MULTIVARIABLE CALCULUS & GEOMETRY

(MTH3C11)

SELF LEARNING MATERIAL

III SEMESTER

M.Sc. MATHEMATICS
(2019 Admission)

UNIVERSITY OF CALICUT

SCHOOL OF DISTANCE EDUCATION
CALICUT UNIVERSITY P.O.
MALAPPURAM - 673 635, KERALA

190561

School of Distance Education
University of Calicut

Self Learning Material

III Semester

M.Sc. Mathematics
(2019 Admission)

MTH3C11 : Multivariable Calculus & Geometry

Prepared by:

ASHBIN MATHEW
Asst. Professor
Department of Mathematics
St. Thomas College, Thrissur.

Scrutinized by:

Dr. JICY N.
Assistant Professor
Department of Mathematics
SDE, University of Calicut.

DISCLAIMER

"The author(s) shall be solely responsible for the content and views expressed in this book".

Contents

Syllabus **5**

1 Functions of Several Variables **7**
1.1 Euclidean Spaces 7
1.2 Basis and dimension 8
1.3 Linear Transformation and Matrices 11
1.4 Differentiation 18
1.5 The Contraction Principle 26
1.6 The Inverse Function Theorem 27
1.7 The Implicit Function Theorem 29
1.8 Determinants 32

2 Curves and Surfaces **37**
2.1 What is a curve 37
2.2 Arc-length 41
2.3 Reparametrization 44
2.4 Closed curves 47
2.5 Level curves versus parametrized curves 50
2.6 Curvature 52
2.7 Plane curves 56
2.8 Space curves 63
2.9 What is a surface? 69
2.10 Smooth surfaces 77
2.11 Smooth maps 79
2.12 Tangents and derivatives 81
2.13 Normals and orientability 84

3 Curvature of Surfaces **87**
3.1 Level surfaces 87
3.2 Ruled surfaces and surfaces of revolution 88
3.3 Applications of the inverse function theorem 92

3.4 The first fundamental form 95
3.5 Area of a surface patch 99
3.6 The second fundamental form 100
3.7 The Gauss and Weingarten maps 101
3.8 Normal and geodesic curvatures 103
3.9 Gaussian and mean curvatures 106
3.10 Principal curvatures of a surface 111

Reference **119**

Syllabus

MTH3C11: MULTIVARIABLE CALCULUS AND GEOMETRY
No. of Credits:4
No. of hours of Lectures/week: 5

TEXT 1 : RUDIN W., PRINCIPLES OF MATHEMATICAL ANALYSIS, (3rd Edn.), Mc. Graw Hill, 1986
TEXT 2: ANDREW PRESSLEY, ELEMENTARY DIFFERENTIAL GEOMETRY(2nd Edn.), Springer-Verlag, 2010 .

Module:1

Functions of Several Variables Linear Transformations, Differentiation, The Contraction Principle, The Inverse Function Theorem, the Implicit Function Theorem. [Chapter 9 Sections 1-29, 33-37 from Text -1]

Module:2

What is a curve? Arc-length, Reparametrization, Closed curves, Level curves versus parametrized curves. Curvature, Plane curves, Space curves What is a surface, Smooth surfaces, Smooth maps, Tangents and derivatives, Normals and orientability. [Chapter 1 Sections $1-5$, Chapter 2 Sections $1-3$, Chapter 4 Sections $1-5$ from Text - 2]

Module:3

Level surfaces, Ruled surfaces and surfaces of revolution, Applications of the inverse function theorem, Lengths of curves on surfaces, Equiareal maps and a theorem of Archimedes, The second fundamental form, The Gauss and Weingarten maps, Normal and geodesic curvatures. Gaussian and mean curvatures, Principal curvatures of a surface. [Chapter 5 Sections $1, 3\&6$, Chapter 6 Sections 1 and 4 (up to and including 6.4.3) Chapter 7 Sections $1-3$, Chapter 8 Sections $1-2$ from Text - 2]

References

1. M. P. do Carmo: Differential Geometry of Curves and Surfaces;
2. W. Klingenberg: A course in Differential Geometry;
3. J. R. Munkres: Analysis on Manifolds; Westview Press; 1997
4. C. C. Pugh: Real Mathematical Analysis, Springer; 2010
5. M. Spivak: A Comprehensive Introduction to Differential Geometry-Vol. I; Publish or Perish, Boston; 1970
6. M. Spivak: Calculus on Manifolds; Westview Press; 1971
7. V.A. Zorich: Mathematical Analysis-I; Springer; 2008

Module 1

Functions of Several Variables

1.1 Euclidean Spaces

Definition 1.1.1. For each positive integer n, Let $\mathbb{R}^n$ be the set of all ordered $n-$ tuples $\mathbf{x} = (x_1, x_2, ...x_n)$, where $x_1, x_2, ..x_n$ are real numbers, called the *coordinates* of $\mathbf{x}$.

The elements of $\mathbb{R}^n$ are called points, or vectors, especially when $n > 1$. If $\mathbf{y} = (y_1, y_2, ..y_n)$ and if c is a real number, put

$$\mathbf{x} + \mathbf{y} = (x_1 + y_1, x_2 + y_2, ...x_n + y_n),$$
$$c\mathbf{x} = (cx_1, cx_2, ...cx_n)$$

so that $\mathbf{x} + \mathbf{y} \in \mathbb{R}^n$ and $c\mathbf{x} \in \mathbb{R}^n$.This defines addition of vectors, as well as multiplication of a vector by a real number.These two operations satisfy the commutative, associative, and distributive laws and make $\mathbb{R}^n$ into a *vector space over the real field.* The zero element of $\mathbb{R}^n$ is the point $\mathbf{0}$, all of whose coordinate are 0.
We also define the inner product or scalar product of $\mathbf{x}$ and $\mathbf{y}$ by

$$\mathbf{x} \cdot \mathbf{y} = \sum_{i=1}^{n} x_i y_i$$

and the *norm* of $\mathbf{x}$ by

$$| \mathbf{x} | = (\mathbf{x} \cdot \mathbf{x})^{\frac{1}{2}} = \left(\sum_{i=1}^{n} x_i^2 \right)^{\frac{1}{2}}$$

The structure now defined (the vector space $\mathbb{R}^n$ with the above inner product and norm) is called euclidean $n-$ space.

Example 1.1.2. if $\mathbf{x} = (1, 0, -1, 2), \mathbf{y} = (3, 2, 1, 4)$ then $\mathbf{x} + \mathbf{y} = (4, 2, 0, 6)$, $2\mathbf{x} = (2, 0, -2, 4), \mid \mathbf{x} \mid = 6^{\frac{1}{2}}$,

Result 1.1.3. Suppose $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ and c is real then

(a) $|\mathbf{x}| \geq 0$

(b) $|\mathbf{x}| = 0$ if and only if $\mathbf{x} = \mathbf{0}$

(c) $|c\mathbf{x}| = |c||\mathbf{x}|$

(d) $|\mathbf{x} \cdot \mathbf{y}| \leq |\mathbf{x}||\mathbf{y}|$ (Schwarz inequality)

(e) $|\mathbf{x} + \mathbf{y}| \leq |\mathbf{x}| + |\mathbf{y}|$ (Triangle inequality)

1.2 Basis and dimension

Definition 1.2.1. A non empty set $X \subset \mathbb{R}^n$ is a *vector space* if $\mathbf{x} + \mathbf{y} \in X$ and $c\mathbf{x} \in X$ for all $\mathbf{x} \in X, \mathbf{y} \in X$ and for all scalars c

Definition 1.2.2. If $\mathbf{x}_1, \mathbf{x}_2, ...\mathbf{x}_k \in \mathbb{R}^n$ and $c_1, c_2, .., c_k$ are scalars,then the vector

$$c_1\mathbf{x}_1 + c_2\mathbf{x}_2 + ... + c_k\mathbf{x}_k$$

is called a *linear combination* of $\mathbf{x}_1, \mathbf{x}_2, ...\mathbf{x}_k$. If $S \subset \mathbb{R}^n$ and if E is the set of all linear combinations of elements of S, we say that *S spans E* or *E is the span of S*

Exercise 1.2.3. If S is a nonempty subset of a vector space X, prove that the span of S is a vector space

Example 1.2.4. if $\mathbf{x}_1 = (1, 2, 4), \mathbf{x}_2 = (-1, 3, 1)$ then $\mathbf{x} = (1, 12, 14)$ is a linear combination of the vectors $\mathbf{x}_1$ and $\mathbf{x}_2$. Since $\mathbf{x} = (1, 12, 14) = 3\mathbf{x}_1 + 2\mathbf{x}_2$

Definition 1.2.5. A set consisting of vectors $\{\mathbf{x}_1, \mathbf{x}_2, ...\mathbf{x}_k\}$ is said to be independent if the relation $c_1\mathbf{x}_1 + c_2\mathbf{x}_2 + ... + c_k\mathbf{x}_k = \mathbf{0}$ implies that $c_1 = c_2 = \ldots = c_k = 0$ .Otherwise $\{\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_k\}$ is said to be dependent.

Observe that no independent set contains the null vector.

Example 1.2.6. $A = \{(1,0,0),(1,1,0),(1,1,1)\}$ is an independent set whereas $B = \{(1,2),(3,4),(5,6)\}$ is a dependent set.

Exercise 1.2.7. Prove that Every subset of a linearly independent set is linearly independent.

Definition 1.2.8. If a Vector space X contains an independent set of r vectors but contains no independent set of $r+1$ vectors, we say that X has *dimension* r and write it as $\dim X = r$.

Observe that if X contains an independent set of n elements, then $\dim X \geq n$.

Exercise 1.2.9. Prove that if $\dim X = r$ then we can find linearly independent subsets of X which contains elements less than r.

Definition 1.2.10. An independent subset of a vector space X which spans X is called a *basis* of X.

Observe that if $B = \{\mathbf{x}_1, \mathbf{x}_2, ...\mathbf{x}_r\}$ is a basis of X,then every $\mathbf{x} \in X$ has a unique representation of the form $\mathbf{x} = \sum c_j \mathbf{x}_j$, Such a representation exists since B spans X, and it is unique since B is independent. The numbers $c_1, ...c_r$ are called the coordinates of x with respect to the basis B.
The most familiar example of a basis is the set $\mathbf{e}_1, ...\mathbf{e}_n$ where $\mathbf{e}_j$ is the vector in $\mathbb{R}^n$ whose jth coordinate is 1 and whose other coordinates are all 0.If $\mathbf{x} = (x_1, ...x_n)$ then $\mathbf{x} = \sum_{j=1}^{n} x_j \mathbf{e}_j$.We shall call $\{\mathbf{e}_1, ...\mathbf{e}_n\}$ the *standard basis* of $\mathbb{R}^n$.

Theorem 1.2.11. Let r be a positive integer.If a vector space X is spanned by a set of r vectors then $\dim X \leq r$.

Proof. Assume that this result is false. So there exist a vector space X which is spanned by a set of r vectors, say, $S_0 = \{\mathbf{x}_1, ..., \mathbf{x}_r\}$ and $\dim X > r$. Since $\dim X > r$, X contains an independent set $Q = \{\mathbf{y}_1, ...\mathbf{y}_{r+1}\}$ contains $r+1$ elements.
Suppose $0 \leq i < r$ and suppose a set S_i, has been constructed which spans X and which consists of all $\mathbf{y}_j$ with $1 \leq j \leq i$ plus a certain collection of $r-i$ members of S_0 , say $\mathbf{x}_1, ...\mathbf{x}_{r-i}$, (In other words S_i, is obtained from S_0 by replacing i of its elements by members of Q, without altering the span.) Since S_i, spans X, $\mathbf{y}_{i+1}$ is in the span of S_i; hence there are scalars $a_1, ...a_{i+1}, b_1, ..b_{r-i}$ with $a_{i+1} = 1$, such that

$$\sum_{j=1}^{i+1} a_j \mathbf{y}_j + \sum_{k=1}^{r-i} b_k \mathbf{x}_k = \mathbf{0}$$

If all b_k's were 0, the independence of Q would force all a_j's to be 0, a contradiction. It follows that some $x_k \in S_i$, is a linear combination of the other members of $T_i = S_i \cup \{y_{i+1}\}$. Remove this x_k from T_i and call the remaining set S_{i+1}. Then S_{i+1} spans the same set as T_i, namely X. Starting with S_0, we thus construct sets $S_1, ..., S_r$.The set S_r consists of $\mathbf{y}_1, ...\mathbf{y}_r$ and our constriction shows that it spans X. But Q is independent ; hence $\mathbf{y}_{r+1}$ is not in the span of S_r,which is a contradiction. □

(Let's try to understand the construction of S_i's.
Span of $S_0 = X$ and $\mathbf{y}_1 \in X$ implies that there exist scalars $a_1, b_1, b_2, ...b_r$ with $a_1 = 1$ such that $a_1\mathbf{y}_1 + b_1\mathbf{x}_1 + ... + b_{r-1}\mathbf{x}_{r-1} + b_r\mathbf{x}_r = \mathbf{0}$. If all b_k's were 0,the independence of Q would force that $a_1 = 0$, a contradiction. So without loss of generality assume that $b_r \neq 0$.
It follows that $x_r \in S_0$, is a linear combination of $\{\mathbf{x}_1, ..., \mathbf{x}_{r-1}, \mathbf{y}_1\}$. Let $T_0 = S_0 \cup \{y_1\} = \{\mathbf{x}_1, ..., \mathbf{x}_{r-1}, \mathbf{x}_r, \mathbf{y}_1\}$ and let $S_1 = \{\mathbf{x}_1, ..., \mathbf{x}_{r-1}, \mathbf{y}_1\}$. Since x_r is a linear combination of $\{\mathbf{x}_1, ..., \mathbf{x}_{r-1}, \mathbf{y}_1\}$ imply that span of T_0 is same as span of S_1. But span of S_0 is X and $T_0 = S_0 \cup \{y_1\}$ imply that span of T_0 is X. So span of S_1 is X.

Span of $S_1 = X$ and $\mathbf{y}_2 \in X$ implies that there exist scalars $a_1, a_2, b_1, b_2, ...b_{r-1}$ with $a_2 = 1$ such that $a_1\mathbf{y}_1 + a_2\mathbf{y}_2 + b_1\mathbf{x}_1 + ... + b_{r-1}\mathbf{x}_{r-1} = \mathbf{0}$. If all b_k's were 0,the independence of Q would force that $a_2 = 0$, a contradiction. So without loss of generality assume that $b_{r-1} \neq 0$. It follows that $x_{r-1} \in S_1$, is a linear combination of $\{\mathbf{x}_1, ..., \mathbf{x}_{r-2}, \mathbf{y}_1, \mathbf{y}_2\}$.
Let $T_1 = S_1 \cup \{y_2\} = \{\mathbf{x}_1, ..., \mathbf{x}_{r-1}, \mathbf{y}_1, \mathbf{y}_2\}$ and let $S_2 = \{\mathbf{x}_1, ..., \mathbf{x}_{r-2}, \mathbf{y}_1, \mathbf{y}_2\}$. Then span of T_1 is same as span of S_2. But span of T_1 is X. So span of S_2 is X. etc....).

Corollary 1.2.12. dim $\mathbb{R}^n = n$.

Proof. Since $\{\mathbf{e}_1, ..., \mathbf{e}_n\}$ spans $\mathbb{R}^n$, the theorem 1.2.11 shows that dim $\mathbb{R}^n \leq n$. Since $\{\mathbf{e}_1, ..., \mathbf{e}_n\}$ is independent, by definition 1.2.8 dim $\mathbb{R}^n \geq n$. □

Theorem 1.2.13. Suppose X is a vector space, and dim $\mathbb{R}^n = n$.

(a) A set E of n vectors in X spans X if and only if E is independent.

(b) X has a basis and every basis consists of n vectors.

(c) If $1 \leq r \leq n$ and $\{\mathbf{y}_1, .., \mathbf{y}_r\}$ is an independent set in X, then X has a basis containing $\{\mathbf{y}_1, .., \mathbf{y}_r\}$.

Proof.

(a) Suppose $E = \{\mathbf{x}_1, ..., \mathbf{x}_n\}$. Since dim $X = n$, the set $\{\mathbf{x}_1, ..., \mathbf{x}_n, \mathbf{y}\}$ is dependent, for every $\mathbf{y} \in X$.
Let $\mathbf{y} \in X$. If E is independent and the set $\{\mathbf{x}_1, ..., \mathbf{x}_n, \mathbf{y}\}$ is dependent then it follows that $a_1\mathbf{y} + b_1\mathbf{x}_1 + ... + b_n\mathbf{x}_n = \mathbf{0}$ for some $a_1 \neq 0$. So $\mathbf{y}$ is in the span of E, hence E spans X.
Conversely, if E is dependent, one of its members can be removed without changing the span of E (by similar construction of S_{i+1} from T_i in the proof of the Theorem 1.2.11). Hence, span of E is same as span of E' for some set $E' \subset E$ containing $n-1$ elements. Hence, by Theorem 1.2.11, dim (span of E') $\leq n-1$.Hence E cannot span X.

(b) Since dim $X = n$, X contains an independent set of n vectors, and (a) shows that every such set is a basis of X. Suppose that E' is a basis of X. Then, by the definition of basis we know that E' is an independent set and span of E' is X. Since $\dim X = n$, every independent set has at- most n elements by the Definition 1.2.8. So the number of elements in E' is at-most n . If number of elements in E' is less than n, then by Theorem 1.2.11, $\dim X \leq n-1$, which is not possible. So number of elements in E' is n.

