\operatorname{tr} \mathcal{W}$.

Let $p_0 = \sigma(u_0, v_0)$ for some $(u_0, v_0) \in U$. Then:

(i) $p_0$ is **elliptic** if $K > 0$.

(ii) $p_0$ is **hyperbolic** if $K < 0$.

(iii) $p_0$ is **parabolic** if $K = 0$ and $H \neq 0$.

(iv) $p_0$ is **planar** if $K = H = 0$.

We compute some examples of the shape operator.

(1) **(Plane.)** Let $\sigma(u,v) = p_0 + uw_1 + vw_2$ where $\{w_1, w_2\}$ is orthonormal and $p_0$ is a point. We found that

$$\mathcal{F}_{\mathrm{I}} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \qquad \mathcal{F}_{\mathrm{II}} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix},$$

which implies that the shape operator is

$$\mathcal{W} = \mathcal{F}_{\mathrm{I}}^{-1} \mathcal{F}_{\mathrm{II}} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}.$$

We have that $K = \det \mathcal{W} = 0$ and $H = \frac{1}{2} \operatorname{tr} \mathcal{W} = 0$ everywhere, so every point on the plane is planar!

(2) **(Cylinder.)** For the cylinder $x^2 + y^2 = 1$ with smooth coordinate chart $\sigma(u,v) = (\cos u, \sin u, v)$ where $(u,v) \in (0, 2\pi) \times \mathbb{R}$, we saw that

$$\mathcal{F}_{\mathrm{I}} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \qquad \mathcal{F}_{\mathrm{II}} = \begin{bmatrix} -1 & 0 \\ 0 & 0 \end{bmatrix}.$$

Then the shape operator is given by

$$\mathcal{W} = \mathcal{F}_{\mathrm{I}}^{-1} \mathcal{F}_{\mathrm{II}} = \begin{bmatrix} -1 & 0 \\ 0 & 0 \end{bmatrix},$$

so $K = \det \mathcal{W} = 0$ and $H = \frac{1}{2} \operatorname{tr} \mathcal{W} = -\frac{1}{2} \neq 0$. Thus, every point on the cylinder is parabolic.

The missing points on the cylinder are covered by the chart $\sigma(u,v) = (\cos u, \sin u, v)$ over $(u,v) \in (-\pi, \pi) \times \mathbb{R}$, which has the same first and second fundamental forms and thus the same shape operator.

**Remark.** Cylinders locally look like parabolic cylinders. For example, the set

$$S = \{(x,y,z) \in \mathbb{R}^3 : z = x^2\} \subset \mathbb{R}^3$$

is a parabolic cylinder. Since $S$ is the graph of the smooth scalar function $f : \mathbb{R}^2 \to \mathbb{R}$ defined by $(x,y) \mapsto x^2$, we see that $S$ is a surface which can be covered by the coordinate chart $\sigma(u,v) = (u,v,u^2)$ for $(u,v) \in \mathbb{R}^2$. Then $\sigma_u = (1,0,2u)$ and $\sigma_v = (0,1,0)$ so that

$$\mathcal{F}_{\mathrm{I}} = \begin{bmatrix} 1 + 4u^2 & 0 \\ 0 & 1 \end{bmatrix}.$$

Moreover, we have $\sigma_{uu} = (0,0,2)$, $\sigma_{uv} = \sigma_{vv} = (0,0,0)$ and

$$N_\sigma = \frac{\sigma_u \times \sigma_v}{\|\sigma_u \times \sigma_v\|} = \frac{1}{\sqrt{1+4u^2}}(-2u, 0, 1),$$

so the second fundamental form is

$$\mathcal{F}_{\mathrm{II}} = \begin{bmatrix} 2/\sqrt{1+4u^2} & 0 \\ 0 & 0 \end{bmatrix}.$$

It follows that the shape operator is

$$\mathcal{W} = \mathcal{F}_{\mathrm{I}}^{-1} \mathcal{F}_{\mathrm{II}} = \begin{bmatrix} 1/(1+4u^2) & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 2/\sqrt{1+4u^2} & 0 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} 2/(1+4u^2)^{3/2} & 0 \\ 0 & 0 \end{bmatrix}.$$

We obtain $K = \det \mathcal{W} = 0$ and $H = \frac{1}{2} \operatorname{tr} \mathcal{W} = 1/(1+4u^2)^{3/2} \neq 0$, so every point on $S$ is parabolic.

(3) **(Sphere.)** Consider the sphere $x^2 + y^2 + z^2 = a^2$ where $a > 0$ with smooth coordinate chart

$$\sigma(\theta, \varphi) = (a\cos\theta\sin\varphi, a\sin\theta\sin\varphi, a\cos\varphi)$$

for $(\theta, \varphi) \in (0, 2\pi) \times (-\pi, \pi)$. We have seen that

$$\mathcal{F}_{\mathrm{I}} = \begin{bmatrix} a^2\sin^2\varphi & 0 \\ 0 & a^2 \end{bmatrix}, \qquad \mathcal{F}_{\mathrm{II}} = \begin{bmatrix} a\sin^2\varphi & 0 \\ 0 & a \end{bmatrix}.$$

Then the shape operator is given by

$$\mathcal{W} = \mathcal{F}_{\mathrm{I}}^{-1}\,\mathcal{F}_{\mathrm{II}} = \begin{bmatrix} 1/a & 0 \\ 0 & 1/a \end{bmatrix},$$

so $K = \det\mathcal{W} = 1/a^2$ and $H = \frac{1}{2}\operatorname{tr}\mathcal{W} = 1/a$ everywhere. Thus, every point on the sphere is elliptic.

**Remark.** By using different spherical coordinates, we can cover the entire sphere. In each case, we have that $K = 1/a^2$ and $H = \pm 1/a$.

(4) **(Saddle surface.)** Consider the surface $z = x^2 - y^2$ where $x, y \in \mathbb{R}$. This is the graph of the smooth function $f(x, y) = x^2 - y^2$, so we obtain the smooth coordinate chart $\sigma(u, v) = (u, v, u^2 - v^2)$ where $(u, v) \in \mathbb{R}^2$. A direct computation shows that

$$\mathcal{F}_{\mathrm{I}} = \begin{bmatrix} 1 + 4u^2 & -4uv \\ -4uv & 1 + 4v^2 \end{bmatrix}, \qquad \mathcal{F}_{\mathrm{II}} = \frac{1}{\sqrt{1 + 4u^2 + 4v^2}} \begin{bmatrix} 2 & 0 \\ 0 & -2 \end{bmatrix}.$$

Note that to find the Gaussian curvature $K = \det\mathcal{W}$, we need not compute $\mathcal{W}$ explicitly. Recalling that the determinant is multiplicative and $\det(A^{-1}) = (\det A)^{-1}$ for a matrix $A$, we have

$$K = \det\mathcal{W} = \det(\mathcal{F}_{\mathrm{I}}^{-1}\,\mathcal{F}_{\mathrm{II}}) = \det(\mathcal{F}_{\mathrm{I}}^{-1})\det\mathcal{F}_{\mathrm{II}} = (\det\mathcal{F}_{\mathrm{I}})^{-1}\det\mathcal{F}_{\mathrm{II}}.$$

We see that $\det\mathcal{F}_{\mathrm{I}} = (1 + 4u^2)(1 + 4v^2) - (-4uv)^2 = 1 + 4u^2 + 4v^2$ and

$$\det\mathcal{F}_{\mathrm{II}} = \frac{1}{(\sqrt{1 + 4u^2 + 4v^2})^2}(2(-2) - 0^2) = -\frac{4}{1 + 4u^2 + 4v^2},$$

which implies that the Gaussian curvature is

$$K = -\frac{4}{(1 + 4u^2 + 4v^2)^2} < 0$$

for all $(u, v) \in \mathbb{R}^2$. Therefore, every point on the saddle surface is hyperbolic.