(c) Let$\{\mathbf{x}_1, ..., \mathbf{x}_n\}$ be a basis of X. The set

$$S = \{\mathbf{y}_1, ..., \mathbf{y}_r, \mathbf{x}_1, ..., \mathbf{x}_n\}$$

spans X and is dependent, since it contains more than n vectors. The argument used in the proof of Theorem 1.2.11 shows that one of the $\mathbf{x}_i$'s is a linear combination of the other members of S. If we remove this $\mathbf{x}_i$, from S, the remaining set still spans X. This process can be repeated r times and leads to a basis of X which contains $\{\mathbf{y}_1, .., \mathbf{y}_r\}$ by (a).

□

1.3 Linear Transformation and Matrices

Definition 1.3.1. A mapping A of a vector space X into a vector space Y is said to be a linear transformation if

$$A(\mathbf{x}_1 + \mathbf{x}_2) = A(\mathbf{x}_1) + A(\mathbf{x}_2) \quad , \qquad A(c\mathbf{x}) = cA(\mathbf{x})$$

for all $\mathbf{x}, \mathbf{x}_1, \mathbf{x}_2 \in X$ and all scalars c. Note that one often writes $A\mathbf{x}$ instead of $A(\mathbf{x})$ if A is linear.

Exercise 1.3.2. Prove that if the mapping A of a vector space X into a vector space Y is linear then $A(\mathbf{0}) = \mathbf{0}$.

Observe that a linear transformation A of X into Y is completely determined by its action on any basis. If $\{\mathbf{x}_1, ..., \mathbf{x}_n\}$ is a basis of X, then every $\mathbf{x} \in X$ has a unique representation of the form

$$\mathbf{x} = \sum_{i=1}^{n} c_i \mathbf{x}_i$$

and the linearity of A allows us to compute $A\mathbf{x}$ from the vectors $A\mathbf{x}_1, ..., A\mathbf{x}_n$, and the coordinates $c_1, ..c_n$ by the formula

$$A\mathbf{x} = \sum_{i=1}^{n} c_i A\mathbf{x}_i \tag{1.1}$$

Linear transformations of X into X are often called *linear operators* on X. If A is a linear operator on X which is one-to-one and maps X onto X, we say that A is *invertible*. In this case we can define an operator A^{-1} on X by requiring that $A^{-1}(A\mathbf{x}) = \mathbf{x}$ for all $\mathbf{x} \in X$

Exercise 1.3.3. Prove that If A is a linear operator on X which is one-to-one and maps X onto X then A^{-1} is linear and invertible.

Exercise 1.3.4. Assume that A is a linear transformations of X into Y and $A\mathbf{x} = \mathbf{0}$ only when $\mathbf{x} = \mathbf{0}$. Prove that A is one-one.

Theorem 1.3.5. A linear operator A on a finite-dimensional vector space X is one-to-one if and only if the range of A is all of X.

Proof. Let $\{\mathbf{x}_1, ..., \mathbf{x}_n\}$ be a basis of X. The linearity of A shows that its range $\mathscr{R}(A)$ is the span of the set $Q = \{A\mathbf{x}_1, ..., A\mathbf{x}_n\}$(by equation 1.1). Therefore from Theorem 1.2.13(a), Q spans X if and only if Q is independent. That is $\mathscr{R}(A) = X$ if and only if Q is independent. We are going to prove that this happens if and only if A is one-to-one.
Suppose A is one-to-one and $\sum c_i A\mathbf{x}_i = \mathbf{0}$.
Then by the linearity of A, $A(\sum c_i \mathbf{x}_i) = \mathbf{0}$. $A(\mathbf{0}) = \mathbf{0}$ and A is one-to-one imply that $\sum c_i \mathbf{x}_i = \mathbf{0}$.
Since $\{\mathbf{x}_1, ..., \mathbf{x}_n\}$ is the basis of X , we will get $c_1 = ... = c_n = 0$. So we conclude that Q is independent.
Conversely, suppose Q is independent and $A(\sum c_i \mathbf{x}_i) = \mathbf{0}$. Then $\sum c_i A\mathbf{x}_i = \mathbf{0}$. Hence $c_1 = ... = c_n = 0$ and we conclude: $A\mathbf{x} = 0$ only if $\mathbf{x} = 0$.If $A\mathbf{x} = A\mathbf{y}$, then $A(\mathbf{x} - \mathbf{y}) = \mathbf{0}$,so that $\mathbf{x} - \mathbf{y} = \mathbf{0}$ and this says that A is one-to-one.

Thus we have proved that A is one-to-one if and only if Q is independent. From Theorem 1.2.13(a) we know that Q is independent if and only if span of Q is X. We know that span of Q is $\mathscr{R}(A)$. From this we can conclude that A is one-to-one if and only if $\mathscr{R}(A) = X$. □

Definition 1.3.6.

(a) Let $L(X, Y)$ be the set of all linear transformations of the vector space X into the vector space Y.
Instead of $L(X, X)$ we shall simply write $L(X)$. If $A_1, A_2 \in L(X, Y)$ and if c_1, c_2 are scalars, define $c_1A_1 + c_2A_2$ by

$$(c_1A_1 + c_2A_2)\mathbf{x} = c_1A_1\mathbf{x} + c_2A_2\mathbf{x} \quad (\mathbf{x} \in X)$$

It is then clear that $c_1A_1 + c_2A_2 \in L(X, Y)$.

(b) If X, Y, Z are vector spaces, and if $A \in L(X, Y)$ and $B \in L(Y, Z)$, we define their product BA to be the composition of A and B:

$$(BA)\mathbf{x} = B(A)\mathbf{x}, \quad (\mathbf{x} \in X).$$

Then $BA \in L(X, Z)$.

Note that BA need not be the same as AB even if $X = Y = Z$.

(c) For $A \in L(\mathbb{R}^n, \mathbb{R}^m)$, define the *norm* $\|A\|$ of A to be the sup of all numbers $|A\mathbf{x}|$, where $\mathbf{x}$ ranges over all vectors in $\mathbb{R}^n$ with $|x| \leq 1$.

Exercise 1.3.7. Prove that $|A\mathbf{x}| \leq \|A\||\mathbf{x}|$ holds for all $\mathbf{x} \in \mathbb{R}^n$.

(Hint: For any $\mathbf{x} \in \mathbb{R}^n$, let $\mathbf{y} = \frac{\mathbf{x}}{|\mathbf{x}|}$ then $|A\mathbf{y}| \leq \|A\|$)

Exercise 1.3.8. If λ is such that $|A\mathbf{x}| \leq \lambda|\mathbf{x}|$ for all $x \in \mathbb{R}^n$,then $\|A\| \leq \lambda$.

(Hint: For $|\mathbf{x}| \leq 1, |A\mathbf{x}| \leq \lambda$ and use the property of supremum)

Example 1.3.9. For every $A \in L(\mathbb{R}^n, \mathbb{R}^1)$ corresponds a unique $\mathbf{y} \in \mathbb{R}^n$ such that $A\mathbf{x} = \mathbf{x} \cdot \mathbf{y}$ and $\|A\| = |y|$

Let $\{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$ be the standard basis of $\mathbb{R}^n$, and let

$$\mathbf{y} = A(\mathbf{e}_1)\,\mathbf{e}_1 + \cdots + A(\mathbf{e}_n)\,\mathbf{e}_n$$

Then for any $\mathbf{x} = c_1\mathbf{e}_1 + \cdots + c_n\mathbf{e}_n$ we have

$$\begin{aligned} A(\mathbf{x}) &= c_1A(\mathbf{e}_1) + \cdots + c_nA(\mathbf{e}_n) \\ &= \mathbf{y} \cdot \mathbf{x} \end{aligned}$$

There can be at most one such $\mathbf{y}$, since if $A(\mathbf{x}) = \mathbf{z} \cdot \mathbf{x}$, then

$$|\mathbf{y} - \mathbf{z}|^2 = \mathbf{y} \cdot \mathbf{y} - \mathbf{y} \cdot \mathbf{z} - \mathbf{z} \cdot \mathbf{y} + \mathbf{z} \cdot \mathbf{z} = A(\mathbf{y}) - A(\mathbf{y}) - A(\mathbf{z}) + A(\mathbf{z}) = 0$$

By the Schwarz inequality (Result 1.1.3) (d) we have

$$|A(\mathbf{x})| = |\mathbf{y} \cdot \mathbf{x}| \leq |\mathbf{y}||\mathbf{x}|$$

for all $\mathbf{x}$, so by exercise 1.3.8 $\|A\| \leq |\mathbf{y}|$.
On the other hand $A(\mathbf{y}) = \mathbf{y} \cdot \mathbf{y} = |\mathbf{y}|^2$, so by exercise 1.3.7 $|A\mathbf{y}| \leq \|A\||\mathbf{y}|$ implies that $\|A\| \geq |\mathbf{y}|$.

Theorem 1.3.10.

(a) If $A \in L(\mathbb{R}^n, \mathbb{R}^m)$, then $\|A\| < \infty$ and A is a uniformly continuous mapping of $\mathbb{R}^n$ into $\mathbb{R}^m$.

(b) If $A, B \in L(\mathbb{R}^n, \mathbb{R}^m)$ and c is a scalar then

$$\|A + B\| \leq \|A\| + \|B\|, \quad \|cA\| = |c|\|A\|$$

with the distance between A and B defined as $\|A - B\|$, $L(\mathbb{R}^n, \mathbb{R}^m)$ is a metric space.

(c) If $A \in L(\mathbb{R}^n, \mathbb{R}^m)$ and $B \in L(\mathbb{R}^m, \mathbb{R}^k)$ then

$$\|BA\| \leq \|B\|\|A\|$$

Proof.

(a) Let $\{\mathbf{e}_1, ..., \mathbf{e}_n\}$ be the standard basis in $\mathbb{R}^n$ and suppose $\mathbf{x} = \sum c_i \mathbf{e}_i$, $|\mathbf{x}| \leq 1$,so that $|c_i| \leq 1$ for $i = 1, ..., n$. Then by results 1.1.3 (c) and 1.1.3 (e)

$$|A\mathbf{x}| = \left|\sum c_i A\mathbf{e}_i\right| \leq \sum |c_i||A\mathbf{e}_i| \leq \sum |A\mathbf{e}_i|$$

So that

$$\|A\| \leq \sum_{i=1}^{n} |A\mathbf{e}_i| < \infty$$

By Exercise 1.3.7, $|A\mathbf{x} - A\mathbf{y}| = |A(\mathbf{x} - \mathbf{y})| \leq \|A\||\mathbf{x} - \mathbf{y}|$ if $\mathbf{x}$, $\mathbf{y} \in \mathbb{R}^n$. So A is uniformly continuous

(b)
$$|(A+B)\mathbf{x}| = |A\mathbf{x}+B\mathbf{x}| \le |A\mathbf{x}| + |B\mathbf{x}| \le (\|A\| + \|B\|)|\mathbf{x}|$$
So by Exercise 1.3.8, $\|A+B\| \le \|A\| + \|B\|$
The second part is proved in the same manner.
If $A, B, C \in L(\mathbb{R}^n, \mathbb{R}^m)$ then we have

(i) $\|A-B\| \ge 0$
$\|A-B\| = 0$ imply that $A(x) = B(x)$ for every $\mathbf{x} \in \mathbb{R}^n$ with $|\mathbf{x}| \le 1$.
So for $\mathbf{0} \ne \mathbf{x} \in \mathbb{R}^n$, if $\|A-B\| = 0$ imply that $A\left(\frac{\mathbf{x}}{|\mathbf{x}|}\right) = B\left(\frac{\mathbf{x}}{|\mathbf{x}|}\right)$.
So $A = B$

(ii) $\|A-B\| = \|B-A\|$

(iii) we have the triangle inequality
$$\|A-C\| = \|(A-B)+(B-C)\| \le \|A-B\| + \|B-C\|$$
So $L(\mathbb{R}^n, \mathbb{R}^m)$ is a metric space with *norm* metric

(c) by Exercise 1.3.7 and 1.3.8, $|BA\mathbf{x}| = |B(A\mathbf{x})| \le \|B\||Ax| \le \|B\|\|A\||\mathbf{x}|$.
So
$$\|BA\| \le \|B\|\|A\|$$

□

Theorem 1.3.11. Let Ω be the set of all invertible operators on $\mathbb{R}^n$

(a) If $A \in \Omega, B \in L(\mathbb{R}^n)$, and
$$\|B-A\|\ \|A^{-1}\| < 1$$
Then $B \in \Omega$.

(b) Ω is an open subset of $L(\mathbb{R}^n)$ and the mapping $A \to A^{-1}$ is continuous on Ω.

Proof. (a) Put $\|A^{-1}\| = \frac{1}{\alpha}$, put $\|B-A\| = \beta$ then $\beta < \alpha$. For every $\mathbf{x} \in \mathbb{R}^n$
$$\begin{aligned}\alpha|\mathbf{x}| &= \alpha|A^{-1}A\mathbf{x}| \le \alpha\|A^{-1}\|\ |Ax| \\ &= |A\mathbf{x}| \le |(A-B)\mathbf{x}| + |B\mathbf{x}| \le \beta|\mathbf{x}| + |B\mathbf{x}|\end{aligned}$$
So that
$$(\alpha-\beta)|\mathbf{x}| \le |B\mathbf{x}| \qquad (\mathbf{x} \in \mathbb{R}^n) \tag{1.2}$$
Since $\alpha - \beta > 0$; equation 1.2 shows that $B\mathbf{x} \ne \mathbf{0}$ if $\mathbf{x} \ne \mathbf{0}$. Hence B is 1-1. By Theorem 1.3.5, $B \in \Omega$.

(b) Let $A \in \Omega$ with $\|A^{-1}\| = \frac{1}{\alpha}$. Let $N_\alpha(A)$ consisting of all $C \in L(\mathbb{R}^n)$ with $\|C - A\| < \alpha$. Then by (a) $N_\alpha(A) \subset \Omega$. So Ω is an open subset of $L(\mathbb{R}^n)$.

Next we have to prove that the function $f : \Omega \to \Omega$ defined by $f(A) = A^{-1}$ is continuous. For that replace $\mathbf{x}$ by $B^{-1}\mathbf{y}$ in equation 1.2 the resulting inequality

$$(\alpha - \beta)|B^{-1}\mathbf{y}| \leq |BB^{-1}\mathbf{y}| = |\mathbf{y}| \quad (\mathbf{y} \in \mathbb{R}^n)$$

imply that

$$|B^{-1}\mathbf{y}| \leq |\frac{|\mathbf{y}|}{(\alpha - \beta)} \quad (\mathbf{y} \in \mathbb{R}^n)$$

So by exercise 1.3.8, $\|B^{-1}\| \leq \frac{1}{(\alpha-\beta)}$. The identity

$$B^{-1} - A^{-1} = B^{-1}(A - B)A^{-1}$$

combined with Theorem 1.3.10 (c), implies therefore that

$$\|B^{-1} - A^{-1}\| \leq \|B^{-1}\|\|A - B\|\|A^{-1}\| \leq \frac{\beta}{\alpha(\alpha - \beta)}$$

If $B \to A$ then $\beta \to 0$ imply $\|B^{-1} - A^{-1}\| \to 0$.That is $f(B) \to f(A)$.

□

Matrices:
Suppose $\{\mathbf{x}_1, ..., \mathbf{x}_n\}$ and $\{\mathbf{y}_1, ..., \mathbf{y}_m\}$ are bases of vector spaces X and Y respectively. Then every $A \in L(X, Y)$ determines a set of numbers a_{ij} such that

$$A\mathbf{x}_j = \sum_{i=1}^{m} a_{ij}\mathbf{y}_i \quad (1 \leq j \leq n) \tag{1.3}$$

It is convenient to visualize these numbers in a rectangular array of m rows and n columns, called an m by n matrix

$$[A] = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}$$

Observe that the coordinates a_{ij} of the vector $A\mathbf{x}_j$ (with respect to the basis $\{y_1, \ldots, y_n\}$ appear in the j th column of $[A]$. The vectors $A\mathbf{x}_j$ are therefore sometimes called the column vectors of $[A]$. With this terminology, the range of A is spanned by the column vectors of $[A]$.

If $\mathbf{x} = \Sigma c_j \mathbf{x}_j$, the linearity of A, combined with 1.3, shows that

$$A\mathbf{x} = \sum_{i=1}^{m} \left(\sum_{i=1}^{n} a_{ij} c_j \right) \mathbf{y}_i \tag{1.4}$$

Thus the coordinates of $A\mathbf{x}$ are $\Sigma_j a_{ij} c_j$. Note that in 1.3 the summation ranges over the first subscript of a_{ij}, but that we sum over the second subscript when computing coordinates.

Suppose next that an m by n matrix is given, with real entries a_{ij}. If A is then defined by 1.4, it is clear that $A \in L(X, Y)$ and that $[A]$ is the given matrix. Thus there is a natural 1-1 correspondence between $L(X, Y)$ and the set of all real m by n matrices.

Example 1.3.12. Let $A : \mathbb{R}^2 \to \mathbb{R}^2$ by $A(x_1, x_2) = (2x_1 + 3x_2, 4x_1 - x_2)$. Then

$$[A] = \begin{bmatrix} 2 & 3 \\ 4 & -1 \end{bmatrix}$$

Example 1.3.13. Let

$$[B'] = \begin{bmatrix} 2 & 4 \\ 6 & 8 \end{bmatrix}$$

Then by 1.4 define $B : \mathbb{R}^2 \to \mathbb{R}^2$ by $B(x_1, x_2) = (2x_1 + 4x_2, 6x_1 + 8x_2)$ then $B \in L(\mathbb{R}^2)$ and $[B] = [B']$.

We emphasize, that $[A]$ depends not only on A but also on the choice of bases in X and Y. The same A may give rise to many different matrices if we change bases and vice versa.

If Z is a third vector space, with basis $\{\mathbf{z}_1, \ldots, \mathbf{z}_p\}$, if A is given by 1.3, and if

$$B\mathbf{y}_i = \sum_k b_{ki} \mathbf{z}_k, \quad (BA)\mathbf{x}_j = \sum_k c_{kj} \mathbf{z}_k,$$

then $A \in L(X, Y), B \in L(Y, Z), BA \in L(X, Z)$, and since

$$\begin{aligned} B\left(A\mathbf{x}_j\right) &= B \sum_i a_{ij} \mathbf{y}_i = \sum_i a_{ij} B\mathbf{y}_i \\ &= \sum a_{ij} \sum_k b_{ki} \mathbf{z}_k = \sum_k \left(\sum_i b_{ki} a_{ij} \right) \mathbf{z}_k \end{aligned}$$

the independence of $\{\mathbf{z}_1, \ldots, \mathbf{z}_p\}$ implies that

$$c_{kj} = \sum_i b_{ki} a_{ij} \quad (1 \le k \le p, 1 \le j \le n) \tag{1.5}$$

This shows how to compute the p by n matrix $[BA]$ from $[B]$ and $[A]$. If we define the product $[B][A]$ to be $[BA]$, then 1.5 describes the usual rule of matrix multiplication.

Finally, suppose $\{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$ and $\{\mathbf{y}_1, \ldots, \mathbf{y}_m\}$ are standard bases of $\mathbb{R}^n$ and $\mathbb{R}^m$, and A is given by 1.4.The Schwarz inequality shows that

$$|A\mathbf{x}|^2 = \sum_i \left(\sum_j a_{ij} c_J \right)^2 \leq \sum_i \left(\sum_j a_{ij}^2 \ \sum_j c_j^2 \right) = \sum_{i,j} a_{ij}^2 |\mathbf{x}|^2.$$

Thus by exercise 1.3.8

$$\|A\| \leq \left\{ \sum_{1,j} a_{ij}^2 \right\}^{1/2} \tag{1.6}$$

Theorem 1.3.14. If S is a metric space, if $a_{11}, \ldots, a_{mn}$ are real continuous functions on S, and if, for each $p \in S$, A_p is the linear transformation of $\mathbb{R}^n$ into $\mathbb{R}^m$ whose matrix has entries $a_{ij}(p)$, then the mapping $p \to A_p$ is a continuous mapping of S into $L(\mathbb{R}^n, \mathbb{R}^m)$.

Proof. Here a_{ij}'s are continuous functions from S to $\mathbb{R}$ for $1 \leq i \leq m$, $1 \leq j \leq n$. Let $p \in S$, then consider the m by n matrix

$$[A_p] = \begin{bmatrix} a_{11}(p) & a_{12}(p) & \cdots & a_{1n}(p) \\ a_{21}(p) & a_{22}(p) & \cdots & a_{2n}(p) \\ \cdots & \cdots & \cdots & \cdots \\ a_{m1}(p) & a_{m2}(p) & \cdots & a_{mn}(p) \end{bmatrix}$$

Let $A_p \in L(\mathbb{R}^n, \mathbb{R}^m)$ is the corresponding linear transformation. If $q \to p$, then $a_{ij}(q) \to a_{ij}(p)$. So $a_{ij}(q) - a_{ij}(p) \to 0$. So, if we apply equation 1.6 to $A_q - A_p$ in place of A, we will get $\|A_q - A_p\| \to 0$ as $q \to p$. □

1.4 Differentiation

Definition 1.4.1. Suppose E is an open set in $\mathbb{R}^n$, $\mathbf{f}$ maps E into $\mathbb{R}^m$, and $\mathbf{x} \in E$. If there exist a linear transformation A of $\mathbb{R}^n$ into $\mathbb{R}^m$ such that

$$\lim_{\mathbf{h} \to \mathbf{0}} \frac{|\mathbf{f}(\mathbf{x}+\mathbf{h}) - \mathbf{f}(\mathbf{x}) - A\mathbf{h}|}{|\mathbf{h}|} = 0 \tag{1.7}$$

then we say that $\mathbf{f}$ is *differentiable* at $\mathbf{x}$, and we write

$$\mathbf{f}'(\mathbf{x}) = A \tag{1.8}$$

If $\mathbf{f}$ is differentiable at every $\mathbf{x} \in E$, we say that $\mathbf{f}$ is *differentiable* in E.

Remarks

(a) The equation 1.7 can be rewritten in the form

$$\mathbf{f}(\mathbf{x}+\mathbf{h}) - \mathbf{f}(\mathbf{x}) = \mathbf{f}\,'(\mathbf{x})\mathbf{h} + \mathbf{r}(\mathbf{h}) \tag{1.9}$$

where the remainder $\mathbf{r}(\mathbf{h})$ satisfies

$$\lim_{\mathbf{h}\to\mathbf{0}} \frac{|\mathbf{r}(\mathbf{h})|}{|\mathbf{h}|} = 0 \tag{1.10}$$

(b) Suppose $\mathbf{f}$ and E are as in Definition 1.4.1 and $\mathbf{f}$ is differentiable in E. For every $\mathbf{x} \in E$, $\mathbf{f}\,'(x)$ is then a function, namely, a linear transformation of $\mathbb{R}^n$ into $\mathbb{R}^m$. But $\mathbf{f}\,'$ is also a function: $\mathbf{f}\,'$ maps E into $L(\mathbb{R}^n, \mathbb{R}^m)$.

(c) Equation 1.9 shows that $\mathbf{f}$ is continuous at any point at which f is differentiable.
(Hint:Every Linear transformation is continuous by Theorem 1.3.10 and $\lim_{\mathbf{h}\to 0} \mathbf{r}(\mathbf{h}) = \mathbf{0}$)

(d) The derivative defined by 1.7 or 1.9 is often called the *differential* of $\mathbf{f}$ at $\mathbf{x}$, or the *total derivative* of $\mathbf{f}$ at $\mathbf{x}$.

Theorem 1.4.2. Suppose E and $\mathbf{f}$ are as in Definition 1.4.1 and $\mathbf{x} \in E$ and equation 1.7 holds with $A = A_1$ and $A = A_2$ Then $A_1 = A_2$.

Proof. If $B = A_1 - A_2$. the inequality

$$|B\mathbf{h}| \le |\mathbf{f}(\mathbf{x}+\mathbf{h}) - \mathbf{f}(x) - A_1\mathbf{h}| + |\mathbf{f}(\mathbf{x}+\mathbf{h}) - \mathbf{f}(x) - A_2\mathbf{h}|$$

shows that $\dfrac{|B\mathbf{h}|}{|\mathbf{h}|} \to 0$ as $\mathbf{h} \to 0$.
For fixed $\mathbf{h} \neq \mathbf{0}$, $t\mathbf{h} \to \mathbf{0}$ as $t \to 0$ it follows that

$$\frac{|B(t\mathbf{h})|}{|t\mathbf{h}|} \to 0 \text{ as } \mathbf{t} \to 0. \tag{1.11}$$

The linearity of B shows that the left side of 1.11 is $\dfrac{|t||B(\mathbf{h})|}{|t||\mathbf{h}|} = \dfrac{|B(\mathbf{h})|}{|\mathbf{h}|}$ independent of t. Thus $B\mathbf{h} = \mathbf{0}$ for every $\mathbf{h} \in \mathbb{R}^n$. Hence $B = 0$. □

Example 1.4.3. If $A \in L(\mathbb{R}^n, \mathbb{R}^m)$ and $\mathbf{x} \in \mathbb{R}^n$ then $A'(\mathbf{x}) = A$.

Proof. Since $A(\mathbf{x}+\mathbf{h}) - A(\mathbf{x}) = A\mathbf{h}$ by the linearity of A. With $\mathbf{f}(\mathbf{x}) = A\mathbf{x}$, the numerator in equation 1.7 is thus $\mathbf{0}$ for every $\mathbf{h} \in \mathbb{R}^n$. □

Theorem 1.4.4. Suppose E is an open set in $\mathbb{R}^n$, $\mathbf{f}$ maps E into $\mathbb{R}^m$, $\mathbf{f}$ is differentiable at $\mathbf{x}_0 \in E$, $\mathbf{g}$ maps an open set containing $\mathbf{f}(E)$ into $\mathbb{R}^k$, and $\mathbf{g}$ is differentiable at $\mathbf{f}(\mathbf{x}_0)$. Then the mapping $\mathbf{F}$ of E into $\mathbb{R}^k$ defined by

$$\mathbf{F}(\mathbf{x}) = \mathbf{g}(\mathbf{f}(\mathbf{x}))$$

is differentiable at $\mathbf{x}_0$. and

$$\mathbf{F}\,'(\mathbf{x}_0) = \mathbf{g}\,'(\mathbf{f}(\mathbf{x}_0))\mathbf{f}\,'(\mathbf{x}_0).$$

Proof. Put $\mathbf{y}_0 = \mathbf{f}(\mathbf{x}_0), A = \mathbf{f}\,'(\mathbf{x}_0), B = \mathbf{g}\,'(\mathbf{y}_0)$ and define

$$\begin{aligned}\mathbf{u}(\mathbf{h}) &= \mathbf{f}(\mathbf{x}_0 + \mathbf{h}) - \mathbf{f}(\mathbf{x}_0) - A\mathbf{h}\\ \mathbf{v}(\mathbf{k}) &= \mathbf{g}(\mathbf{y}_0 + \mathbf{k}) - \mathbf{g}(\mathbf{y}_0) - B\mathbf{k}\end{aligned}$$

for all $\mathbf{h} \in \mathbb{R}^n$ and $\mathbf{k} \in \mathbb{R}^m$ for which $\mathbf{f}(\mathbf{x}_0+\mathbf{h})$ and $\mathbf{g}(\mathbf{y}_0+\mathbf{k})$ are defined. Then

$$|\mathbf{u}(\mathbf{h})| = \epsilon(\mathbf{h})|\mathbf{h}|, \quad |\mathbf{v}(\mathbf{k})| = \eta(\mathbf{k})|\mathbf{k}| \tag{1.12}$$

where $\epsilon(\mathbf{h}) \to 0$ as $\mathbf{h} \to \mathbf{0}$ and $\eta(\mathbf{k}) \to 0$ as $\mathbf{k} \to \mathbf{0}$
Given $\mathbf{h}$ put $\mathbf{k} = \mathbf{f}(\mathbf{x}_0 + \mathbf{h}) - \mathbf{f}(\mathbf{x}_0)$ Then

$$|\mathbf{k}| = |A\mathbf{h} + \mathbf{u}(\mathbf{h})| \leq [\|A\| + \epsilon(\mathbf{h})]|\mathbf{h}|, \tag{1.13}$$

and

$$\begin{aligned}\mathbf{F}(\mathbf{x}_0 + \mathbf{h}) - \mathbf{F}(\mathbf{x}_0) - BA\mathbf{h} &= \mathbf{g}(\mathbf{f}(\mathbf{x}_0 + \mathbf{h})) - \mathbf{g}(\mathbf{f}(\mathbf{x}_0)) - BA\mathbf{h}\\ &= \mathbf{g}(\mathbf{y}_0 + \mathbf{k}) - \mathbf{g}(\mathbf{y}_0) - BA\mathbf{h}\\ &= B(\mathbf{k} - A\mathbf{h}) + \mathbf{v}(\mathbf{k})\\ &= B\mathbf{u}(\mathbf{h}) + \mathbf{v}(\mathbf{k})\end{aligned}$$

Hence by equation 1.12 and equation 1.13 imply, for $\mathbf{h} \neq \mathbf{0}$, that

$$\frac{|\mathbf{F}(\mathbf{x}_0 + \mathbf{h}) - \mathbf{F}(\mathbf{x}_0) - BA\mathbf{h}|}{|\mathbf{h}|} \leq \|B\|\epsilon(\mathbf{h}) + [\|A\| + \epsilon(\mathbf{h})]\eta(\mathbf{k})$$

Let $\mathbf{h} \to \mathbf{0}$ Then $\epsilon(\mathbf{h}) \to 0$. Also , $\mathbf{k} \to \mathbf{0}$ by equation 1.13, so that $\eta(\mathbf{k}) \to 0$. So

$$\lim_{\mathbf{h}\to\mathbf{0}} \frac{|\mathbf{F}(\mathbf{x}_0 + \mathbf{h}) - \mathbf{F}(\mathbf{x}_0) - BA\mathbf{h}|}{|\mathbf{h}|} = 0$$

It follows that $\mathbf{F}\,'(\mathbf{x}_0) = BA$. □

Partial derivatives We again consider a function $\mathbf{f}$ that maps an open set $E \subset \mathbb{R}^n$ into $\mathbb{R}^m$. Let $\{\mathbf{e}_1, ..., \mathbf{e}_n\}$ and $\{\mathbf{u}_1, ..., \mathbf{u}_m\}$ be the standard bases of $\mathbb{R}^n$ and $\mathbb{R}^m$. The components of $\mathbf{f}$ are the real functions $f_1, ...f_m$ defined by

$$\mathbf{f}(x) = \sum_{i=1}^{m} f_i(\mathbf{x})\mathbf{u}_i \quad (\mathbf{x} \in E) \tag{1.14}$$

or, equivalently, by $f_i(\mathbf{x}) = \mathbf{f}(x) \cdot \mathbf{u}_i, 1 \leq i \leq m$
For $\mathbf{x} \in E, 1 \leq i \leq m, 1 \leq j \leq n$ we define

$$(D_j f_i)(\mathbf{x}) = \lim_{t \to 0} \frac{f_i(\mathbf{x} + t\mathbf{e}_j) - f_i(\mathbf{x})}{t} \tag{1.15}$$

provided the limit exists,is called *partial derivative* of f_i with respect to x_j.

Theorem 1.4.5. Suppose $\mathbf{f}$ maps an open set $E \subset \mathbb{R}^n$ into $\mathbb{R}^m$ and $\mathbf{f}$ is differentiable at a point $\mathbf{x} \in E$. Then the partial derivatives $(D_j f_i)(\mathbf{x})$ exist, and

$$\mathbf{f}\,'(\mathbf{x})\mathbf{e}_j = \sum_{i=1}^{m} (D_j f_i)(\mathbf{x})\mathbf{u}_i \quad (1 \leq j \leq n) \tag{1.16}$$

where $\{\mathbf{e}_1, , ..., \mathbf{e}_n\}$ and $\{\mathbf{u}_1, ..., \mathbf{u}_m\}$ be the standard bases of $\mathbb{R}^n$ and $\mathbb{R}^m$.

Proof. Fix j.Since $\mathbf{f}$ is differentiable at $\mathbf{x}$, by equation 1.9

$$\mathbf{f}(\mathbf{x} + t\mathbf{e}_j) - \mathbf{f}(\mathbf{x}) = \mathbf{f}\,'(\mathbf{x})(t\mathbf{e}_j) + \mathbf{r}(t\mathbf{e}_j)$$

where $\dfrac{|\mathbf{r}(t\mathbf{e}_j)|}{t} \to 0$ as $t \to 0$.The linearity of $\mathbf{f}\,'(\mathbf{x})$ shows that

$$\lim_{t \to 0} \frac{\mathbf{f}(\mathbf{x} + t\mathbf{e}_j) - \mathbf{f}(\mathbf{x})}{t} = \mathbf{f}\,'(x)\mathbf{e}_j \tag{1.17}$$

If we now represent $\mathbf{f}$ in terms of its components, as in equation 1.14, then equation 1.17 becomes

$$\lim_{t \to 0} \sum_{i=1}^{m} \frac{\mathbf{f}_i(\mathbf{x} + t\mathbf{e}_j) - \mathbf{f}_i(\mathbf{x})}{t} \mathbf{u}_i = \mathbf{f}\,'(\mathbf{x})\mathbf{e}_j \tag{1.18}$$

It follows that each quotient in this sum has a limit, as $t \to 0$ so that each $(D_j f_i)(\mathbf{x})$ exists, and then equation 1.16 follows from equation 1.18. □

Let $[\mathbf{f}\,'(\mathbf{x})]$ be the matrix of the linear transformation $\mathbf{f}\,'(\mathbf{x})$ with respect to our standard bases.

Then $\mathbf{f}'(\mathbf{x})\mathbf{e}_j$ is the j^{th} column vector of $[\mathbf{f}'(\mathbf{x})]$ and 1.16 shows therefore that the number $(D_j f_i)(\mathbf{x})$ occupies the spot in the i^{th} row and j^{th} column of $[\mathbf{f}'(\mathbf{x})]$ Thus

$$[\mathbf{f}'(\mathbf{x})] = \begin{bmatrix} (D_1 f_1)(\mathbf{x}) & \cdots & (D_n f_1)(\mathbf{x}) \\ \cdots & \cdots & \cdots \\ (D_1 f_m)(\mathbf{x}) & \cdots & (D_n f_m)(\mathbf{x}) \end{bmatrix}$$

If $\mathbf{h} = \sum h_j \mathbf{e}_j$ is any vector in $\mathbb{R}^n$ then equation 1.16 implies that

$$\mathbf{f}'(\mathbf{x})\mathbf{h} = \sum_{i=1}^{m} \left\{ \sum_{j=1}^{n} (D_j f_i)(\mathbf{x}) h_j \right\} \mathbf{u}_i$$

Example 1.4.6. From Theorem 1.4.5 , if $\mathbf{f}$ is differentiable at a point $\mathbf{x} \in E$ then all its partial derivative exists. Consider the function f on $\mathbb{R}^2$ by

$$f(x,y) = \begin{cases} 0 & \text{if } (x,y) = (0,0) \\ \dfrac{xy}{x^2+y^2} & \text{if } (x,y) \neq (0,0) \end{cases}$$

By using equation 1.15

$$(D_1 f)(0,0) = \lim_{t \to 0} \frac{f(t,0) - f(0,0)}{t} = 0$$

$$(D_2 f)(0,0) = \lim_{t \to 0} \frac{f(0,t) - f(0,0)}{t} = 0$$

However, $f(x,y)$ is not continuous at $(0,0)$, since if $(x,y) \to (0,0)$ along the line $y = x$, then $f(x,y) = \frac{1}{2}$ and if $(x,y) \to (0,0)$ along the $x-$ axis , then $f(x,y) = 0$. So $\lim\limits_{(x,y)\to(0,0)} f(x,y)$ does not exist.
So f is not differentiable at $(0,0)$.