Let's describe a few more properties of the shape operator. First off, note that although $\mathcal{F}_{\mathrm{I}}$ and $\mathcal{F}_{\mathrm{II}}$ are always symmetric, it is possible that $\mathcal{W}$ is not symmetric.

For example, consider the saddle surface $z = xy$ parametrized by $\sigma(u, v) = (u, v, uv)$ where $u, v \in \mathbb{R}$. A direct computation then gives

$$\mathcal{F}_{\mathrm{I}} = \begin{bmatrix} 1 + v^2 & uv \\ uv & 1 + u^2 \end{bmatrix}, \qquad \mathcal{F}_{\mathrm{II}} = \frac{1}{(1 + u^2 + v^2)^{3/2}} \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

so that the shape operator is

$$\mathcal{W} = \mathcal{F}_{\mathrm{I}}^{-1}\,\mathcal{F}_{\mathrm{II}} = \frac{1}{(1 + u^2 + v^2)^{3/2}} \begin{bmatrix} uv & 1 + v^2 \\ 1 + u^2 & uv \end{bmatrix}.$$

In particular, we have $1 + v^2 \neq 1 + u^2$ away from points where $u = \pm v$, so $\mathcal{W}$ is not symmetric.

Nonetheless, we have the following result.

**Proposition 3.6**

The shape operator $\mathcal{W}$ is a diagonalizable matrix with real eigenvalues.

**Proof of Proposition 3.6.**

Let $w_1$ and $w_2$ be two orthogonal unit vectors in $T_pS$ so that $w_1 \cdot w_1 = w_2 \cdot w_2 = 1$ and $w_1 \cdot w_2 = 0$. Since $w_1, w_2 \in T_pS = \operatorname{span}_{\mathbb{R}}\{\sigma_u, \sigma_v\}$, we can write

$$\begin{aligned} w_1 &= a_1\sigma_u + b_1\sigma_v, \\ w_2 &= a_2\sigma_u + b_2\sigma_v \end{aligned}$$

for some $a_1, a_2, b_1, b_2 \in \mathbb{R}$. For $i, j \in \{1, 2\}$, observe that

$$w_i \cdot w_j = \begin{bmatrix} a_i & b_i \end{bmatrix} \mathcal{F}_{\mathrm{I}} \begin{bmatrix} a_j \\ b_j \end{bmatrix} = \delta_{ij},$$

where $\delta_{ij}$ is the Kronecker delta. Setting

$$C := \begin{bmatrix} w_1 \mid w_2 \end{bmatrix} = \begin{bmatrix} a_1 & a_2 \\ b_1 & b_2 \end{bmatrix},$$

we have that

$$C^T \mathcal{F}_{\mathrm{I}} C = \begin{bmatrix} a_1 & b_1 \\ a_2 & b_2 \end{bmatrix} \mathcal{F}_{\mathrm{I}} \begin{bmatrix} a_1 & a_2 \\ b_1 & b_2 \end{bmatrix} = \begin{bmatrix} w_1 \cdot w_1 & w_1 \cdot w_2 \\ w_2 \cdot w_1 & w_2 \cdot w_2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = I_{2\times 2}.$$

Note that the columns of $C$ are linearly independent, so $C$ is invertible. Therefore, we obtain

$$C^{-1}\mathcal{F}_{\mathrm{I}}^{-1}(C^T)^{-1} = (C^T \mathcal{F}_{\mathrm{I}} C)^{-1} = (I_{2\times 2})^{-1} = I_{2\times 2},$$

which implies that

$$\begin{aligned} C^{-1}\mathcal{W}C &= C^{-1}(\mathcal{F}_{\mathrm{I}}^{-1} \mathcal{F}_{\mathrm{II}})C \\ &= C^{-1}\mathcal{F}_{\mathrm{I}}^{-1}(C^T)^{-1}C^T \mathcal{F}_{\mathrm{II}} C \\ &= C^T \mathcal{F}_{\mathrm{II}} C =: B. \end{aligned}$$

This means that $\mathcal{W}$ is similar to $B = C^T \mathcal{F}_{\mathrm{II}} C$. But $B$ is symmetric since

$$B^T = (C^T \mathcal{F}_{\mathrm{II}} C)^T = C^T \mathcal{F}_{\mathrm{II}}^T (C^T)^T = C^T \mathcal{F}_{\mathrm{II}} C = B.$$

Then $B$ is diagonalizable with real eigenvalues. Since $\mathcal{W}$ is similar to $B$, it follows that $\mathcal{W}$ is also diagonalizable with real eigenvalues. $\square$

Due to Proposition 3.6, it makes sense to define the following.

**Definition 3.7**

The eigenvalues $\kappa_1$ and $\kappa_2$ of $\mathcal{W}$ are called the **principal curvatures** of the coordinate chart $\sigma$.

Observe that we have $K = \det \mathcal{W} = \kappa_1\kappa_2$ and $H = \frac{1}{2}\operatorname{tr}\mathcal{W} = \frac{1}{2}(\kappa_1 + \kappa_2)$.

Therefore, given the principal curvatures $\kappa_1$ and $\kappa_2$ and a point $p_0 = \sigma(u_0, v_0)$, we see that:

(i) $p_0$ is elliptic when $K > 0$, which happens if and only if $\kappa_1$ and $\kappa_2$ are nonzero with the same sign.

(ii) $p_0$ is hyperbolic when $K < 0$, which happens if and only if $\kappa_1$ and $\kappa_2$ are nonzero with opposite signs.

(iii) $p_0$ is parabolic when $K = 0$ and $H \neq 0$, which happens if and only if $\kappa_1 = 0$ and $\kappa_2 \neq 0$ (or vice versa).

(iv) $p_0$ is planar when $K = H = 0$, which happens if and only if $\kappa_1 = \kappa_2 = 0$.

**Remark.** We know that $S$ is locally the graph of a function. After possibly rotating and translating $S$, we can assume that it is the graph of a function $f(x, y)$ near $p_0$ and that $p_0$ corresponds to a critical point of $f(x, y)$. We have seen that at $p_0$, we have $\mathcal{F}_{\mathrm{I}} = I_{2\times 2}$ and $\mathcal{F}_{\mathrm{II}} = H(f)$ so that $\mathcal{W} = H(f)$. We can now apply the second derivative test.

- If $\det H(f) > 0$ and $f_{xx} > 0$, then $p_0$ is a local minimum.
- If $\det H(f) > 0$ and $f_{xx} < 0$, then $p_0$ is a local maximum.
- If $\det H(f) < 0$, then $p_0$ is a saddle point.
- If $\det H(f) = 0$, then the test is inconclusive.

The first two situations correspond to elliptic points. The third corresponds to hyperbolic points, and the fourth corresponds to parabolic or planar points.

## 3.4 Normal and geodesic curvatures

We now study in more detail the curvature of regular curves or surfaces. Since any regular curve can be reparametrized using arclength to be unit speed, we will assume throughout that the curves are unit speed.

Let $\gamma : (\alpha, \beta) \to \sigma(U) = V \subset S$ be a unit speed curve on $S$ included in the coordinate patch $\sigma : U \subset \mathbb{R}^2 \to V \subset S \subset \mathbb{R}^3$. Let $p_0 = \gamma(t_0)$ be a point on the curve. Since $\gamma$ is unit speed, we know by Proposition 2.2 that $\gamma'(t_0) \cdot \gamma''(t_0) = 0$. Under the assumption $\gamma''(t_0) \neq \mathbf{0}$, this means that

$$\gamma'(t_0) \perp \gamma''(t_0).$$

Therefore, we will assume that $\gamma''(t_0) \neq \mathbf{0}$. Since $\gamma'(t_0) \in T_{p_0}S$, we have that

$$\gamma'(t_0) \perp N_\sigma$$

since $N_\sigma \perp T_{p_0}S$, and by the definition of cross product, we obtain

$$\gamma'(t_0) \perp N_\sigma \times \gamma'(t_0).$$

In particular, we see that $\gamma''(t_0) \in \operatorname{span}_{\mathbb{R}}\{N_\sigma, N_\sigma \times \gamma'(t_0)\}$. We can write

$$\gamma''(t_0) = \kappa_n N_\sigma + \kappa_g (N_\sigma \times \gamma'(t_0))$$

for some $\kappa_n, \kappa_g \in \mathbb{R}$. But $\|\gamma'(t_0)\| = \|N_\sigma\| = 1$, so $\|N_\sigma \times \gamma'(t_0)\| = 1$. Since $N_\sigma \cdot (N_\sigma \times \gamma'(t_0)) = 0$, we have

$$\begin{aligned} \gamma''(t_0) \cdot N_\sigma &= (\kappa_n N_\sigma + \kappa_g (N_\sigma \times \gamma'(t_0))) \cdot N_\sigma \\ &= \kappa_n N_\sigma \cdot N_\sigma + \kappa_g (N_\sigma \times \gamma'(t_0)) \cdot N_\sigma \\ &= \kappa_n \cdot 1 + \kappa_g \cdot 0 = \kappa_n. \end{aligned}$$

By a similar computation, we have that

$$\gamma''(t_0) \cdot (N_\sigma \times \gamma''(t_0)) = \kappa_g.$$

Since $\|N_\sigma\| = \|N_\sigma \times \gamma'(t_0)\| = 1$, recall from Section 3.2 that $\gamma''(t_0) \cdot N_\sigma$ is the scalar projection of $\gamma''(t_0)$ onto $N_\sigma$ and $\gamma''(t_0) \cdot (N_\sigma \times \gamma'(t_0))$ is the scalar projection of $\gamma''(t_0)$ onto $(N_\sigma \times \gamma'(t_0))$.

**Definition 3.8**

The **normal curvature** is defined to be

$$\kappa_n := \gamma''(t_0) \cdot N_\sigma.$$

The **geodesic curvature** is defined to be

$$\kappa_g := \gamma''(t_0) \cdot (N_\sigma \times \gamma'(t_0)).$$

We call $\kappa_n N_\sigma$ the **normal component of** $\gamma''(t_0)$.

Note that we can have $\kappa_n = 0$ or $\kappa_g = 0$, but we cannot have $\kappa_n = \kappa_g = 0$ since $\gamma''(t_0) \neq \mathbf{0}$. Moreover, depending on where $\gamma''(t_0)$ lies in the plane $\operatorname{span}_{\mathbb{R}}\{N_\sigma, N_\sigma \times \gamma'(t_0)\}$, we may have $\kappa_n < 0$ or $\kappa_g < 0$.

Since $\gamma''(t_0) \neq \mathbf{0}$, we see that $\kappa = \|\gamma''(t_0)\| > 0$, where $\kappa$ is the usual curvature of $\gamma$. Note that

$$\begin{aligned}
\kappa^2 = \|\gamma''(t_0)\|^2 &= \gamma''(t_0) \cdot \gamma''(t_0) \\
&= (\kappa_n N_\sigma + \kappa_g (N_\sigma \times \gamma'(t_0))) \cdot (\kappa_n N_\sigma + \kappa_g (N_\sigma \times \gamma'(t_0))) \\
&= \kappa_n^2 N_\sigma \cdot N_\sigma + 2\kappa_n \kappa_g (N_\sigma \cdot (N_\sigma \times \gamma'(t_0))) + \kappa_g^2 (N_\sigma \times \gamma'(t_0)) \cdot (N_\sigma \times \gamma'(t_0)) \\
&= \kappa_n^2 + \kappa_g^2,
\end{aligned}$$

which implies that $\kappa = \sqrt{\kappa_n^2 + \kappa_g^2}$. In particular, if $\kappa_n = 0$, then $\kappa = |\kappa_g|$, and if $\kappa_g = 0$, then $\kappa = |\kappa_n|$.

**Definition 3.9**

We call $\gamma$ a **geodesic** if $\kappa_g = 0$ everywhere.

Now, we show that the normal curvature $\kappa_n$ of a smooth unit speed curve $\gamma : (\alpha, \beta) \to V \subset S$ only depends on $\gamma'$ and $\mathcal{F}_{\mathrm{II}}$.