Definition 1.4.7. Let f be a real valued differentiable function with domain E, Let $\mathbf{x} \in E$ then the gradient of f at $\mathbf{x}$, defined by

$$\nabla f(\mathbf{x}) = \sum_{i=1}^{n} (D_i f)(\mathbf{x}) \mathbf{e}_i$$

Definition 1.4.8. Let f be a real valued differentiable function with domain E. Fix an $\mathbf{x} \in E$, let $\mathbf{u} \in \mathbb{R}^n$ be a unit vector (that is, $|\mathbf{u}| = 1$), then

$$\lim_{t \to 0} \frac{f(\mathbf{x} + t\mathbf{u}) - f(\mathbf{x})}{t}$$

is called the directional derivative of f at $\mathbf{x}$, in the direction of the unit vector $\mathbf{u}$, and is denoted by $(D_{\mathbf{u}} f)(\mathbf{x})$.

By simple calculation we can show that

$$(D_{\mathbf{u}}f)(\mathbf{x}) = (\nabla f)(\mathbf{x}) \cdot \mathbf{u} \tag{1.19}$$

If f and $\mathbf{x}$ are fixed, but $\mathbf{u}$ varies, then 1.19 shows that $(D_{\mathrm{u}}f)(\mathbf{x})$ attains its maximum when $\mathbf{u}$ is a positive scalar multiple of $(\nabla f)(\mathbf{x})$.

If $\mathbf{u} = \Sigma u_i \mathbf{e}_i$, then 1.19 shows that $(D_{\mathbf{u}}f)(\mathbf{x})$ can be expressed in terms of the partial derivatives of f at $\mathbf{x}$ by the formula

$$(D_{\mathrm{u}}f)(\mathbf{x}) = \sum_{i=1}^{n} (D_i f)(\mathbf{x}) u_i.$$

Result 1.4.9. Suppose $\mathbf{f}$ is a continuous mapping of $[a, b]$ into $\mathbb{R}^k$ and $\mathbf{f}$ is differentiable in (a, b). Then there exist $x \in (a, b)$ such that

$$|\mathbf{f}(b) - \mathbf{f}(a)| \leq (b - a)|\mathbf{f}\,'(x)|$$

Theorem 1.4.10. Suppose $\mathbf{f}$ maps a convex open set $E \subset \mathbb{R}^n$ into $\mathbb{R}^m$, $\mathbf{f}$ is differentiable in E and there is a real mumber M such that

$$\|\mathbf{f}\,'(\mathbf{x})\| \leq M$$

for every $\mathbf{x} \in E$. Then

$$|\mathbf{f}(\mathbf{b}) - \mathbf{f}(\mathbf{a})| \leq M|\mathbf{b} - \mathbf{a}|$$

for all $\mathbf{a} \in E, \mathbf{b} \in E$.

Proof. Fix $\mathbf{a} \in E, \mathbf{b} \in E$. Define

$$\gamma(t) = (1 - t)\mathbf{a} + t\mathbf{b}$$

for all $t \in R^1$ such that $\gamma(t) \in E$. Since E is convex, $\gamma(t) \in E$ if $0 \leq t \leq 1$. Put

$$\mathbf{g}(t) = \mathbf{f}(\gamma(t))$$

Then

$$\mathbf{g}\,'(t) = \mathbf{f}\,'(\gamma(t))\gamma\,'(t) = \mathbf{f}\,'(\gamma(t))(\mathbf{b} - \mathbf{a})$$

so that

$$|\mathbf{g}\,'(t)| \leq \|\mathbf{f}\,'(\gamma(t))\|\,|\,\mathbf{b} - \mathbf{a}| \leq M|\mathbf{b} - \mathbf{a}|$$

for all $t \in [0, 1]$. By Result 1.4.9,

$$|\mathbf{g}(1) - \mathbf{g}(0)| \leq M|\mathbf{b} - \mathbf{a}|$$

But $\mathbf{g}(0) = \mathbf{f}(\mathbf{a})$ and $\mathbf{g}(1) = \mathbf{f}(\mathbf{b})$. This completes the proof. □

Corollary 1.4.11. If, $\mathbf{f}'(\mathbf{x}) = \mathbf{0}$ for all $\mathbf{x} \in E$, then $\mathbf{f}$ is constant.

Proof. To prove this, note that the hypotheses of the above theorem hold with $M = 0$. So

$$|\mathbf{f}(\mathbf{b}) - \mathbf{f}(\mathbf{a})| \leq \mathbf{0}|\mathbf{b} - \mathbf{a}|$$

for all $\mathbf{a} \in E, \mathbf{b} \in E$.
This imply that $|\mathbf{f}(\mathbf{b}) - \mathbf{f}(\mathbf{a})| = 0$. That is $\mathbf{f}(\mathbf{b}) = \mathbf{f}(\mathbf{a})$ for all $\mathbf{a} \in E, \mathbf{b} \in E$. □

Definition 1.4.12. A differentiable mapping $\mathbf{f}$ of an open set $E \subset \mathbb{R}^n$ into $\mathbb{R}^m$ is said to be *continuously differentiable* in E if $\mathbf{f}'$ is a continuous mapping of E into $L(\mathbb{R}^n, \mathbb{R}^m)$. More explicitly, Every $\mathbf{x} \in E$ and to every $\epsilon > 0$ corresponds a $\delta > 0$ such that

$$\|\mathbf{f}'(\mathbf{y}) - \mathbf{f}'(\mathbf{x})\| < \epsilon$$

if $\mathbf{y} \in E$ and $|\mathbf{x} - \mathbf{y}| < \delta$. If this is so, we also say that $\mathbf{f}$ is a $\mathscr{C}'$-mapping, or that $\mathbf{f} \in \mathscr{C}'(E)$.

Result 1.4.13. Mean Value theorem: If f is a real continuous function on $[a, b]$ which is differentiable in (a, b), then there is a point $x \in (a, b)$ at which

$$f(b) - f(a) = (b - a)f'(x)$$

Theorem 1.4.14. Suppose $\mathbf{f}$ maps an open set $E \subset \mathbb{R}^n$ into $\mathbb{R}^m$. Then $\mathbf{f} \in \mathscr{C}'(E)$ if and only if the partial derivatives $D_j f_i$ exist and are continuous on E for $1 \leq i \leq m$, $1 \leq j \leq n$.

Proof. Assume first that $\mathbf{f} \in \mathscr{C}'(E)$. By equation 1.16,

$$(D_j f_i)(\mathbf{x}) = (\mathbf{f}'(\mathbf{x})\mathbf{e}_j) \cdot \mathbf{u}_i$$

for all i, j, and for all $\mathbf{x} \in E$. Hence

$$(D_j f_i)(\mathbf{y}) - (D_j f_i)(\mathbf{x}) = \{[\mathbf{f}'(\mathbf{y}) - \mathbf{f}'(\mathbf{x})]\,\mathbf{e}_j\} \cdot \mathbf{u}_i$$

and since $|\mathbf{u}_i| = |\mathbf{e}_i| = 1$, it follows from result 1.1.3 (d) and exercise 1.3.7 that

$$\begin{aligned} |(D_j f_i)(\mathbf{y}) - (D_j f_i)(\mathbf{x})| &\leq |[\mathbf{f}'(\mathbf{y}) - \mathbf{f}'(\mathbf{x})]\,e_j| \\ &\leq \|f'(\mathbf{y}) - f'(\mathbf{x})\| \end{aligned}$$

$\mathbf{f} \in \mathscr{C}'(E)$. So , for every $\epsilon > 0$ corresponds a $\delta > 0$ such that

$$|(D_j f_i)(\mathbf{y}) - (D_j f_i)(\mathbf{x})| < \epsilon$$

if $\mathbf{y} \in E$ and $|\mathbf{x} - \mathbf{y}| < \delta$. Hence $D_j f_i$ is continuous.

For the converse, consider the case $m = 1$. Fix $x \in E$ and $\epsilon > 0$. Since E is open, there is an open ball $S \subset E$, with center at $\mathbf{x}$ and radius r, and the continuity of the functions $D_j f$ shows that r can be chosen so that

$$|(D_j f)(\mathbf{y}) - (D_j f)(\mathbf{x})| < \frac{\epsilon}{n} \quad (\mathbf{y} \in S, 1 \leq j \leq n) \tag{1.20}$$

Suppose $\mathbf{h} \in \mathbb{R}^n$ with $|\mathbf{h}| < r$. Let $\mathbf{h} = \sum h_j \mathbf{e}_j$, put $\mathbf{v}_0 = \mathbf{0}$, and $\mathbf{v}_k = h_1\mathbf{e}_1 + \cdots + h_k\mathbf{e}_k$, for $1 \leq k \leq n$. Then

$$f(\mathbf{x}+\mathbf{h}) - f(\mathbf{x}) = \sum_{j=1}^{n} [f(\mathbf{x}+\mathbf{v}_j) - f(\mathbf{x}+\mathbf{v}_{j-1})] \tag{1.21}$$

Since $|\mathbf{v}_k| < |\mathbf{h}| < r$ for $1 \leq k \leq n$ and since S is convex, the segments with end points $\mathbf{x}+\mathbf{v}_{j-1}$ and $\mathbf{x}+\mathbf{v}_j$ lie in S. Since $\mathbf{v}_j = \mathbf{v}_{j-1} + h_j\mathbf{e}_j$, the mean value theorem (Result 1.4.13) shows that the j th summand in 1.21 is equal to

$$h_j (D_j f)(\mathbf{x} + \mathbf{v}_{j-1} + \theta_j h_j \mathbf{e}_j)$$

for some $\theta_j \in (0,1)$, and this differs from $h_j (D_j f)(\mathbf{x})$ by less than $|h_j| \frac{\epsilon}{n}$ using 1.20.

[To understand this , let's consider the case $j = 2$.
Let $\mathbf{x} = (x_1, ...x_n), \mathbf{h} = (h_1, ..., h_n)$. Then $f(\mathbf{x}+\mathbf{v}_2) - f(\mathbf{x}+\mathbf{v}_1)$
$= f((x_1, x_2, ..., x_n) + (h_1, h_2, 0, ..., 0)) - f((x_1, ..., x_n) + (h_1, 0, 0, ..., 0))$
$= f(x_1 + h_1, x_2 + h_2, ..., x_n) - f(x_1 + h_1, x_2, ..., x_n)$. So consider the function $g_2 : [x_2, x_2 + h_2] \to \mathbb{R}$ by $g(t) = f(x_1 + h_1, t, x_3, ..., x_n)$ then by mean value theorem (Result 1.4.13) there exist $c_2 \in (x_2, x_2 + h_2)$ such that

$$g(x_2 + h_2) - g(x_2) = h_2 g'(c_2)$$

$c_2 \in (x_2, x_2 + h_2)$ imply that $c_2 = x_2 + \theta_2 h_2$ for some $\theta_2 \in (0, 1)$]

By equation 1.21, it follows that

$$\left| f(\mathbf{x}+\mathbf{h}) - f(\mathbf{x}) - \sum_{j=1}^{n} h_j (D_j f)(\mathbf{x}) \right| \leq \frac{1}{n} \sum_{j=1}^{n} |h_j| \, \epsilon \leq |\mathbf{h}|\epsilon$$

for all $\mathbf{h}$ such that $|\mathbf{h}| < r$

This says that f is differentiable at $\mathbf{x}$ and that $\mathbf{f}\,'(\mathbf{x})$ is the linear function which assigns the number $\Sigma h_j (D_j f)(\mathbf{x})$ to the vector $\mathbf{h} = \Sigma h_j \mathbf{e}_j$. The matrix $[f\,'(\mathbf{x})]$ consists of the row $(D_1 f)(\mathbf{x}), \ldots, (D_n f)(\mathbf{x})$; and since $D_1 f, \ldots, D_n f$ are continuous functions on E, by Theorem 1.3.14, $f \in \mathscr{C}\,'(E)$. □

1.5 The Contraction Principle

Definition 1.5.1. Let X be a metric space, with metric d. If φ maps X into X and if there is a number $c < 1$ such that

$$d(\varphi(x), \varphi(y)) \leq c\, d(x, y) \tag{1.22}$$

for all $x, y \in X$, then φ is said to be a contraction of X into X.

Theorem 1.5.2. If X is a complete metric space, and if φ is a contraction of X into X, then there exists one and only one $x \in X$ such that $\varphi(x) = x$.

Proof. The uniqueness is a triviality . If $\varphi(x) = x$ and $\varphi(y) = y$, then equation 1.22 gives $d(x, y) \leq c\, d(x, y)$, which happens only when $d(x, y) = 0$
Let's prove the existence part.

Pick $x_0 \in X$ arbitrarily, and define $\{x_n\}$ recursively, by setting

$$x_{n+1} = \varphi\left(x_n\right) \quad (n = 0, 1, 2, \ldots)$$

Choose $c < 1$ so that equation 1.22 holds. For $n \geq 1$ then we have

$$d\left(x_{n+1}, x_n\right) = d\left(\varphi\left(x_n\right), \varphi\left(x_{n-1}\right)\right) \leq c\, d\left(x_n, x_{n-1}\right)$$

Hence induction gives

$$d\left(x_{n+1}, x_n\right) \leq c^n\, d\left(x_1, x_0\right) \quad (n = 0, 1, 2, \ldots)$$

If $n < m$, it follows that,

$$\begin{aligned} d\left(x_n, x_m\right) &\leq \sum_{i=n+1}^{m} d\left(x_i, x_{i-1}\right) \\ &\leq \left(c^n + c^{n+1} + \cdots + c^{m-1}\right) d\left(x_1, x_0\right) \\ &\leq \left[(1-c)^{-1} d\left(x_1, x_0\right)\right] c^n \end{aligned}$$

Thus $\{x_n\}$ is a Cauchy sequence . Since X is complete, $\lim x_n = x$ for some $x \in X$.

Since φ is a contraction, φ is continuous on X. Hence

$$\varphi(x) = \lim_{n\to\infty} \varphi(x_n) = \lim_{n\to\infty} x_{n+1} = x.$$

□

1.6 The Inverse Function Theorem

Theorem 1.6.1. Suppose $\mathbf{f}$ is a $\mathscr{C}'-$ mapping of an open set $E \subset \mathbb{R}^n$ into $\mathbb{R}^m$, $\mathbf{f}'(\mathbf{a})$ is is invertible for some $\mathbf{a} \in E$ and $\mathbf{b} = \mathbf{f}(\mathbf{a})$. Then

(a) there exist open sets U and V in $\mathbb{R}^n$ such that $\mathbf{a} \in U, \mathbf{b} \in V, \mathbf{f}$ is one-to one on U and $\mathbf{f}(U) = V$

(b) if $\mathbf{g}$ is the inverse of $\mathbf{f}$ [which exists, by (a)], defined in V by

$$\mathbf{g}(\mathbf{f}(\mathbf{x})) = \mathbf{x} \quad (\mathbf{x} \in U)$$

then $g \in \mathscr{C}'(V)$.

Proof. (a) Put $\mathbf{f}'(\mathbf{a}) = A$, and choose λ so that

$$2\lambda \|A^{-1}\| = 1 \tag{1.23}$$

Since $\mathbf{f}'$ is continuous at $\mathbf{a}$, there is an open ball $U \subset E$, with center at $\mathbf{a}$, such that

$$\|\mathbf{f}'(\mathbf{x}) - A\| < \lambda \quad (\mathbf{x} \in U) \tag{1.24}$$

We associate to each $\mathbf{y} \in \mathbb{R}^n$ a function φ, defined by

$$\varphi(\mathbf{x}) = \mathbf{x} + A^{-1}(\mathbf{y} - \mathbf{f}(\mathbf{x})) \quad (\mathbf{x} \in E) \tag{1.25}$$

Note that $\mathbf{f}(\mathbf{x}) = \mathbf{y}$ if and only if $\mathbf{x}$ is a fixed point of φ.

Since $\varphi'(\mathbf{x}) = I - A^{-1}\mathbf{f}'(\mathbf{x}) = A^{-1}(A - \mathbf{f}'(\mathbf{x}))$,1.23 and 1.24 imply that

$$\|\varphi'(\mathbf{x})\| < \frac{1}{2} \quad (\mathbf{x} \in U)$$

Hence

$$|\varphi(\mathbf{x}_1) - \varphi(\mathbf{x}_2)| \leq \frac{1}{2}|\mathbf{x}_1 - \mathbf{x}_2| \quad (\mathbf{x}_1, \mathbf{x}_2 \in U) \tag{1.26}$$

by Theorem 1.4.10. It follows that φ has at most one fixed point in U, so that $\mathbf{f}(\mathbf{x}) = \mathbf{y}$ for at most one $\mathbf{x} \in U$. Thus f is $1-1$ in U.

Next, put $V = \mathbf{f}(U)$, and pick $\mathbf{y}_0 \in V$. Then $\mathbf{y}_0 = \mathbf{f}(\mathbf{x}_0)$ for some $\mathbf{x}_0 \in U$. Let B be an open ball with center at $\mathbf{x}_0$ and radius $r > 0$, so small that its closure $\bar{B}$ lies in U. We will show that $\mathbf{y} \in V$ whenever $|\mathbf{y} - \mathbf{y}_0| < \lambda r$. This proves,that V is open.