First, note that since the coordinate chart $\sigma : U \subset \mathbb{R}^2 \to V = \sigma(U) \subset S$ is a diffeomorphism, we know that $\sigma^{-1} : V \subset S \to U \subset \mathbb{R}^2$ is also smooth. We set

$$\bar{\gamma} := \sigma^{-1} \circ \gamma : (\alpha, \beta) \to U \subset \mathbb{R}^2.$$

Then $\bar{\gamma}$ is a smooth curve in $U \subset \mathbb{R}^2$ with component functions $u, v : (\alpha, \beta) \to \mathbb{R}$ such that

$$\bar{\gamma}(t) = (u(t), v(t))$$

for all $t \in (\alpha, \beta)$. Therefore, we can write

$$\gamma(t) = \sigma \circ \bar{\gamma}(t) = \sigma(u(t), v(t)).$$

In other words, any curve $\gamma$ in the coordinate chart $\sigma : U \subset \mathbb{R}^2 \to V \subset S$ is of the form $\gamma(t) = \sigma(u(t), v(t))$ for smooth functions $u, v : (\alpha, \beta) \to \mathbb{R}$. By the chain rule, we have

$$\gamma'(t) = \sigma_u(u(t), v(t))u'(t) + \sigma_v(u(t), v(t))v'(t)$$

with $u'(t), v'(t) \in \mathbb{R}$, so the components of $\gamma'(t)$ are $(u'(t), v'(t))$ with respect to the basis $\{\sigma_u, \sigma_v\}$ of $T_{\gamma(t)}S$. Furthermore, the product rule gives

$$\begin{aligned}
\gamma''(t) &= \frac{\mathrm{d}}{\mathrm{d}t}(\sigma_u(u(t), v(t)))u'(t) + \sigma_u u''(t) + \frac{\mathrm{d}}{\mathrm{d}t}(\sigma_v(u(t), v(t)))v'(t) + \sigma_v v''(t) \\
&= [\sigma_{uu} u'(t) + \sigma_{uv} v'(t)]u'(t) + \sigma_u u''(t) + [\sigma_{vu} u'(t) + \sigma_{vv} v'(t)]v'(t) + \sigma_v v''(t) \\
&= [\sigma_{uu}(u'(t))^2 + 2\sigma_{uv} u'(t)v'(t) + \sigma_{vv}(v'(t))^2] + \sigma_u u''(t) + \sigma_v v''(t).
\end{aligned}$$

**Proposition 3.10**

If $\gamma(t) = \sigma(u(t), v(t))$, then
$$\kappa_n = \begin{bmatrix} u' & v' \end{bmatrix} \mathcal{F}_{\mathrm{II}} \begin{bmatrix} u' \\ v' \end{bmatrix}.$$

**Proof of Proposition 3.10.**

We have by definition and our above computation for $\gamma''(t)$ that

$$\begin{aligned}\kappa_n &= \gamma''(t) \cdot N_\sigma \\ &= (\sigma_{uu} \cdot N_\sigma)(u'(t))^2 + 2(\sigma_{uv} \cdot N_\sigma)u'(t)v'(t) + (\sigma_{vv} \cdot N_\sigma)(v'(t))^2 + (\sigma_u \cdot N_\sigma)u''(t) + (\sigma_v \cdot N_\sigma)v''(t) \\ &= (\sigma_{uu} \cdot N_\sigma)(u'(t))^2 + 2(\sigma_{uv} \cdot N_\sigma)u'(t)v'(t) + (\sigma_{vv} \cdot N_\sigma)(v'(t))^2,\end{aligned}$$

where we used the fact that $\sigma_u \cdot N_\sigma = \sigma_v \cdot N_\sigma = 0$ since $N_\sigma \perp T_{\gamma(t)}S = \operatorname{span}_{\mathbb{R}}\{\sigma_u, \sigma_v\}$. If we set

$$\begin{aligned} L &:= \sigma_{uu} \cdot N_\sigma, \\ M &:= \sigma_{uv} \cdot N_\sigma, \\ N &:= \sigma_{vv} \cdot N_\sigma,\end{aligned}$$

then we see that the second fundamental form is

$$\mathcal{F}_{\mathrm{II}} = \begin{bmatrix} L & M \\ M & N \end{bmatrix}$$

and we obtain

$$\kappa_n = L(u'(t))^2 + 2Mu'(t)v'(t) + N(v'(t))^2 = \begin{bmatrix} u' & v' \end{bmatrix} \begin{bmatrix} L & M \\ M & N \end{bmatrix} \begin{bmatrix} u' \\ v' \end{bmatrix} = \begin{bmatrix} u' & v' \end{bmatrix} \mathcal{F}_{\mathrm{II}} \begin{bmatrix} u' \\ v' \end{bmatrix}. \qquad \square$$

As a corollary, we get the following classical result.

**Theorem 3.11: Meusnier**

Let $\gamma$ be a unit speed curve on $S$. Then the normal component $\kappa_n N_\sigma$ of $\gamma''$ only depends on the unit tangent vector $\gamma'$ and not on $\gamma$. (That is, $\kappa_n$ only depends on $\gamma'$ and not in $\gamma$.)

Using this, we can prove the following result.

**Proposition 3.12**

The principal curvatures $\kappa_1$ and $\kappa_2$ (the eigenvalues of $\mathcal{W} = \mathcal{F}_{\mathrm{I}}^{-1}\mathcal{F}_{\mathrm{II}}$) are the maximum and minimum values of $\kappa_n$ for any unit speed curves on $S$ passing through $p_0$.

**Proof of Proposition 3.12.**

We prove this using the Lagrange multiplier method. By Meusnier's theorem (Theorem 3.11), we know that $\kappa_n$ is completely determined by $\gamma'(t) \in \operatorname{span}_{\mathbb{R}}\{\sigma_u, \sigma_v\}$. Suppose that $\gamma'(t) = a\sigma_u + b\sigma_v$ where $a, b \in \mathbb{R}$. Since we are considering unit speed curves, we have $\|\gamma'(t)\| = 1$ and hence

$$1 = \|\gamma'(t)\|^2 = \gamma'(t) \cdot \gamma'(t) = \begin{bmatrix} a & b \end{bmatrix} \mathcal{F}_{\mathrm{I}} \begin{bmatrix} a \\ b \end{bmatrix}.$$

By Proposition 3.10, we also know that

$$\kappa_n = \begin{bmatrix} a & b \end{bmatrix} \mathcal{F}_{\mathrm{II}} \begin{bmatrix} a \\ b \end{bmatrix}.$$

For all $a, b \in \mathbb{R}$, define the functions

$$f(a,b) = \begin{bmatrix} a & b \end{bmatrix} \mathcal{F}_{\mathrm{II}} \begin{bmatrix} a \\ b \end{bmatrix}, \qquad g(a,b) = \begin{bmatrix} a & b \end{bmatrix} \mathcal{F}_{\mathrm{I}} \begin{bmatrix} a \\ b \end{bmatrix}.$$

Our goal is to find the maximum and minimum values of $f(a,b)$ subject to the constraint that $g(a,b) = 1$. For ease of notation, we set

$$\mathcal{F}_{\mathrm{I}} = \begin{bmatrix} \sigma_u \cdot \sigma_u & \sigma_u \cdot \sigma_v \\ \sigma_v \cdot \sigma_u & \sigma_v \cdot \sigma_v \end{bmatrix} = \begin{bmatrix} E & F \\ F & G \end{bmatrix}, \qquad \mathcal{F}_{\mathrm{II}} = \begin{bmatrix} \sigma_{uu} \cdot N_\sigma & \sigma_{uv} \cdot N_\sigma \\ \sigma_{vu} \cdot N_\sigma & \sigma_{vv} \cdot N_\sigma \end{bmatrix} = \begin{bmatrix} L & M \\ M & N \end{bmatrix}.$$

Then we obtain

$$f(a,b) = \begin{bmatrix} a & b \end{bmatrix} \begin{bmatrix} L & M \\ M & N \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = a^2 L + 2abM + b^2 N,$$
$$g(a,b) = \begin{bmatrix} a & b \end{bmatrix} \begin{bmatrix} E & F \\ F & G \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = a^2 E + 2abF + b^2 G.$$

We see that

$$\nabla f(a,b) = \left( \frac{\partial f}{\partial a}, \frac{\partial f}{\partial b} \right) = (2aL + 2bM, 2aM + 2bN) = 2\,\mathcal{F}_{\mathrm{II}} \begin{bmatrix} a \\ b \end{bmatrix},$$
$$\nabla g(a,b) = \left( \frac{\partial g}{\partial a}, \frac{\partial g}{\partial b} \right) = (2aE + 2bF, 2aF + 2bG) = 2\,\mathcal{F}_{\mathrm{I}} \begin{bmatrix} a \\ b \end{bmatrix}.$$

By the Lagrange multiplier method, the maximum and minimum of $f(a,b)$ subject to $g(a,b) = 1$ occur at points $(a,b) \in \mathbb{R}^2$ that are solutions of the system

$$\begin{cases} \nabla f(a,b) = \lambda \nabla g(a,b) \\ g(a,b) = 1 \end{cases}$$

for some $\lambda \in \mathbb{R}$. We see that $\nabla f(a,b) = \lambda \nabla g(a,b)$ if and only if

$$\mathcal{F}_{\mathrm{II}} \begin{bmatrix} a \\ b \end{bmatrix} = \lambda\, \mathcal{F}_{\mathrm{I}} \begin{bmatrix} a \\ b \end{bmatrix},$$

and since $\mathcal{F}_{\mathrm{I}}$ is invertible, this gives

$$\mathcal{W} \begin{bmatrix} a \\ b \end{bmatrix} = \mathcal{F}_{\mathrm{I}}^{-1} \mathcal{F}_{\mathrm{II}} \begin{bmatrix} a \\ b \end{bmatrix} = \lambda \begin{bmatrix} a \\ b \end{bmatrix}.$$

In particular, we have that $(a,b)^T$ is an eigenvector of $\mathcal{W} = \mathcal{F}_{\mathrm{I}}^{-1} \mathcal{F}_{\mathrm{II}}$, and $\lambda$ is an eigenvalue of $\mathcal{W}$ so that $\lambda \in \{\kappa_1, \kappa_2\}$. Using the constraint $g(a,b) = 1$, we find that

$$\kappa_n = \begin{bmatrix} a & b \end{bmatrix} \mathcal{F}_{\mathrm{II}} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} a & b \end{bmatrix} \mathcal{F}_{\mathrm{I}} \left( \mathcal{F}_{\mathrm{I}}^{-1} \mathcal{F}_{\mathrm{II}} \begin{bmatrix} a \\ b \end{bmatrix} \right) = \begin{bmatrix} a & b \end{bmatrix} \mathcal{F}_{\mathrm{I}}\, \lambda \begin{bmatrix} a \\ b \end{bmatrix} = \lambda g(a,b) = \lambda.$$

Therefore, the maximum and minimum values of $\kappa_n = f(a,b)$ subject to $g(a,b) = 1$ are the eigenvalues of $\mathcal{W}$, namely the principal curvatures $\kappa_1$ and $\kappa_2$. $\square$

As a corollary, we obtain the following.

**Corollary 3.13**

For $i \in \{1,2\}$, let $(a_i, b_i)^T$ be an eigenvector of $\kappa_i$ so that $\mathcal{W}t_i = \kappa_i t_i$. If $\kappa_1 \neq \kappa_2$, then the vectors $X_i = a_i\sigma_u + b_i\sigma_v \in T_{p_0}S$ are perpendicular.

**Proof of Corollary 3.13.**

We wish to show that $X_1 \cdot X_2 = 0$. Note that

$$X_1 \cdot X_2 = \begin{bmatrix} a_1 & b_1 \end{bmatrix} \mathcal{F}_{\mathrm{I}} \begin{bmatrix} a_2 \\ b_2 \end{bmatrix},$$

which implies that

$$\kappa_2(X_1 \cdot X_2) = \kappa_2 \begin{bmatrix} a_1 & b_1 \end{bmatrix} \mathcal{F}_{\mathrm{I}} \begin{bmatrix} a_2 \\ b_2 \end{bmatrix} = \begin{bmatrix} a_1 & b_1 \end{bmatrix} \left( \kappa_2 \mathcal{F}_{\mathrm{I}} \begin{bmatrix} a_2 \\ b_2 \end{bmatrix} \right).$$

But $(a_i, b_i)^T$ is a $\kappa_i$-eigenvector of $\mathcal{W} = \mathcal{F}_{\mathrm{I}}^{-1}\mathcal{F}_{\mathrm{II}}$, so $\mathcal{F}_{\mathrm{I}}^{-1}\mathcal{F}_{\mathrm{II}}(a_i, b_i)^T = \kappa_i(a_i, b_i)^T$ and hence

$$\mathcal{F}_{\mathrm{II}} \begin{bmatrix} a_i \\ b_i \end{bmatrix} = \kappa_i \mathcal{F}_{\mathrm{I}} \begin{bmatrix} a_i \\ b_i \end{bmatrix}$$

by left multiplying $\mathcal{F}_{\mathrm{I}}$. Setting $t_i = (a_i, b_i)^T$ for $i \in \{1,2\}$, this gives us

$$\begin{aligned} \kappa_2(X_1 \cdot X_2) &= t_1^T \mathcal{F}_{\mathrm{II}}\, t_2 = (t_1^T \mathcal{F}_{\mathrm{II}}\, t_2)^T \\ &= t_2^T \mathcal{F}_{\mathrm{II}}\, t_1 = t_2^T(\kappa_1 \mathcal{F}_{\mathrm{I}}\, t_1) \\ &= \kappa_1 t_2 \mathcal{F}_{\mathrm{I}}\, t_1^T = \kappa_1(X_2 \cdot X_1) = \kappa_1(X_1 \cdot X_2), \end{aligned}$$

where the second equality follows since $t_1^T \mathcal{F}_{\mathrm{II}}\, t_2$ is a real number and taking the transpose has no effect. This implies that $(\kappa_1 - \kappa_2)(t_1 \cdot t_2) = 0$, and $\kappa_1 \neq \kappa_2$ implies that $t_1 \cdot t_2 = 0$. ☐

As a consequence of Corollary 3.13, we can always pick an orthonormal basis $\{t_1, t_2\}$ of $T_{p_0}S$ such that the components of $t_i$ are $\kappa_i$-eigenvectors.