Fix $\mathbf{y}, |\mathbf{y} - \mathbf{y}_0| < \lambda x$. With φ as in 1.25.

$$|\varphi(\mathbf{x}_0) - \mathbf{x}_0| = \left|A^{-1}(\mathbf{y} - \mathbf{y}_0)\right| < \left\|A^{-1}\right\| \lambda r = \frac{r}{2}$$

If $\mathbf{x} \in \bar{B}$, it therefore follows from 1.26 that

$$\begin{aligned} |\varphi(\mathbf{x}) - \mathbf{x}_0| &\leq |\varphi(\mathbf{x}) - \varphi(\mathbf{x}_0)| + |\varphi(\mathbf{x}_0) - \mathbf{x}_0| \\ &< \frac{1}{2}|\mathbf{x} - \mathbf{x}_0| + \frac{r}{2} \leq r; \end{aligned}$$

hence $\varphi(\mathbf{x}) \in B$. Note that 1.26 holds if $\mathbf{x}_1 \in \bar{B}, \mathbf{x}_2 \in \bar{B}$.

Thus φ is a contraction of $\bar{B}$ into $\bar{B}$. Being a closed subset of $\mathbb{R}^n$, $\bar{B}$ is complete. Theorem 1.5.2 implies therefore that φ has a fixed point $\mathbf{x} \in \bar{B}$. For this $\mathbf{x}, \mathbf{f}(\mathbf{x}) = \mathbf{y}$. Thus $\mathbf{y} \in \mathbf{f}(\bar{B}) \subset f(U) = V$

This proves part (a) of the theorem.

(b) Pick $\mathbf{y} \in V, \mathbf{y} + \mathbf{k} \in V$. Then there exist $\mathbf{x} \in U, \mathbf{x} + \mathbf{h} \in U$ so that $\mathbf{y} = \mathbf{f}(\mathbf{x}), \mathbf{y} + \mathbf{k} = \mathbf{f}(\mathbf{x} + \mathbf{h})$. With φ as in 1.25,

$$\varphi(\mathbf{x} + \mathbf{h}) - \varphi(\mathbf{x}) = \mathbf{h} + A^{-1}[\mathbf{f}(\mathbf{x}) - \mathbf{f}(\mathbf{x} + \mathbf{h})] = \mathbf{h} - A^{-1}\mathbf{k}$$

By 1.26, $|\mathbf{h} - A^{-1}\mathbf{k}| \leq \frac{1}{2}|\mathbf{h}|$. Hence $|A^{-1}\mathbf{k}| \geq \frac{1}{2}|\mathbf{h}|$, and

$$|\mathbf{h}| \leq 2\left\|A^{-1}\right\| |\mathbf{k}| = \lambda^{-1}|\mathbf{k}| \tag{1.27}$$

By 1.23, 1.24, and Theorem 1.3.11, $\mathbf{f}'(\mathbf{x})$ has an inverse, say T. Since

$$\mathbf{g}(\mathbf{y} + \mathbf{k}) - \mathbf{g}(\mathbf{y}) - T\mathbf{k} = \mathbf{h} - T\mathbf{k} = -T\left[\mathbf{f}(\mathbf{x} + \mathbf{h}) - \mathbf{f}(\mathbf{x}) - \mathbf{f}'(\mathbf{x})\mathbf{h}\right]$$

1.27 implies

$$\frac{|\mathbf{g}(\mathbf{y} + \mathbf{k}) - \mathbf{g}(\mathbf{y}) - T\mathbf{k}|}{|\mathbf{k}|} \leq \frac{\|T\|}{\lambda} \cdot \frac{|\mathbf{f}(\mathbf{x} + \mathbf{h}) - \mathbf{f}(\mathbf{x}) - \mathbf{f}'(\mathbf{x})\mathbf{h}|}{|\mathbf{h}|}$$

As $\mathbf{k} \to \mathbf{0}$,1.27 shows that $\mathbf{h} \to 0$. The right side of the last inequality thus tends to 0 . Hence the same is true of the left. We have thus proved that $\mathbf{g}'(\mathbf{y}) = T$. But T was chosen to be the inverse of $\mathbf{f}'(x) = \mathbf{f}'(\mathbf{g}(\mathbf{y}))$. Thus

$$\mathbf{g}'(\mathbf{y}) = \{\mathbf{f}'(\mathbf{g}(\mathbf{y}))\}^{-1} \quad (\mathbf{y} \in V) \tag{1.28}$$

Finally, note that g is a continuous mapping of V onto U (since $\mathbf{g}$ is differentiable), that $\mathbf{f}'$ is a continuous mapping of U into the set Ω of all invertible elements of $L(\mathbb{R}^n)$, and that inversion is a continuous mapping of Ω onto Ω, by Theorem 1.3.11. If we combine these facts with equation 1.28, we see that $g \in \mathscr{C}'(V)$. This completes the proof.

□

Theorem 1.6.2. If f is a $\mathscr{C}'-$ mapping of an open set $E \subset \mathbb{R}^n$ into $\mathbb{R}^m$ and if $\mathbf{f}'(\mathbf{x})$ is invertible for every $\mathbf{x} \in E$, then $\mathbf{f}(W)$ is an open subset of $\mathbb{R}^n$ for every open set $W \subset E$.

1.7 The Implicit Function Theorem

Notation 1.7.1. If $\mathbf{x} = (x_1, \ldots, x_n) \in \mathbb{R}^n$ and $\mathbf{y} = (y_1, \ldots, y_m) \in \mathbb{R}^m$, let us write $(\mathbf{x}, \mathbf{y})$ for the point (or vector)

$$(x_1, \ldots, x_n, y_1, \ldots, y_m) \in \mathbb{R}^{n+m}$$

That is, the first entry in $(\mathbf{x}, \mathbf{y})$ will always be a vector in $\mathbb{R}^n$ and the second will be a vector in $\mathbb{R}^m$.

Every $A \in L(\mathbb{R}^{n+m}, \mathbb{R}^n)$ can be split into two linear transformations A_x and A_y, defined by

$$A_x\mathbf{h} = A(\mathbf{h}, \mathbf{0}), \quad A_y\mathbf{k} = A(\mathbf{0}, \mathbf{k})$$

for any $\mathbf{h} \in \mathbb{R}^n, \mathbf{k} \in \mathbb{R}^m$. Then $A_x \in L(\mathbb{R}^n), A_y \in L(\mathbb{R}^m, \mathbb{R}^n)$, and

$$A(\mathbf{h}, \mathbf{k}) = A_x\mathbf{h} + A_y\mathbf{k}. \tag{1.29}$$

Theorem 1.7.2. Linear version of the implicit function theorem.
If $A \in L(\mathbb{R}^{n+m}, \mathbb{R}^n)$ and if A_x is invertible, then there corresponds to every $\mathbf{k} \in \mathbb{R}^m$ a unique $\mathbf{h} \in \mathbb{R}^n$ such that $A(\mathbf{h}, \mathbf{k}) = \mathbf{0}$. This $\mathbf{h}$ can be computed from $\mathbf{k}$ by the formula

$$\mathbf{h} = -(A_x)^{-1} A_y\mathbf{k} \tag{1.30}$$

Proof. By 1.29,

$$A(\mathbf{h}, \mathbf{k}) = 0$$

if and only if

$$A_x\mathbf{h} + A_y\mathbf{k} = \mathbf{0}$$

which is the same as 1.30 when A_x is invertible. □

Theorem 1.7.3. Implicit function theorem.
Let f be a $\mathscr{C}'$-mapping of an open set $E \subset \mathbb{R}^{n+m}$ into $\mathbb{R}^n$, such that $\mathbf{f}(\mathbf{a}, \mathbf{b}) = \mathbf{0}$ for some point $(\mathbf{a}, \mathbf{b}) \in E$. Put $A = \mathbf{f}'(\mathbf{a}, \mathbf{b})$ and assume that A_x is invertible.

Then there exist open sets $U \subset \mathbb{R}^{n+m}$ and $W \in \mathbb{R}^m$, with $(\mathbf{a}, \mathbf{b}) \in U$ and $\mathbf{b} \in W$, having the following property:

To every $\mathbf{y} \in W$ corresponds a unique $\mathbf{x}$ such that

$$(\mathbf{x}, \mathbf{y}) \in U \quad \text{and} \quad \mathbf{f}(\mathbf{x}, \mathbf{y}) = 0$$

If this $\mathbf{x}$ is defined to be $\mathbf{g}(\mathbf{y})$, then $\mathbf{g}$ is a $\mathscr{C}'$-mapping of W into $\mathbb{R}^n, \mathbf{g}(\mathbf{b}) = \mathbf{a}$,

$$\mathbf{f}(\mathbf{g}(\mathbf{y}), \mathbf{y}) = 0 \quad (\mathbf{y} \in W) \tag{1.31}$$

and

$$\mathbf{g}'(\mathbf{b}) = -(-A_x)^{-1} A_y \tag{1.32}$$

Proof. Define $\mathbf{F}$ by

$$\mathbf{F}(\mathbf{x},\mathbf{y}) = (\mathbf{f}(\mathbf{x},\mathbf{y}),\mathbf{y}) \quad ((\mathbf{x},\mathbf{y}) \in E) \tag{1.33}$$

Then $\mathbf{F}$ is a $\mathscr{C}'$-mapping of E into $\mathbb{R}^{n+m}$. We claim that $\mathbf{F}'(\mathbf{a},\mathbf{b})$ is an invertible element of $L(\mathbb{R}^{n+m})$:

Since $\mathbf{f}(\mathbf{a},\mathbf{b}) = \mathbf{0}$, we have

$$\mathbf{f}(\mathbf{a}+\mathbf{h},\mathbf{b}+\mathbf{k}) = A(\mathbf{h},\mathbf{k}) + \mathbf{r}(\mathbf{h},\mathbf{k})$$

where $\mathbf{r}$ is the remainder that occurs in the definition of $\mathbf{f}'(\mathbf{a},\mathbf{b})$, Since

$$\begin{aligned}\mathbf{F}(\mathbf{a}+\mathbf{h},\mathbf{b}+\mathbf{k}) - \mathbf{F}(\mathbf{a},\mathbf{b}) &= (\mathbf{f}(\mathbf{a}+\mathbf{h},\mathbf{b}+\mathbf{k}),\mathbf{k}) \\ &= (\mathbf{A}(\mathbf{h},\mathbf{k}),\mathbf{k}) + (\mathbf{r}(\mathbf{h},\mathbf{k}),0)\end{aligned}$$

it follows that $\mathbf{F}'(\mathbf{a},\mathbf{b})$ is the linear operator on $\mathbb{R}^{n+m}$ that maps $(\mathbf{h},\mathbf{k})$ to $(A(\mathbf{h},\mathbf{k}),\mathbf{k})$. If this image vector is $\mathbf{0}$, then $A(\mathbf{h},\mathbf{k}) = \mathbf{0}$ and $\mathbf{k} = \mathbf{0}$, hence $A(\mathbf{h},\mathbf{0}) = \mathbf{0}$, and Theorem 1.7.2 implies that $\mathbf{h} = \mathbf{0}$. It follows that $\mathbf{F}'(\mathbf{a},\mathbf{b})$ is 1-1; hence it is invertible (Theorem 1.3.5).

The inverse function theorem can therefore be applied to $\mathbf{F}$. It shows that there exist open sets U and V in $\mathbb{R}^{n+m}$, with $(\mathbf{a},\mathbf{b}) \in U, (\mathbf{0},\mathbf{b}) \in V$ such that $\mathbf{F}$ is a 1-1 mapping of U onto V.

We let W be the set of all $\mathbf{y} \in \mathbb{R}^m$ such that $(\mathbf{0},\mathbf{y}) \in V$. Note that $\mathbf{b} \in W$. Consider the function $L : \mathbb{R}^m \to \mathbb{R}^{n+m}$ by $L(\mathbf{y}) = (\mathbf{0},\mathbf{y})$. Then L is continuous and $L^{-1}(V) = W$. So W is open .

If $\mathbf{y} \in W$, then $(\mathbf{0},\mathbf{y}) = \mathbf{F}(\mathbf{x},\mathbf{y})$ for some $(\mathbf{x},\mathbf{y}) \in U$. By 1.33, $\mathbf{f}(\mathbf{x},\mathbf{y}) = \mathbf{0}$ for this $\mathbf{x}$.

Suppose, with the same $\mathbf{y}$, that $(\mathbf{x}',\mathbf{y}) \in U$ and $\mathbf{f}(\mathbf{x}',\mathbf{y}) = \mathbf{0}$. Then

$$\mathbf{F}(\mathbf{x}',\mathbf{y}) = (\mathbf{f}(\mathbf{x}',\mathbf{y}),\mathbf{y}) = (\mathbf{f}(\mathbf{x},\mathbf{y}),\mathbf{y}) = \mathbf{F}(\mathbf{x},\mathbf{y})$$

Since $\mathbf{F}$ is 1-1 in U, it follows that $\mathbf{x}' = \mathbf{x}$. This proves the first part of the theorem.

For the second part, define $\mathbf{g}(\mathbf{y})$, for $\mathbf{y} \in W$, so that $(\mathbf{g}(\mathbf{y}),\mathbf{y}) \in U$ and 1.31 holds. Then

$$\mathbf{F}(\mathbf{g}(\mathbf{y}),\mathbf{y}) = (\mathbf{0},\mathbf{y}) \quad (\mathbf{y} \in W) \tag{1.34}$$

If $\mathbf{G}$ is the mapping of V onto U that inverts $\mathbf{F}$, then $\mathbf{G} \in \mathscr{C}'$, by the inverse function theorem, and 1.34 gives

$$(\mathbf{g}(\mathbf{y}),\mathbf{y}) = \mathbf{G}(\mathbf{0},\mathbf{y}) \quad (\mathbf{y} \in W) \tag{1.35}$$

Since $G \in \mathscr{C}'$,1.35 shows that $\mathbf{g} \in \mathscr{C}'$.

Finally, to compute $\mathbf{g}'(\mathbf{b})$, put $(\mathbf{g}(\mathbf{y}), \mathbf{y}) = \Phi(\mathbf{y})$. Then

$$\Phi'(\mathbf{y})\mathbf{k} = (\mathbf{g}'(\mathbf{y})\mathbf{k}, \mathbf{k}) \quad (\mathbf{y} \in W, \mathbf{k} \in \mathbb{R}^m) \tag{1.36}$$

By 1.31, $\mathbf{f}(\Phi(\mathbf{y})) = \mathbf{0}$ in W. The chain rule shows therefore that

$$\mathbf{f}'(\Phi(\mathbf{y}))\Phi'(\mathbf{y}) = 0$$

When $\mathbf{y} = \mathbf{b}$, then $\Phi(\mathbf{y}) = (\mathbf{a}, \mathbf{b})$, and $\mathbf{f}'(\Phi(\mathbf{y})) = A$. Thus

$$A\Phi'(\mathbf{b}) = 0 \tag{1.37}$$

It now follows from 1.37,1.36, and 1.29, that

$$A_x\mathbf{g}'(\mathbf{b})\mathbf{k} + A_y\mathbf{k} = A\left(\mathbf{g}'(\mathbf{b})\mathbf{k}, \mathbf{k}\right) = A\Phi'(\mathbf{b})\mathbf{k} = \mathbf{0}$$

for every $\mathbf{k} \in \mathbb{R}^m$. Thus

$$A_x\mathbf{g}'(\mathbf{b}) + A_y = 0$$

This is equivalent to 1.32, and completes the proof. □

Example 1.7.4. Take $n = 2, m = 3$, and consider the mapping $\mathbf{f} = (f_1, f_2)$ of $\mathbb{R}^5$ into $\mathbb{R}^2$ given by

$$\begin{aligned} f_1\left(x_1, x_2, y_1, y_2, y_3\right) &= 2e^{x_1} + x_2y_1 - 4y_2 + 3 \\ f_2\left(x_1, x_2, y_1, y_2, y_3\right) &= x_2\cos x_1 - 6x_1 + 2y_1 - y_3 \end{aligned}$$

If $\mathbf{a} = (0, 1)$ and $\mathbf{b} = (3, 2, 7)$, then $\mathbf{f}(\mathbf{a}, \mathbf{b}) = 0$. With respect to the standard bases, the matrix of the transformation $A = \mathbf{f}'(\mathbf{a}, \mathbf{b})$ is

$$[A] = \begin{bmatrix} 2 & 3 & 1 & -4 & 0 \\ -6 & 1 & 2 & 0 & -1 \end{bmatrix}$$

Hence

$$[A_x] = \begin{bmatrix} 2 & 3 \\ -6 & 1 \end{bmatrix}, \quad [A_y] = \begin{bmatrix} 1 & -4 & 0 \\ 2 & 0 & -1 \end{bmatrix}.$$

We see that the column vectors of $[A_x]$ are independent. Hence A_x is invertible and the implicit function theorem asserts the existence of a $\mathscr{C}'$-mapping $\mathbf{g}$, defined in a neighborhood of $(3, 2, 7)$, such that $\mathbf{g}(3, 2, 7) = (0, 1)$ and $\mathbf{f}(\mathbf{g}(\mathbf{y}), \mathbf{y}) = \mathbf{0}$. We can use 1.32 to compute $\mathbf{g}'(3, 2, 7)$: Since

$$\left[(A_x)^{-1}\right] = [A_x]^{-1} = \frac{1}{20}\begin{bmatrix} 1 & -3 \\ 6 & 2 \end{bmatrix}$$

1.32 gives

$$[\mathbf{g}'(3,2,7)] = -\frac{1}{20}\begin{bmatrix} 1 & -3 \\ 6 & 2 \end{bmatrix}\begin{bmatrix} 1 & -4 & 0 \\ 2 & 0 & -1 \end{bmatrix} = \begin{bmatrix} \frac{1}{4} & \frac{1}{5} & -\frac{3}{20} \\ -\frac{1}{2} & \frac{6}{5} & \frac{1}{10} \end{bmatrix}$$

In terms of parial derivatives, the conclusion is that

$$\begin{array}{lll} D_1 g_1 = \frac{1}{4} & D_2 g_1 = \frac{1}{5} & D_3 g_1 = -\frac{3}{20} \\ D_1 g_2 = -\frac{1}{2} & D_2 g_2 = \frac{6}{5} & D_3 g_2 = \frac{1}{10} \end{array}$$

at the point $(3, 2, 7)$

1.8 Determinants

Definition 1.8.1. If $(j_1, \ldots, j_n)$ is an ordered n-tuple of integers, define

$$s(j_1, \ldots, j_n) = \prod_{p<q} \operatorname{sgn}(j_q - j_p) \tag{1.38}$$

where $\operatorname{sgn} x = 1$ if $x > 0, \operatorname{sgn} x = -1$ if $x < 0$, sgn $x = 0$ if $x = 0$. Then $s(j_1, \ldots, j_n) = 1, -1$, or 0, and it changes sign if any two of the j's are interchanged.