Note that $\gamma'(t_0) \in T_{p_0}S$ with $\|\gamma'(t_0)\| = 1$, so we can write

$$\gamma'(t_0) = \cos(\theta)t_1 + \sin(\theta)t_2$$

whee $\theta \in [0, 2\pi)$ is the angle between $\gamma'(t_0)$ and $t_1$. Moreover, if we have $\gamma'(t_0) = a\sigma_u + b\sigma_v$, then a similar computation as before implies that

$$\kappa_n = \begin{bmatrix} a & b \end{bmatrix} \mathcal{F}_{\mathrm{II}} \begin{bmatrix} a \\ b \end{bmatrix} = \cos^2(\theta)\kappa_1 + \sin^2(\theta)\kappa_2 =: f(\theta).$$

It follows that the average value of the $\kappa_n$'s is

$$\begin{aligned} \frac{1}{2\pi}\int_0^{2\pi} f(\theta)\,\mathrm{d}\theta &= \frac{1}{2\pi}\int_0^{2\pi} \kappa_1\cos^2\theta + \kappa_2\sin^2\theta\,\mathrm{d}\theta \\ &= \frac{1}{2}(\kappa_1 + \kappa_2) = H \end{aligned}$$

where the above integral can be solved by using the double angle formulae $\cos^2\theta = \frac{1}{2}(1 + \cos 2\theta)$ and $\sin^2\theta = \frac{1}{2}(1 - \cos 2\theta)$. From this, it makes sense why we call $H$ the mean curvature.

## 3.5 Geodesics

We will work with the same setting as before, where $S$ is a surface, $\sigma : U \subset \mathbb{R}^2 \to V \subset S \subset \mathbb{R}^3$ is a smooth coordinate chart, and $\gamma : (\alpha, \beta) \to V$ is a unit speed curve on $S$. Recall from Definition 3.9 that we called $\gamma$ a geodesic if $\kappa_g = 0$ everywhere. An equivalent definition is as follows:

**Definition 3.14**

A unit speed curve $\gamma$ is a **geodesic** if $\gamma''(t) = \mathbf{0}$ or $\gamma''(t) \parallel N_\sigma$ (i.e. they are parallel) for all $t \in (\alpha, \beta)$.

To see that these are indeed equivalent, let's start with this new definition. Then $\gamma$ is a geodesic if and only if $\sigma_u \cdot \gamma''(t) = \sigma_v \cdot \gamma''(t) = 0$ for all $t \in (\alpha, \beta)$. Since $\gamma''(t) \in \operatorname{span}_{\mathbb{R}}\{N_\sigma, N_\sigma \times \gamma'(t)\}$ and

$$\gamma''(t) = \kappa_n N_\sigma + \kappa_g (N_\sigma \times \gamma'(t))$$

with $\kappa_n = \gamma''(t_0) \cdot N_\sigma$ and $\kappa_g = \gamma''(t_0) \cdot (N_\sigma \times \gamma'(t_0))$, this shows that our new definition is equivalent to having $\kappa_g = 0$ for all $t \in (\alpha, \beta)$.

We look at some examples of geodesics.

(1) **(Lines.)** Let $\gamma(t) = p_0 + tw$ for $t \in \mathbb{R}$, where $p_0$ is a point and $w$ is a vector with $\|w\| = 1$. Then $\gamma''(t_0) = \mathbf{0}$ for all $t \in \mathbb{R}$, so $\gamma$ is a geodesic.

(2) **(Great circles on a sphere.)** Consider the sphere $x^2 + y^2 + z^2 = a^2$ for some $a > 0$. A **great circle** $C$ on $S$ is given by the intersection of $S$ with a plane in $\mathbb{R}^3$ passing through the origin (the center of $S$).

Let us check that great circles are geodesics by looking at intersections with vertical planes. (The other cases are proven similarly after first rotating the sphere so that the plane of intersection becomes vertical.) Recall that the sphere can be parametrized by

$$\sigma(\theta, \varphi) = (a \cos\theta \sin(\varphi/a), a \sin\theta \sin(\varphi/a), a \cos(\varphi/a)),$$

where $(\theta, \varphi) \in (0, 2\pi) \times (0, a\pi)$. We introduced an extra factor of $1/a$ in the angle $\varphi$ to obtain a unit speed curve. A direct computation shows that the standard unit normal is

$$N_\sigma(\theta, \varphi) = -(\cos\theta \sin(\varphi/a), \sin\theta \sin(\varphi/a), a \cos(\varphi/a)).$$

A vertical plane is obtained by fixing $\theta = \theta_0$. The corresponding great circle is the curve

$$C : \gamma(\varphi) = (a \cos\theta_0 \sin(\varphi/a), a \sin\theta_0 \sin(\varphi/a), a \cos(\varphi/a)),$$

which is unit speed with $\gamma''(\varphi) = -\frac{1}{a}(\cos\theta_0 \sin(\varphi/a), \sin\theta_0 \sin(\varphi/a), \cos(\varphi/a)) = \frac{1}{a} N_\sigma(\theta_0, \varphi) \neq \mathbf{0}$. We see that $\gamma''(\varphi) \parallel N_\sigma(\theta_0, \varphi)$ everywhere, so $\gamma$ is a geodesic.

**Geodesic equations.** As we discussed earlier, the geodesics on $S$ are the solutions of the equations

$$\begin{cases} \gamma''(t) \cdot \sigma_u(\gamma(t)) = 0 \\ \gamma''(t) \cdot \sigma_v(\gamma(t)) = 0. \end{cases}$$

This is a second order ODE, which can be solved. Write

$$\mathcal{F}_{\mathrm{I}} = \begin{bmatrix} \sigma_u \cdot \sigma_u & \sigma_u \cdot \sigma_v \\ \sigma_v \cdot \sigma_u & \sigma_v \cdot \sigma_v \end{bmatrix} = \begin{bmatrix} E & F \\ F & G \end{bmatrix}$$

and suppose that $\gamma(t) = \sigma(u(t), v(t))$ where $u, v : (\alpha, \beta) \to \mathbb{R}$ are scalar functions. (We can always write $\gamma$ in this way; see the discussion immediately after Definition 3.9.) The above equations can be rewritten as

$$\gamma''(t) \cdot \sigma_u(\gamma(t)) = 0 \iff \frac{\mathrm{d}}{\mathrm{d}t}(Eu' + Fv') = \frac{1}{2} \begin{bmatrix} u' & v' \end{bmatrix} \begin{bmatrix} E_u & F_u \\ F_u & G_u \end{bmatrix} \begin{bmatrix} u' \\ v' \end{bmatrix},$$

$$\gamma''(t) \cdot \sigma_v(\gamma(t)) = 0 \iff \frac{\mathrm{d}}{\mathrm{d}t}(Fu' + Gv') = \frac{1}{2} \begin{bmatrix} u' & v' \end{bmatrix} \begin{bmatrix} E_v & F_v \\ F_v & G_v \end{bmatrix} \begin{bmatrix} u' \\ v' \end{bmatrix},$$

and we call these the **geodesic equations**.

**Proof of the geodesic equations.**

Note that $\gamma(t) = \sigma(u(t), v(t))$, which gives $\gamma'(t) = \sigma_u u'(t) + \sigma_v v'(t)$ and

$$\gamma''(t) \cdot \sigma_u = \frac{\mathrm{d}}{\mathrm{d}t}(\gamma'(t)) \cdot \sigma_u = \frac{\mathrm{d}}{\mathrm{d}t}(\gamma'(t) \cdot \sigma_u) - \gamma'(t) \cdot \frac{\mathrm{d}}{\mathrm{d}t}(\sigma_u).$$

This means that $\gamma''(t) \cdot \sigma_u = 0$ if and only if

$$\frac{\mathrm{d}}{\mathrm{d}t}(\gamma'(t) \cdot \sigma_u) = \gamma'(t) \cdot \frac{\mathrm{d}}{\mathrm{d}t}(\sigma_u).$$

For the left-hand side, we see that

$$\gamma'(t) \cdot \sigma_u = (\sigma_u u'(t) + \sigma_v v'(t)) \cdot \sigma_u = (\sigma_u \cdot \sigma_u) u'(t) + (\sigma_v \cdot \sigma_u) v'(t) = Eu'(t) + Fv'(t).$$

On the other hand, note that $\frac{\mathrm{d}}{\mathrm{d}t}[\sigma_u(u(t), v(t))] = \sigma_{uu} u'(t) + \sigma_{uv} v'(t)$ and hence

$$\begin{aligned}\gamma'(t) \cdot \frac{\mathrm{d}}{\mathrm{d}t}(\sigma_u) &= (\sigma_u u'(t) + \sigma_v v'(t)) \cdot (\sigma_{uu} u'(t) + \sigma_{uv} v'(t)) \\ &= (\sigma_u \cdot \sigma_{uu})(u'(t))^2 + (\sigma_u \cdot \sigma_{uv} + \sigma_v \cdot \sigma_{uu}) u'(t) v'(t) + (\sigma_v \cdot \sigma_{uv})(v'(t))^2.\end{aligned}$$

But we notice that

$$\begin{aligned} E_u &= (\sigma_u \cdot \sigma_u)_u = 2\sigma_u \cdot \sigma_{uu}, \\ F_u &= (\sigma_u \cdot \sigma_v)_u = \sigma_u \cdot \sigma_{uv} + \sigma_u \cdot \sigma_{vv}, \\ G_u &= (\sigma_v \cdot \sigma_v)_u = 2\sigma_v \cdot \sigma_{uv}, \end{aligned}$$

which gives us

$$\gamma'(t) \cdot \frac{\mathrm{d}}{\mathrm{d}t}(\sigma_u) = \frac{1}{2} E_u (u'(t))^2 + F_u u'(t) v'(t) + \frac{1}{2} G_u (v'(t))^2 = \frac{1}{2} \begin{bmatrix} u' & v' \end{bmatrix} \begin{bmatrix} E_u & F_u \\ F_u & G_u \end{bmatrix} \begin{bmatrix} u' \\ v' \end{bmatrix}.$$

This proves the first geodesic equation, and the second is proved similarly. $\square$

The geodesic equations tell us the following:

- Geodesics *always* exist because they are solutions to nice differential equations.
- Geodesics are completely determined by $\mathcal{F}_{\mathrm{I}}$. That is, if two surfaces have coordinate patches with the same first fundamental form, then they have the "same" geodesics.

**Definition 3.15**

A diffeomorphism $\Phi : S_1 \to S_2$ between two surfaces $S_1$ and $S_2$ is called an **isometry** if for any smooth coordinate chart $\sigma : U \subset \mathbb{R}^2 \to V \subset S_1 \subset \mathbb{R}^3$ of $S_1$ and corresponding smooth coordinate chart

$$\tilde{\sigma} := \Phi \circ \sigma : U \subset \mathbb{R}^2 \to \Phi(V) \subset S_2 \subset \mathbb{R}^3$$

of $S_2$, we have that $\mathcal{F}_{\mathrm{I}} = \tilde{\mathcal{F}}_{\mathrm{I}}$.

Note that translations and rotations are isometries, whereas dilations and contractions are not. For another example, let $S_1 = \{(x, y, z) \in \mathbb{R}^3 : z = 0,\ 0 < x < 2\pi\}$ be a subset of the $xy$-plane and let $S_2 = \{(x, y, z) \in \mathbb{R}^3 : x^2 + y^2 = 1,\ x \neq 1\}$ be the upright cylinder of radius 1 minus the line $x = 1$. Consider the diffeomorphism

$$\begin{aligned} \Phi : S_1 &\to S_2 \\ (x, y, 0) &\mapsto (\cos x, \sin x, y). \end{aligned}$$

Let $\sigma : \mathbb{R}^2 \to S_1$ be defined by $(u, v) \mapsto (u, v, 0)$ and let

$$\begin{aligned}\tilde{\sigma} = \Phi \circ \sigma : \mathbb{R}^2 &\to S_2 \\ (u, v) &\mapsto (\cos u, \sin u, v).\end{aligned}$$

Then $\tilde{\sigma}$ is the usual parametrization of the cylinder, so we have that $\mathcal{F}_{\mathrm{I}} = \tilde{\mathcal{F}}_{\mathrm{I}} = I_{2\times 2}$ and $\Phi$ is an isometry.

We can use isometries to find geodesics because the image of a geodesic under an isometry is again a geodesic (using the fact that the geodesic equations depend only on $\mathcal{F}_{\mathrm{I}}$).

(1) **(Geodesics on a plane.)** For all $(x, y) \in \mathbb{R}^2$, set $\sigma(u, v) = p_0 + uw_1 + vw_2$ where $\{w_1, w_2\}$ is an orthonormal set of direction vectors for the plane. Then $\sigma$ is a coordinate chart for the plane such that

$$\mathcal{F}_{\mathrm{I}} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} E & F \\ F & G \end{bmatrix},$$

which implies that

$$\begin{bmatrix} E_u & F_u \\ F_u & G_u \end{bmatrix} = \begin{bmatrix} E_v & F_v \\ F_v & G_v \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}.$$

Then the geodesic equations become

$$\frac{\mathrm{d}}{\mathrm{d}t}(u') = \frac{\mathrm{d}}{\mathrm{d}t}(v') = 0,$$

which means that $u'' = v'' = 0$. In particular, we have $u(t) = at + b$ and $v(t) = ct + d$ for some $a, b, c, d \in \mathbb{R}$. Then all geodesics are of the form

$$\gamma(t) = \sigma(at + b, ct + d)$$

where $\sigma(u, v) = p_0 + uw_1 + vw_2$, so $\gamma$ is a line!