Example 1.8.2.

(a) $s(2,3,1) = \operatorname{sgn}(1-3).\operatorname{sgn}(1-2).\operatorname{sgn}(3-2) = \operatorname{sgn}(-2).\operatorname{sgn}(-1).\operatorname{sgn}(1) = (-1).(-1).1 = 1$

(b) $s(2,3,2) = 0$

Let $[A]$ be the matrix of a linear operator A on $\mathbb{R}^n$, relative to the standard basis $\{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$, with entries $a(i,j)$ in the ith row and j th column. The determinant of $[A]$ is defined to be the number

$$\det[A] = \sum s(j_1, \ldots, j_n)\, a(1, j_1)\, a(2, j_2) \cdots a(n, j_n) \tag{1.39}$$

The sum in 1.39 extends over all ordered n-tuples of integers $(j_1, \ldots, j_n)$ with $1 \le j_r \le n$

The column vectors $\mathbf{x}_j$ of $[A]$ are

$$\mathbf{x}_j = \sum_{i=1}^{n} a(i,j)\mathbf{e}_i \quad (1 \le j \le n) \tag{1.40}$$

It will be convenient to think of det $[A]$ as a function of the column vectors of $[A]$, If we write

$$\det(\mathbf{x}_1, \ldots, \mathbf{x}_n) = \det[A]$$

det is now a real function on the set of all ordered n-tuples of vectors in $\mathbb{R}^n$.

Example 1.8.3.

$$[A] = \begin{bmatrix} a(1,1) & a(1,2) \\ a(2,1) & a(2,2) \end{bmatrix}$$

then $\det[A] =$

$s(1,1)a(1,1)a(2,1) + s(1,2)a(1,1)a(2,2) + s(2,1)a(1,2)a(2,1) + s(2,2)a(1,2)a(2,2)$
$= a(1,1)a(2,2) - a(1,2)a(2,1)$

Theorem 1.8.4.

(a) If I is the identity operator on $\mathbb{R}^n$, then

$$\det[I] = \det(\mathbf{e}_1, \ldots, \mathbf{e}_n) = 1.$$

(b) det is a linear function of each of the colunn vectors $\mathbf{x}_j$, if the others are held fixed.
That is $\det(\mathbf{x}_1, ..., c\mathbf{x}_j, ..., \mathbf{x}_n) = c\det(\mathbf{x}_1, ..., \mathbf{x}_j, ..., \mathbf{x}_n)$ and

$$\det(\mathbf{x}_1, ..., \mathbf{x}_j + \mathbf{y}_j, ..., \mathbf{x}_n) = \det(\mathbf{x}_1, ..., \mathbf{x}_j, ..., \mathbf{x}_n) + \det(\mathbf{x}_1, ..., \mathbf{y}_j, ..., \mathbf{x}_n)$$

(c) If $[A]_1$ is obtained from $[A]$ by interchanging two columns, then

$$\det[A]_1 = -\det[A]$$

(d) If $[A]$ has two equal columns, then $\det[A] = 0$.

Proof.

(a) If $A = I$ then $a(i,i) = 1$ and $a(i,j) = 0$ for $i \neq j$. Hence

$$\det[I] = s(1,2,\ldots,n)a(1,1)a(2,2)...a(n,n) = 1.$$

(b) By 1.38, $s(j_1, \ldots, j_n) = 0$ if any two of the j's are equal. Each of the remaining $n!$ products in 1.39 contains exactly one factor from each column. This proves (b).

(c) It is an immediate consequence of the fact that $s(j_1, \ldots, j_n)$ changes sign if any two of the j's s are interchanged.

(d) It is a corollary of (c).

□

Theorem 1.8.5. If $[A]$ and $[B]$ are n by n matrices, then

$$\det([B][A]) = \det[B]\det[A].$$

Proof. If $\mathbf{x}_1, \ldots, \mathbf{x}_n$ are the columns of $[A]$, define

$$\Delta_B(\mathbf{x}_1, \ldots, \mathbf{x}_n) = \Delta_B[A] = \det([B][A] \tag{1.41}$$

The columns of $[B][A]$ are the vectors $B\mathbf{x}_1, \ldots, B\mathbf{x}_n$. Thus

$$\Delta_B(\mathbf{x}_1, \ldots, \mathbf{x}_n) = \det(B\mathbf{x}_1, \ldots, B\mathbf{x}_n) \tag{1.42}$$

By 1.42 and Theorem 1.8.4, Δ_B also has properties 1.8.4(b) to (d). By (b) and 1.40,

$$\Delta_B[A] = \Delta_B\left(\sum_i a(i,1)\mathbf{e}_i, \mathbf{x}_2, \ldots, \mathbf{x}_n\right) = \sum_i a(i,1)\Delta_B(\mathbf{e}_i, \mathbf{x}_2, \ldots, \mathbf{x}_n)$$

Repeating this process with $\mathbf{x}_2, \ldots, \mathbf{x}_n$, we obtain

$$\Delta_B[A] = \sum a(i_1,1)\, a(i_2,2) \cdots a(i_n,n)\, \Delta_B(\mathbf{e}_{i_1}, \ldots, \mathbf{e}_{i_n}) \tag{1.43}$$

the sum being extended over all ordered n-tuples $(i_1, \ldots, i_n)$ with $1 \leq i_r, \leq n$. by (c) and (d)

$$\Delta_B(\mathbf{e}_{i_1}, \ldots, \mathbf{e}_{i_n}) = t(i_1, \ldots, i_n)\, \Delta_B(\mathbf{e}_1, \ldots, \mathbf{e}_n) \tag{1.44}$$

where $t = 1, 0$, or -1, and since $[B][I] = [B]$,1.41 shows that

$$\Delta_B(\mathbf{e}_1, \ldots, \mathbf{e}_n) = \det[B] \tag{1.45}$$

Substituting 1.45 and 1.44 into 1.43, we obtain

$$\det([B][A]) = \left\{\sum a(i_1,1) \cdots a(i_n,n)\, t(i_1, \ldots, i_n)\right\} \det[B]$$

for all n by n matrices $[A]$ and $[B]$. Taking $B = I$, we see that the above sum in braces is $\det[A]$. This proves the theorem. □

Theorem 1.8.6. A linear operator A on $\mathbb{R}^n$ is invertible if and only if $\det[A] \neq 0$.

Proof. If A is invertible, Theorem 1.8.5 shows that

$$\det[A]\det\left[A^{-1}\right] = \det\left[AA^{-1}\right] = \det[I] = 1,$$

so that $\det[A] \neq 0$.

If A is not invertible, the columns $\mathbf{x}_1, \ldots, \mathbf{x}_n$ of $[A]$ are dependent (Theorem 1.3.5); hence there is one, say, x_k, such that

$$\mathbf{x}_k + \sum_{j \neq k} c_j \mathbf{x}_j = 0 \tag{1.46}$$

for certain scalars c_j. By Theorem 1.8.4(b) and (d), $\mathbf{x}_k$ can be replaced by $\mathbf{x}_k + c_j\mathbf{x}_j$ without altering the determinant, if $j \neq k$. Repeating, we see that $\mathbf{x}_k$ can be replaced by the left side of 1.46, i.e., by $\mathbf{0}$,without altering the determinant. But a matrix which has $\mathbf{0}$ for one column has determinant $\mathbf{0}$. Hence $\det[A] = 0$. □

Theorem 1.8.7. Determinant of a linear operator A on $\mathbb{R}^n$ is independent from the choices of basis on $\mathbb{R}^n$

Proof. Suppose $\{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$ and $\{\mathbf{u}_1, \ldots, \mathbf{u}_n\}$ are bases in $\mathbb{R}^n$. Every linear operator A on $\mathbb{R}^n$ determines matrices $[A]$ and $[A]_U$, with entries a_{ij} and α_{ij}, glven by

$$A\mathbf{e}_j = \sum_i a_{ij}\mathbf{e}_i, \quad A\mathbf{u}_j = \sum_i \alpha_{ij}\mathbf{u}_i.$$

If $\mathbf{u}_j = B\mathbf{e}_j = \sum b_{ij}\mathbf{e}_j$, then the matix $[B]$ is invertible and $A\mathbf{u}_j$ is equal to

$$\sum_k \alpha_{kj} B\mathbf{e}_k = \sum_k \alpha_{kj} \sum_i b_{ik}\mathbf{e}_i = \sum_i \left(\sum_k b_{ik}\alpha_{kj} \right) \mathbf{e}_i$$

and also to

$$AB\mathbf{e}_j = A\sum_k b_{kj}\mathbf{e}_k = \sum_i \left(\sum_k a_{ik}b_{kj} \right) \mathbf{e}_i.$$

Thus $\Sigma b_{ik}\alpha_{kj} = \Sigma a_{ik}b_{kj}$, or

$$[B][A]_U = [A][B] \tag{1.47}$$

Since B is invertible, $\det[B] \neq 0$. Hence 1.47, combined with Theorem 1.8.5, shows that

$$\det[A]_U = \det[A].$$

□

Module 2

Curves and Surfaces

2.1 What is a curve

Definition 2.1.1. For a function $f(x, y)$,

$$\mathcal{C} = \{(x, y) \in \mathbb{R}^2 | f(x, y) = c\}$$

where c is a constant is called level curves of f.

Definition 2.1.2. A parametrized curve in $\mathbb{R}^n$ is a map $\gamma : (\alpha, \beta) \to \mathbb{R}^n$ for some α, β, with $-\infty \leq \alpha \leq \beta \leq \infty$.

The symbol (α, β) denotes the open interval $(\alpha, \beta) = \{t \in \mathbb{R} \mid \alpha < t < \beta\}$.

A parametrized curve, whose image is contained in a level curve $\mathcal{C}$, is called a parametrization of (part of) $\mathcal{C}$.

Example 2.1.3. Parametrization $\gamma(t) = (\gamma_1(t), \gamma_2(t))$, of the parabola $y = x^2$ must satisfy $\gamma_2(t) = \gamma_1(t)^2$.
Take $\gamma_1(t) = t$ then $\gamma_2(t) = t^2$. To get every point on the parabola we must allow t to take every real number value. So a parametrization of the parabola $y = x^2$ is

$$\gamma : (-\infty, \infty) \to \mathbb{R}^2, \gamma(t) = (t, t^2).$$

But this is not the only parametrization of the parabola. Another choice is $\gamma(t) = (t^3, t^6)$. Yet another is $(2t, 4t^2)$, and of course there are (infinitely many) others. So the parametrization of a given level curve is not unique.

Example 2.1.4. Now Let's try to find the parametrization the circle $x^2 + y^2 = 1$. If we take $x = t$ as in the previous example, we will get

$y = \sqrt{1-t^2}$ (we could have taken $y = -\sqrt{1-t^2}$). So we get the parametrization

$$\gamma(t) = \left(t, \sqrt{1-t^2}\right)$$

But this is only a parametrization of the upper half of the circle because $\sqrt{1-t^2}$ is always ≥ 0. Similarly, if we had taken $y = -\sqrt{1-t^2}$, we would only have covered the lower half of the circle. If we want a parametrization of the whole circle, we must try again. We need functions $\gamma_1(t)$ and $\gamma_2(t)$ such that

$$\gamma_1(t)^2 + \gamma_2(t)^2 = 1$$

for all $t \in (\alpha, \beta)$, and such that every point on the circle is equal to $(\gamma_1(t), \gamma_2(t))$ for some $t \in (\alpha, \beta)$. Let $\gamma_1(t) = \cos t$ and $\gamma_2(t) = \sin t$ (since $\cos^2 t + \sin^2 t = 1$ for all values of t). We can take $(\alpha, \beta) = (-\infty, \infty)$, although this is overkill: any open interval (α, β) whose length is greater than 2π will be sufficient.

Example 2.1.5. Consider the parametrized curve (called an *Astroid*)

$$\gamma(t) = \left(\cos^3 t, \sin^3 t\right), \quad t \in \mathbb{R}$$

Since $\cos^2 t + \sin^2 t = 1$ for all t, the coordinates $x = \cos^3 t, y = \sin^3 t$ of the point $\gamma(t)$ satisfy

$$x^{2/3} + y^{2/3} = 1$$

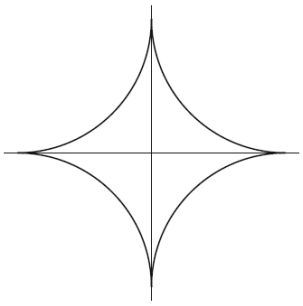


Figure 2.1: Astroid

Definition 2.1.6. A function $f : (\alpha, \beta) \to \mathbb{R}$ is said to be smooth if the derivative $\dfrac{d^n f}{dt^n}$ exists for all $n \geq 1$ and all $t \in (a, \beta)$.

If $f(t)$ and $g(t)$ are smooth functions, it follows from standard results of calculus that the sum $f(t) + g(t)$, product $f(t)g(t)$, quotient $f(t)/g(t)$, and composite $f(g(t))$ are smooth functions, where they are defined.

To differentiate a vector-valued function such as $\gamma(t)$ (as in Definition 2.1.2), we differentiate component wise: if

$$\gamma(t) = (\gamma_1(t), \gamma_2(t), \ldots, \gamma_n(t)),$$

then

$$\frac{d\gamma}{dt} = \left(\frac{d\gamma_1}{dt}, \frac{d\gamma_2}{dt}, \ldots, \frac{d\gamma_n}{dt}\right), \frac{d^2\gamma}{dt^2} = \left(\frac{d^2\gamma_1}{dt^2}, \frac{d^2\gamma_2}{dt^2}, \ldots, \frac{d^2\gamma_n}{dt^2}\right), \text{ etc.}$$

We often denote $\dfrac{d\gamma}{dt}$ by $\dot{\gamma}(t)$, $\dfrac{d^2\gamma}{dt^2}$ by $\ddot{\gamma}(t)$, etc. We say that γ is smooth if the derivative $\dfrac{d^n\gamma}{dt^n}$ exist for all $n \geq 1$ and all $t \in (\alpha, \beta)$; this is equivalent to requiring that each of the components $\gamma_1, \gamma_2, \ldots, \gamma_n$ of γ is smooth.
From now on, all parametrized curves will be assumed to be smooth.

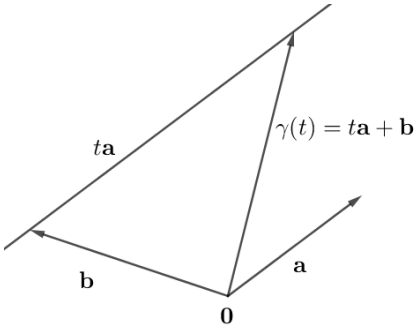
Definition 2.1.7. If γ is a parametrized curve, its first derivative $\dot{\gamma}(t)$ is called the tangent vator of γ at the point $\gamma(t)$.

Proposition 2.1.8. If the tangent vector of a parametrized curve is constant, the image of the curve is (part of) a straight line.

Proof. If $\dot{\gamma}(t) = \mathbf{a}$ for all t, where $\mathbf{a}$ is a constant vector, we have, integrating component wise,

$$\gamma(t) = \int \frac{d\gamma}{dt} dt = \int \mathbf{a} dt = t\mathbf{a} + \mathbf{b},$$

where $\mathbf{b}$ is another constant vector. If $\mathbf{a} \neq 0$, this is the parametric equation of the straight line parallel to $\mathbf{a}$ and passing through the point $\mathbf{b}$:



If $\mathbf{a} = 0$, the image of γ is a single point (namely, $\mathbf{b}$). □

Example 2.1.9. The limacon is the parametrized curve

$$\gamma(t) = ((1 + 2\cos t)\cos t, (1 + 2\cos t)\sin t), \quad t \in \mathbb{R}.$$

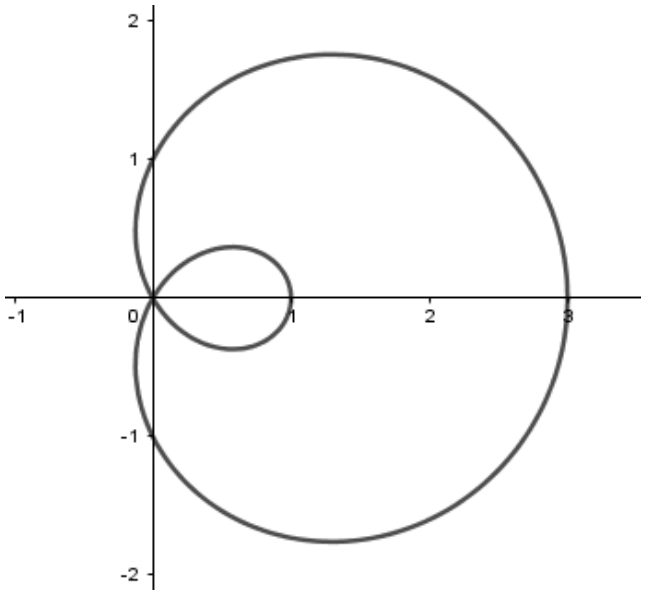


Figure 2.2: Limacon

Note that γ has a self-intersection at the origin in the sense that $\gamma(t) = \mathbf{0}$ for $t = \frac{2\pi}{3}$ and for $t = \frac{4\pi}{3}$. The tangent vector is

$$\dot{\gamma}(t) = (-\sin t - 2\sin 2t, \cos t + 2\cos 2t)$$

.

In particular,

$$\dot{\gamma}(2\pi/3) = (\sqrt{3}/2, -3/2), \quad \dot{\gamma}(4\pi/3) = (-\sqrt{3}/2, -3/2)$$

.

This example shows that we must be careful while talking about a 'point' of a parametrized curve γ: strictly speaking, this should be the same thing as a value of the curve parameter t, and not the corresponding geometric point $\gamma(t) \in \mathbb{R}^n$. Thus, Definition 2.1.7 should more properly read "If γ is a parametrized curve, its first derivative $\dot{\gamma}(t)$ is called the tangent vector of γ at the parameter value t.

Exercise 2.1.10. Find parametrizations of the following level curves:

(i) $y^2 - x^2 = 1$;

(ii) $\frac{x^2}{4} + \frac{y^2}{9} = 1$.

Solution:

(i) $\gamma(t) = (\tan t, \sec t)$ with $-\pi/2 < t < \pi/2$ and $\pi/2 < t < 3\pi/2$. Note that γ is defined on the union of two disjoint intervals: this corresponds to the fact that the hyperbola $y^2 - x^2 = 1$ is in two pieces, where $y \geq 1$ and where $y \leq -1$.

(ii) $\gamma(t) = (2\cos t, 3\sin t)$.

Exercise 2.1.11. Find the Cartesian equations of the following parametrized curves:

(i) $\gamma(t) = \big(\cos^2 t, \sin^2 t\big)$;

(ii) $\gamma(t) = (e^t, t^2)$.

Solution:

(i) $x + y = 1$.

(ii) $y = (\ln x)^2$.

2.2 Arc-length

Definition 2.2.1. The arc-length of a curve γ starting at the point $\gamma(t_0)$ is the function $s(t)$ given by

$$s(t) = \int_{t_0}^{t} \|\dot{\gamma}(u)\| du$$

Thus, $s(t_0) = 0$ and $s(t)$ is positive or negative according to whether t is larger or smaller than t_0. If we choose a different starting point $\gamma(\tilde{t}_0)$, the resulting arc-length $\tilde{s}$ differs from s by the constant $\displaystyle\int_{t_0}^{\tilde{t}_0} \|\dot{\gamma}(u)\|\ du$ because

$$\int_{t_0}^{t} \|\dot{\gamma}(u)\|\ du = \int_{\tilde{t}_0}^{t} \|\dot{\gamma}(u)\|\ du + \int_{t_0}^{\tilde{t}_0} \|\dot{\gamma}(u)\|\ du$$

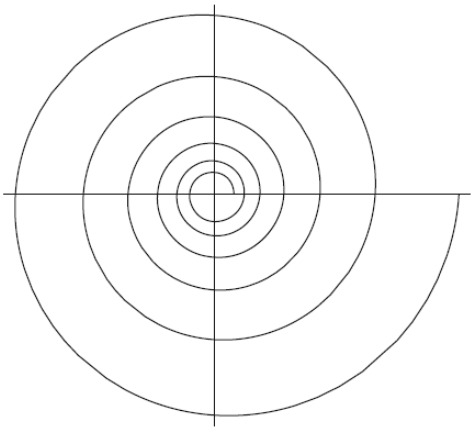


Figure 2.3: Logarithmic spiral

Example 2.2.2. For a logarithmic spiral

$$\gamma(t) = (e^{kt}\cos t, e^{kt}\sin t)$$

where k is a non-zero constant, we have

$$\dot{\gamma} = (e^{kt}(k\cos t - \sin t), (e^{kt}(k\sin t + \cos t))$$
$$\therefore \ \|\dot{\gamma}\|^2 = e^{2kt}(k\cos t - \sin t)^2 + e^{2kt}(k\sin t + \cos t)^2 = (k^2+1)e^{2kt}$$

Hence, the arc-length of γ starting at $\gamma(0) = (1, 0)$ is

$$s = \int_0^t \sqrt{k^2+1}\ e^{ku}\ du = \frac{\sqrt{k^2+1}}{k}(e^{kt} - 1)$$

The arc-length is a differentiable function. If s is the arc-length of a curve γ starting at $\gamma(t_0)$, we have

$$\frac{ds}{dt} = \frac{d}{dt}\int_{t_0}^t \|\dot{\gamma}(u)\|\ du = \|\dot{\gamma}(t)\| \tag{2.1}$$

Definition 2.2.3. If $\gamma : (\alpha, \beta) \to \mathbb{R}^n$ is a parametrized curve, its speed at the point $\gamma(t)$ is $\|\dot{\gamma}(t)\|$, and γ is said to be a unit-speed curve if $\dot{\gamma}(t)$ is a unit vector for all $t \in (\alpha, \beta)$.

We recall that the dot product (or scalar product) of vectors $\mathbf{a} = (a_1, \ldots, a_n)$ and $\mathbf{b} = (b_1, \ldots, b_n)$ in $\mathbb{R}^n$ is

$$\mathbf{a} \cdot \mathbf{b} = \sum_{i=1}^n a_i b_i$$

.

If $\mathbf{a}$ and $\mathbf{b}$ are smooth functions of a parameter t, we shall make use of the 'product formula'

$$\frac{d}{dt}(\mathbf{a}\cdot\mathbf{b}) = \frac{d\mathbf{a}}{dt}\cdot\mathbf{b} + \mathbf{a}\cdot\frac{d\mathbf{b}}{dt}$$

This follows easily from the definition of the dot product and the usual product formula for scalar functions,

$$\frac{d}{dt}(a_i b_i) = \frac{da_i}{dt}b_i + a_i\frac{db_i}{dt}$$

Proposition 2.2.4. Let $\mathbf{n}(t)$ be a unit vector that is a smooth function of a parameter t. Then, the dot product

$$\dot{\mathbf{n}}(t)\cdot\mathbf{n}(t) = 0$$

for all t, i.e., $\dot{\mathbf{n}}(t)$ is zero or perpendicular to $\mathbf{n}(t)$ for all t.

In particular, if γ is a unit-speed curve, then $\ddot{\gamma}$ is zero or perpendicular to $\dot{\gamma}$.

Proof. Using the product formula to differentiate both sides of the equation $\mathbf{n}\cdot\mathbf{n} = 1$ with respect to t gives

$$\dot{\mathbf{n}}\cdot\mathbf{n} + \mathbf{n}\cdot\dot{\mathbf{n}} = 0$$

so $2\dot{\mathbf{n}}\cdot\mathbf{n} = 0$. The last part follows by taking $\mathbf{n} = \dot{\gamma}$. □

Exercise 2.2.5. Calculate the arc-length of the catenary $\gamma(t) = (t, \cosh t)$ starting at the point $(0, 1)$.

Solution:

$\dot{\gamma}(t) = (1, \sinh t)$ so $\|\dot{\gamma}\| = \cosh t$ and the arc-length is $s = \displaystyle\int_0^t \cosh u\, du = \sinh t$

Exercise 2.2.6. Show that the following curves are unit-speed:

(i) $\gamma(t) = \left(\frac{1}{3}(1+t)^{3/2}, \frac{1}{3}(1-t)^{3/2}, \frac{t}{\sqrt{2}}\right)$;

(ii) $\gamma(t) = \left(\frac{4}{5}\cos t, 1 - \sin t, -\frac{3}{5}\cos t\right)$.

Solution:

(i) $\|\dot{\gamma}\|^2 = \frac{1}{4}(1+t) + \frac{1}{4}(1-t) + \frac{1}{2} = 1$.

(ii) $\|\dot{\gamma}\|^2 = \frac{16}{25}\sin^2 t + \cos^2 t + \frac{9}{25}\sin^2 t = \cos^2 t + \sin^2 t = 1$.

2.3 Reparametrization

Definition 2.3.1. A parametrized curve $\tilde{\gamma} : (\tilde{\alpha}, \tilde{\beta}) \to \mathbb{R}^n$ is a reparametrization of a parametrized curve $\gamma : (\alpha, \beta) \to \mathbb{R}^n$ if there is a smooth bijective map $\phi : (\tilde{\alpha}, \tilde{\beta}) \to (\alpha, \beta)$ (the reparametrization map) such that the inverse map $\phi^{-1} : (\alpha, \beta) \to (\tilde{\alpha}, \tilde{\beta})$ is also smooth and

$$\tilde{\gamma}(\tilde{t}) = \gamma(\phi(\tilde{t})) \quad \text{for all} \quad \tilde{t} \in (\tilde{\alpha}, \tilde{\beta}) \tag{2.2}$$

Note that, since ϕ has a smooth inverse, γ is a reparatmetrization of $\tilde{\gamma}$

$$\tilde{\gamma}\left(\phi^{-1}(t)\right) = \gamma\left(\phi\left(\phi^{-1}(t)\right)\right) = \gamma(t) \text{ for all } t \in (\alpha, \beta)$$

Two curves that are reparametrizations of each other have the same image, so they should have the same geometric properties.

Example 2.3.2. In Example 2.1.4, we found that the circle $x^2 + y^2 = 1$ has a parametrization $\gamma(t) = (\cos t, \sin t)$. Let

$$\tilde{\gamma}(t) = (\sin t, \cos t)$$

then $\tilde{\gamma}(t)$ is another parametrization of the circle $x^2 + y^2 = 1$
(since $\sin^2 t + \cos^2 t = 1$) . Let $\phi(t) = \pi/2 - t$. then

$$(\cos \phi(t), \sin \phi(t)) = (\sin t, \cos t)$$

So $\tilde{\gamma}$ is a reparametrizatian of γ.

Definition 2.3.3. A point $\gamma(t)$ of a parametrized curve γ is called a regular point if $\dot{\gamma}(t) \neq \mathbf{0}$; otherwise $\gamma(t)$ is a singular point of γ. A curve is regular if all of its points are regular.

Proposition 2.3.4. Any reparametrization of a regular curve is regular.

Proof. Suppose that γ and $\tilde{\gamma}$ are related as in Definition 2.3.1, let $t = \phi(\tilde{t})$ and $\psi = \phi^{-1}$ so that $\tilde{t} = \psi(t)$. Differentiating both sides of the equation $\phi(\psi(t)) = t$ with respect to t and using the chain rule gives

$$\frac{d\phi}{d\tilde{t}} \frac{d\psi}{dt} = 1.$$

This shows that $d\phi/d\tilde{t}$ is never zero. Since $\tilde{\gamma}(\tilde{t})$=$\gamma(\phi(\tilde{t}))$, another application of the chain rule gives

$$\frac{d\tilde{\gamma}}{d\tilde{t}} = \frac{d\gamma}{dt} \frac{d\phi}{d\tilde{t}}$$

which shows that $d\tilde{\gamma}/d\tilde{t}$ is never zero, if $d\gamma/dt$ is never zero. □

Proposition 2.3.5. If $\gamma(t)$ is a regular curve, its arc-length s (see Definition 2.2.1), starting at any point of γ, is a smooth function of t.

Proof. We have already seen that (whether or not γ is regular) s is a differentiable function of t and

$$\frac{ds}{dt} = \|\dot{\gamma}(t)\|$$

To simplify the notation, assume from now onwards that γ is a plane curve, say

$$\gamma(t) = (u(t), v(t))$$

where u and v are smooth functions of t, so that

$$\frac{ds}{dt} = \sqrt{\dot{u}^2 + \dot{v}^2}$$

we know that the function $f(x) = \sqrt{x}$ is a smooth function on the open interval $(0, \infty)$. It is easy to prove by induction on $n \geq 1$ that

$$\frac{d^n f}{dx^n} = (-1)^{n-1} \frac{1.3.5.\ldots.(2n-1)}{2^n} x^{-(2n+1)/2}$$

Since u and v are smooth functions of t so are $\dot{u}$ and $\dot{v}$ and hence is $\dot{u}^2 + \dot{v}^2$. Since γ is regular, $\dot{u}^2 + \dot{v}^2 \ > 0$ for all values of t, so the composite function

$$\frac{ds}{dt} = f\left(\dot{u}^2 + \dot{v}^2\right)$$

is a smooth function of t and hence s itself is smooth. □

Proposition 2.3.6. A parametrized curve has a unit-speed reparametrization if and only if it is regular.

Proof. Suppose first that a parametrized curve $\gamma : (\alpha, \beta) \to \mathbb{R}^n$ has a unit-speed reparametrization $\tilde{\gamma}$ with reparametrization map ϕ. Letting $t = \phi(\tilde{t})$, we have $\tilde{\gamma}(\tilde{t}) = \gamma(t)$ and so

$$\frac{d\tilde{\gamma}}{d\tilde{t}} = \frac{d\gamma}{dt}\frac{dt}{d\tilde{t}}$$

$$\therefore \left\|\frac{d\tilde{\gamma}}{d\tilde{t}}\right\| = \left\|\frac{d\gamma}{dt}\right\| \left|\frac{dt}{d\tilde{t}}\right|$$

Since $\tilde{\gamma}$ is unit-speed, $\left\|\frac{d\tilde{\gamma}}{d\tilde{t}}\right\| = 1$,so $\frac{d\gamma}{dt}$ cannot be zero.

Conversely, suppose that the tangent vector $d\gamma/dt$ is never zero. By Equation 2.1, $\dfrac{ds}{dt} > 0$ for all t, where s is the arc-length of γ starting at any point of the curve and by proposition 2.3.5 s is a smooth function of t. It follows from the inverse function theorem (Refer:Theorem 3.3.1) that $s : (\alpha, \beta) \to \mathbb{R}$ is injective, that its image is an open interval $(\tilde{\alpha}, \tilde{\beta})$, and that the inverse map $s^{-1} : (\tilde{\alpha}, \tilde{\beta}) \to (\alpha, \beta)$ is smooth. We take $\phi = s^{-1}$ and let $\tilde{\gamma}$ be the corresponding reparametrization of γ, so that $\tilde{\gamma}(s) = \gamma(t)$ (see Equation 2.2). Then,

$$\frac{d\tilde{\gamma}}{ds}\frac{ds}{dt} = \frac{d\gamma}{dt},$$

$$\therefore \left\|\frac{d\tilde{\gamma}}{ds}\right\| \frac{ds}{dt} = \left\|\frac{d\gamma}{dt}\right\| = \frac{ds}{dt} \quad (\text{ by Equation 2.1})$$

$$\therefore \left\|\frac{d\tilde{\gamma}}{ds}\right\| = 1$$

□

Corollary 2.3.7. Let γ be a regular curve and let $\tilde{\gamma}$ be a unit-speed reparametrization of γ :

$$\tilde{\gamma}(u(t)) = \gamma(t) \quad \text{for all } t,$$

where u is a smooth function of t. Then, if s is the arc-length of γ (starting at any point), we have

$$u = \pm s + c \tag{2.3}$$

where c is a constant. Conversely, if u is given by Equation 2.3 for some value of c and with either sign, then $\tilde{\gamma}$ is a unit-speed reparametrization of γ.

Proof. The calculation in the first part of the proof of Proposition 2.3.6 shows that u gives a unit-speed reparametrization of γ if and only if

$$\frac{du}{dt} = \pm\left\|\frac{d\gamma}{dt}\right\| = \pm\frac{ds}{dt} \quad (\text{ by Equation 2.1}),$$

which is equivalent to $u = \pm s + c$ for some constant c. □

Example 2.3.8. For the parametrization $\gamma(t) = (t, t^2)$ of the parabola $y = x^2, \dot{\gamma}(t) = (1, 2t)$ is obviously never zero, so γ is regular.
But $\tilde{\gamma}(t) = (t^3, t^6)$ is also a parametrization of the same parabola. This time, $\dot{\tilde{\gamma}} = (3t^2, 6t^5)$, and this is zero when $t = 0$, so $\tilde{\gamma}$ is not regular. That is a given level curve can have both regular and non-regular parametrization.

Exercise 2.3.9. Which of the following curves are regular?

(i) $\gamma(t) = \left(\cos^2 t, \sin^2 t\right)$ for $t \in \mathbb{R}$.

(ii) The same curve as in (i), but with $0 < t < \pi/2$.

(iii) $\gamma(t) = (t, \cosh t)$ for $t \in \mathbb{R}$.

Solution:

(i) $\dot{\gamma} = \sin 2t(-1, 1)$ vanishes when t is an integer multiple of $\pi/2$, so γ is not regular.

(ii) γ is regular since $\dot{\gamma} \neq 0$ for $0 < t < \pi/2$.

(iii) $\dot{\gamma} = (1, \sinh t)$ is obviously never zero, so γ is regular.

Exercise 2.3.10. Show that:

(i) If $\tilde{\gamma}$ is a reparametrization of a curve γ, then γ is a reparametrization of $\tilde{\gamma}$.

(ii) If $\tilde{\gamma}$ is a reparametrization of γ, and $\hat{\gamma}$ is a reparametrization of $\tilde{\gamma}$, then $\hat{\gamma}$ is a reparametrization of γ.