(2) **(Geodesics on the sphere.)** Solving the geodesic equations on the sphere $x^2 + y^2 + z^2 = a^2$ shows that the great circles are the only geodesics (see Pressley, Exercise 8.4 on page 144).

(3) **(Geodesics on the cylinder.)** Consider the cylinder $x^2 + y^2 = 1$ in $\mathbb{R}^3$. As with our previous example, let $S_1 = \{(x, y, z) \in \mathbb{R}^3 : z = 0,\ 0 < x < 2\pi\}$ and $S_2 = \{(x, y, z) \in \mathbb{R}^3 : x^2 + y^2 = 1,\ x \neq 1\}$. We saw that the map $\Phi : S_1 \to S_2$ defined by $(x, y, z) \mapsto (\cos x, \sin x, y)$ is an isometry. This gives us a one-to-one correspondence between geodesics on $S_1$ and $S_2$. Therefore, the geodesics on the cylinder are the images of lines on $S_1$ under $\Phi$. In particular, they are given by

$$\gamma(t) = \Phi(at + b, ct + d) = (\cos(at + b), \sin(at + b), ct + d).$$

This is a line if $a = 0$, a circle if $c = 0$, and a helix if $a$ and $c$ are both nonzero.

**Geodesics as shortest paths.** Let $S$ be a surface. Let $\sigma : U \subset \mathbb{R}^2 \to V \subset S \subset \mathbb{R}^3$ be a smooth coordinate patch, and for simplicity, let $\gamma : [0, 1] \to V \subset S$ be a unit speed curve on $S$. We can write $\gamma(t) = \sigma(u(t), v(t))$ for some scalar functions $u, v : [0, 1] \to \mathbb{R}$.

Suppose that $p = \gamma(0)$ and $q = \gamma(1)$. We want to find necessary conditions for $\gamma$ to be the shortest path on $S$ between $p$ and $q$. That is, we want $\gamma$ to have the smallest arclength of any family of curves on $S$ starting at $p$ and ending at $q$. We will find that if $\gamma$ minimizes distance between $p$ and $q$, then $\gamma$ is a geodesic.

We start by constructing such a family. Let $\delta > 0$. For any $\varepsilon \in (-\delta, \delta)$, we set

$$\gamma_\varepsilon(t) := \sigma(u(t) + \varepsilon\alpha(t), v(t) + \varepsilon\beta(t))$$

where $\alpha, \beta : [0, 1] \to \mathbb{R}$ are smooth functions such that $\alpha(0) = \beta(0) = 0$ and $\alpha(1) = \beta(1) = 0$. Then we have that $\gamma_\varepsilon(0) = \sigma(u(0), v(0)) = \gamma(0) = p$ and $\gamma_\varepsilon(1) = \sigma(u(1), v(1)) = \gamma(1) = q$ for all $\varepsilon \in (-\delta, \delta)$, so each of these curves passes through $p$ and $q$. Note that $\gamma(t) = \gamma_0(t)$.

The arclength of $\gamma_\varepsilon(t)$ between $p$ and $q$ is

$$L(\varepsilon) := \int_0^1 \|\gamma_\varepsilon'(t)\| \,\mathrm{d}t.$$

If $\gamma = \gamma_0$ minimizes arclength, then it must be a global minimum of $L(\varepsilon)$; in particular, it is a critical point of $L(\varepsilon)$, which implies that for any choice of $\alpha, \beta : [0,1] \to \mathbb{R}$, we have

$$\frac{\mathrm{d}}{\mathrm{d}\varepsilon}(L(\varepsilon))\Big|_{\varepsilon=0} = 0.$$

**Proposition 3.16**

We have $\frac{\mathrm{d}}{\mathrm{d}\varepsilon}L(\varepsilon)|_{\varepsilon=0} = 0$ if and only if $\gamma$ is a geodesic.

**Proof of Proposition 3.16.**

By a direct computation (in the notes), it can be shown that

$$\frac{\mathrm{d}}{\mathrm{d}\varepsilon}(L(\varepsilon))\Big|_{\varepsilon=0} = -\int_0^1 [\alpha(t)(\sigma_u \cdot \gamma''(t)) + \beta(t)(\sigma_v \cdot \gamma''(t))] \,\mathrm{d}t.$$

We know that $\gamma$ is a geodesic if and only if $\sigma_u \cdot \gamma''(t) = \sigma_v \cdot \gamma''(t) = 0$, which implies that

$$\frac{\mathrm{d}}{\mathrm{d}\varepsilon}(L(\varepsilon))\Big|_{\varepsilon=0} = 0.$$

Conversely, suppose that $\frac{\mathrm{d}}{\mathrm{d}\varepsilon}L(\varepsilon)|_{\varepsilon=0} = 0$ but $\gamma$ is not a geodesic. This means that $\sigma_u \cdot \gamma''(t_0) \neq 0$ or $\sigma_v \cdot \gamma''(t_0) \neq 0$ for some $t_0 \in [0,1]$. Without loss of generality, suppose that $\sigma_u \cdot \gamma''(t_0) > 0$. By continuity, we have that $\sigma_u \cdot \gamma''(t) > 0$ on $[t_0 - c, t_0 + c]$ for some $c > 0$. Set $\beta \equiv 0$ and let $\alpha$ be a bump function on $[0,1]$ such that $\alpha$ is 1 on $[t_0 - c/2, t_0 + c/2]$ and 0 otherwise. This gives us

$$\frac{\mathrm{d}}{\mathrm{d}\varepsilon}L(\varepsilon)|_{\varepsilon=0} = -\int_0^1 \alpha(t)(\sigma_u \cdot \gamma''(t)) \,\mathrm{d}t < 0$$

since $\alpha(t)(\sigma_u \cdot \gamma''(t))$ is nonnegative on $[0,1]$ and strictly positive on $[t_0 - c/2, t_0 + c/2]$. This gives us a contradiction, so $\gamma$ must be a geodesic. $\square$

This tells us that any curve that minimizes distance is a geodesic. However, it is important to note that the converse is false. Although geodesics minimize arclength locally (i.e. around a point), they may not do so globally. For example, consider the unit sphere $\mathbb{S}^2$. Let $p$ and $q$ be two points on $\mathbb{S}^2$ that are not antipodal (i.e. on opposite ends of the sphere). The geodesic passing through $p$ and $q$ is a great circle, but only one of the circle arcs minimizes the distance between $p$ and $q$.

Another reason why we are interested in geodesics is due to the Gauss-Bonnet theorem.

(1) Draw a "geodesic triangle" and let $\alpha$, $\beta$, and $\gamma$ be the angles between them. Then we will have

$$\alpha + \beta + \gamma = \pi + \int_T K \,\mathrm{d}A.$$

We will define what this integral means later.

(2) If $S$ is a compact "oriented" surface, then Gauss-Bonnet tells us that

$$\frac{1}{2\pi}\int_S K \,\mathrm{d}A = 2(g-1).$$

This is called the **Euler characteristic**. The number $g$ is called the **genus**, which is the number of holes in the surface.