Solution:

(i) If $\tilde{\gamma}(t) = \gamma(\varphi(t))$, let ψ be the inverse of the reparametrization map φ. Then $\tilde{\gamma}(\psi(t)) = \gamma(\varphi(\psi(t))) = \gamma(t)$.

(ii) If $\tilde{\gamma}(t) = \gamma(\varphi(t))$ and $\hat{\gamma}(t) = \tilde{\gamma}(\psi(t))$, where φ and ψ are reparametrization maps, then $\hat{\gamma}(t) = \gamma((\varphi \circ \psi)(t))$ and $\varphi \circ \psi$ is a reparametrization map because it is smooth and $\frac{d}{dt}(\varphi(\psi(t)) = \dot{\varphi}(\psi(t))\dot{\psi}(t) \neq 0$ as $\dot{\varphi}$ and $\dot{\psi}$ are both $\neq 0$.

2.4 Closed curves

Definition 2.4.1. Let $\gamma : \mathbb{R} \to \mathbb{R}^n$ be a smooth curve and let $T \in \mathbb{R}$. We say that γ is T-periodic if

$$\gamma(t + T) = \gamma(t) \text{ for all } t \in \mathbb{R}$$

If γ is not constant and is T-periodic for some $T \neq 0$, then γ is said to be closed.

Thus, if γ is T-periodic, a point moving around γ returns to its starting point after time T, whatever the starting point is. Of course, every curve is 0-periodic

Remark
If γ is T-periodic, it is clear that γ is determined by its restriction to any interval of length $|T|$. Conversely, closed curves are often given to us as curves defined on a closed interval, say $\gamma : [a, b] \to \mathbb{R}^n$. If γ and all its derivatives take the same value at a and b there is a unique way to extend γ to a $(b - a)$ - periodic (smooth) curve $\gamma : \mathbb{R} \to \mathbb{R}^n$.

It is clear that if a curve γ is T-periodic then it is $(-T)$-periodic because

$$\gamma(t - T) = \gamma((t - T) + T) = \gamma(t).$$

It follows that if γ is T-periodic for some $T \neq 0$, then it is T -periodic for some $T > 0$

Definition 2.4.2. The period of a closed curve γ is the smallest positive number T (Exercise 2.4.11) such that γ is T-periodic.

Example 2.4.3. The ellipse $\gamma(t) = (p \cos t, q \sin t)$ is a closed curve with period 2π

Definition 2.4.4. If γ is a regular closed curve of period T , define the length of γ to be

$$\ell(\gamma) = \int_0^T \|\dot{\gamma}(t)\| dt$$

Proposition 2.4.5. If γ is a regular closed curve of period T, then its unit-speed reparametrization of γ is always closed.

Proof. By the proof of Proposition 2.3.6 , using the arc-length

$$s = \int_0^t \|\dot{\gamma}(u)\| \, du$$

of γ as the parameter gives a unit-speed reparametrization $\tilde{\gamma}$ of γ (so that $\tilde{\gamma}(s) = \gamma(t)$). Note that

$$s(t + T) = \int_0^{t+T} \|\dot{\gamma}(u)\| \, du = \int_0^T \|\dot{\gamma}(u)\| \, du + \int_T^{t+T} \|\dot{\gamma}(u)\| \, du = \ell(\gamma) + s(t),$$

since, putting $v = u - T$ and using $\gamma(u - T) = \gamma(u)$ and hence by differentiation $\dot{\gamma}(u - T) = \dot{\gamma}(u)$), we get

$$\int_T^{t+T} \|\dot{\gamma}(u)\| \, du = \int_0^t \|\dot{\gamma}(v)\| \, dv = s(t)$$

Hence,

$$\tilde{\gamma}(s(t)) = \tilde{\gamma}(s(t')) \iff \gamma(t) = \gamma(t') \iff t' - t = kT \iff s(t') - s(t) = k\ell(\gamma)$$

where k is an integer. This shows that $\tilde{\gamma}$ is a closed curve with period $\ell(\gamma)$. □

Definition 2.4.6. A curve γ is said to have a self-intersection at a point $\mathbf{p}$ of the curve if there exist parameter values $a \neq b$ such that

1. $\gamma(a) = \gamma(b) = \mathbf{p}$, and
2. if γ is closed with period T, then $a - b$ is not an integer multiple of T.

Example 2.4.7. The limacon in Example 2.1.9 is a closed curve with period 2π. It is clear from the picture that it has exactly one self-intersection, at the origin.

Exercise 2.4.8. Give an example to show that a reparametrization of a closed curve need not be closed.
Solution:
The curve $\tilde{\gamma}(t) = (\cos(t^3 + t), \sin(t^3 + t))$ is a reparametrization of the circle $\gamma(t) = (\cos t, \sin t)$ but it is not closed.

Exercise 2.4.9. Show that if a curve γ is T_1-periodic and T_2-periodic, then it is $(k_1T_1 + k_2T_2)$-periodic for any integers k_1, k_2.
Solution:
If γ is T-periodic then it is kT-periodic for all $k \neq 0$ (this can be proved by induction on k if $k > 0$, or on $-k$ if $k < 0$). If γ is T_1-periodic and T_2-periodic then it is k_1T_1- and k_2T_2-periodic for all non-zero integers k_1, k_2, so $\gamma(t + k_1T_1 + k_2T_2) = \gamma(t + k_1T_1)$ as γ is k_2T_2-periodic, which $= \gamma(t)$ as γ is k_1T_1-periodic.

Exercise 2.4.10. Let $\gamma : \mathbb{R} \to \mathbb{R}^n$ be a curve and suppose that T_0 is the smallest positive number such that γ is T_0-periodic. Prove that γ is T-periodic if and only if $T = kT_0$ for some integer k.
Solution: If γ is T-periodic write $T = kT_0 + T_1$ where k is an integer and $0 \leq T_1 < T_0$. By Exercise 2.4.9 γ is T_1-periodic; if $T_1 > 0$ this contradicts the definition of T_0.

Exercise 2.4.11. Suppose that a non-constant function $\gamma : \mathbb{R} \to \mathbb{R}$ is T-periodic for some $T \neq 0$. This exercise shows that there is a smallest positive T_0 such that γ is T_0-periodic. The proof uses a little real analysis. Suppose for a contradiction that there is no such T_0.

(i) Show that there is a sequence $T_1, T_2, T_3, \ldots$ such that $T_1 > T_2 > T_3 > \cdots > 0$ and that γ is T_r-periodic for all $r \geq 1$.

(ii) Show that the sequence $\{T_r\}$ in (i) can be chosen so that $T_r \to 0$ as $r \to \infty$

(iii) Show that the existence of a sequence $\{T_r\}$ as in (i) such that $T_r \to 0$ as $r \to \infty$ implies that γ is constant.

Solution:

(i) Choose $T_1 > 0$ such that γ is T_1-periodic; then T_1 is not the smallest positive number with this property, so there is a $T_2 > 0$ such that γ is T_2– periodic. Iterating this argument gives the desired sequence.

(ii) The sequence $\{T_r\}_{r\geq 1}$ in (i) is decreasing and bounded below, so must converge to some $T_\infty \geq 0$. Then γ is T_∞-periodic because (using continuity of γ) $\gamma(t + T_\infty) = \lim_{r\to\infty} \gamma(t + T_r) = \lim_{r\to\infty} \gamma(t) = \gamma(t)$. γ is $(T_r - T_\infty)$-periodic for all $r \geq 1$, and this sequence of positive numbers converges to 0

(iii) If $\{T_r\}$ is as in (i) and $T_r \to 0$ as $r \to \infty$, then by the mean value theorem $0 = (f(t + T_r) - f(t))/T_r = \dot{f}(t + \lambda T_r)$ for some $0 < \lambda < 1$. Letting $r \to \infty$ gives $\dot{f}(t) = 0$ for all t, so f is constant.

2.5 Level curves versus parametrized curves

Theorem 2.5.1. Let $f(x, y)$ be a smooth function of two variables (which means that all the partial derivatives of f, of all orders, exist and are continuous functions). Assume that, at every point of the level curve

$$\mathcal{C} = \left\{(x, y) \in \mathbb{R}^2 \mid f(x, y) = 0\right\},$$

$\frac{\partial f}{\partial x}$ and $\frac{\partial f}{\partial y}$ are not both zero. If $\mathbf{p}$ is a point of $\mathcal{C}$, with coordinates (x_0, y_0) then there is a regular parametrized curve $\gamma(t)$, defined on an open interval containing 0, such that γ passes through $\mathbf{p}$ when $t = 0$ and $\gamma(t)$ is contained in $\mathcal{C}$ for all t.

Proof. Suppose that $(x_0 + \Delta x, y_0 + \Delta y)$ is a point of $\mathcal{C}$ near $\mathbf{p}$, so that

$$f(x_0 + \Delta x, y_0 + \Delta y) = 0$$

By the two-variable form of Taylor's theorem,

$$f(x_0 + \Delta x, y_0 + \Delta y) = f(x_0, y_0) + \Delta x \frac{\partial f}{\partial x} + \Delta y \frac{\partial f}{\partial y}$$

neglecting products of the small quantities Δx and Δy (the partial derivatives are evaluated at (x_0, y_0)). Hence,

$$\Delta x \frac{\partial f}{\partial x} + \Delta y \frac{\partial f}{\partial y} = 0 \tag{2.4}$$

Since Δx and Δy are small, the vector $(\Delta x, \Delta y)$ is nearly tangent to $\mathcal{C}$ at $\mathbf{p}$. So equation 2.4 says that the vector $\mathbf{n} = \left(\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}\right)$ is perpendicular to $\mathcal{C}$ at $\mathbf{p}$. The hypothesis in the Theorem tells us that the vector $\mathbf{n}$ is non-zero at every point of $\mathcal{C}$. Suppose, for example, that $\frac{\partial f}{\partial x} \neq 0$ at $\mathbf{p}$. Then, $\mathbf{n}$ is not parallel to the x-axis at $\mathbf{p}$, so the tangent to $\mathcal{C}$ at $\mathbf{p}$ is not parallel to the y -axis. This implies that vertical lines $x =$ constant near $x = x_0$ all intersect $\mathcal{C}$ in a unique point (x, y) near $\mathbf{p}$. In other words, the equation

$$f(x, y) = 0 \tag{2.5}$$

has a unique solution y near y_0 for every x near x_0. That there is a function (Theorem 3.3.1) $g(x)$, defined for x near x_0, such that $y = g(x)$ is the unique solution of Equation 2.5 near y_0. We can now define a parametrization γ of the part of $\mathcal{C}$ near $\mathbf{p}$ by

$$\gamma(t) = (t, g(t))$$

By inverse function theorem g is smooth and γ is regular since $\dot{\gamma} = (1.\dot{g})$ is never zero. □

Theorem 2.5.2. Let γ be a regular parametrized plane curve, and let $\gamma(t_0) = (x_0, y_0)$ be a point in the image of γ. Then, there is a smooth real-valued function $f(x, y)$, defined for x and y in open intervals containing x_0 and y_0, such that $\gamma(t)$ is contained in the level curve $f(x, y) = 0$ for all values of t in some open interval containing t_o.

Proof. Let

$$\gamma(t) = (u(t), v(t))$$

where u and v are smooth functions , Since γ is regular, at least one of $\dot{u}(t_0)$ and $\dot{v}(t_0)$ is non-zero, say $\dot{u}(t_0)$. This means that the graph of u as a function of t is not parallel to the t-axis at t_0 : As in the proof of Theorem 2.5.1, this implies that any line parallel to the t-axis close to $u = x_0$ intersects the

graph of u at a unique point $u(t)$ with t close to t_0. This gives a function $h(x)$, defined for x in an open interval containing x_0, such that $t = h(x)$ is the unique solution of $u(t) = x$ if x is near x_0 and t is near t_0. By the inverse function theorem , h is smooth. So the function

$$f(x, y) = y - v(h(x))$$

has the properties we want. □

2.6 Curvature

Definition 2.6.1. If γ is a unit-speed curve with parameter t, its curvature $\kappa(t)$ at the point $\gamma(t)$ is defined to be $\|\ddot{\gamma}(t)\|$.

Example 2.6.2. If $\gamma(t) = t + \mathbf{b}$, a straight line, then $\ddot{\gamma} = \mathbf{0}$ everywhere .So curvature of a straight line is always zero.

Example 2.6.3. Consider the circle in $\mathbb{R}^2$ centred at (x_0, y_0) and of radius R. This has a unit-speed parametrization

$$\gamma(t) = \left(x_0 + R\cos\frac{t}{R}, y_0 + R\sin\frac{t}{R}\right).$$

We have $\dot{\gamma}(t) = \left(-\sin\frac{t}{R}, \cos\frac{t}{R}\right)$, and so

$$\|\dot{\gamma}(t)\| = \sqrt{\left(-\sin\frac{t}{R}\right)^2 + \left(\cos\frac{t}{R}\right)^2} = 1$$

confirming that γ is unit-speed, and hence $\ddot{\gamma}(t) = \left(-\frac{1}{R}\cos\frac{t}{R}, -\frac{1}{R}\sin\frac{t}{R}\right)$, so the curvature

$$\|\ddot{\gamma}(t)\| = \sqrt{\left(-\frac{1}{R}\cos\frac{t}{R}\right)^2 + \left(-\frac{1}{R}\sin\frac{t}{R}\right)^2} = \frac{1}{R}$$

is the reciprocal of the radius of the circle. That is small circles should have large curvature and large circles should have small curvature.

By Proposition2.3.6, Every regular curve γ has a unit-speed parametrization $\tilde{\gamma}$. If $\hat{\gamma}$ is another unit-speed pararnetrization of γ, then by Corollary 2.3.7

$$\tilde{\gamma}(t) = \hat{\gamma}(u)$$

where $u = \pm t + c$ and c is a constant. Then, by the chain rule,

$$\frac{d\tilde{\gamma}}{dt} = \pm\frac{d\hat{\gamma}}{du}$$

so

$$\frac{d^2\tilde{\gamma}^2}{dt} = \pm\frac{d}{du}\left(\pm\frac{d\hat{\gamma}}{du}\right) = \frac{d^2\hat{\gamma}}{du^2},$$

which shows that $\tilde{\gamma}$ and $\hat{\gamma}$ have the same curvature.

Proposition 2.6.4. Let $\gamma(t)$ be a regular curve in $\mathbb{R}^3$. Then, its curvature is

$$\kappa = \frac{\| \ddot{\gamma} \times \dot{\gamma} \|}{\| \dot{\gamma} \|^3} \tag{2.6}$$

where the $\times$ indicates the vector (cross) product and the dot denotes $\frac{d}{dt}$.

Proof. Let s be a unit-speed parameter for γ. Then, by the chain rule,

$$\dot{\gamma} = \frac{d\gamma}{dt} = \frac{d\gamma}{ds}\frac{ds}{dt},$$

so

$$\kappa = \left\|\frac{d^2\gamma}{ds^2}\right\| = \left\|\frac{d}{ds}\left(\frac{d\gamma/dt}{ds/dt}\right)\right\| = \left\|\frac{\frac{d}{dt}\left(\frac{d\gamma/dt}{ds/dt}\right)}{ds/dt}\right\| = \left\|\frac{\frac{ds}{dt}\frac{d^2\gamma}{dt^2} - \frac{d^2s}{dt^2}\frac{d\gamma}{dt}}{(ds/dt)^3}\right\| \tag{2.7}$$

Now,

$$\left(\frac{ds}{dt}\right)^2 = \|\dot{\gamma}\|^2 = \dot{\gamma}\cdot\dot{\gamma},$$

and differentiating with respect to t gives

$$\frac{ds}{dt}\frac{d^2s}{dt^2} = \dot{\gamma}\cdot\ddot{\gamma}.$$

Using this and Equation 2.7, we get

$$\kappa = \left\|\frac{(\frac{ds}{dt})^2\ddot{\gamma} - \frac{d^2s}{dt^2}\frac{ds}{dt}\dot{\gamma}}{(ds/dt)^4}\right\| = \frac{\|(\dot{\gamma}\cdot\dot{\gamma})\ddot{\gamma} - (\dot{\gamma}\cdot\ddot{\gamma})\dot{\gamma}\|}{\|\dot{\gamma}\|^4}$$

Using the vector triple product identity

$$\mathbf{a}\times(\mathbf{b}\times\mathbf{c}) = (\mathbf{a}\cdot\mathbf{c})\mathbf{b} - (\mathbf{a}\cdot\mathbf{b})\mathbf{c}$$

(where $\mathbf{a},\mathbf{b},\mathbf{c}\in\mathbb{R}^3$), we get

$$\dot{\gamma}\times(\ddot{\gamma}\times\dot{\gamma}) = (\dot{\gamma}\cdot\dot{\gamma})\ddot{\gamma} - (\dot{\gamma}\cdot\ddot{\gamma})\dot{\gamma}$$

Further, $\dot{\gamma}$ and $\ddot{\gamma}\times\dot{\gamma}$ are perpendicular vectors, so

$$\|\dot{\gamma}\times(\ddot{\gamma}\times\dot{\gamma})\| = \|\dot{\gamma}\|\|\ddot{\gamma}\times\dot{\gamma}\|$$

Hence,

$$\frac{\|(\dot\gamma\cdot\dot\gamma)\ddot\gamma-(\dot\gamma\cdot\ddot\gamma)\dot\gamma\|}{\|\dot\gamma\|^4}=\frac{\|\dot\gamma\times(\ddot\gamma\times\dot\gamma)\|}{\|\dot\gamma\|^4}=\frac{\|\dot\gamma\|\|\ddot\gamma\times\dot\gamma\|}{\|\dot\gamma\|^4}=\frac{\|\ddot\gamma\times\dot\gamma\|}{\|\dot\gamma\|^3}.$$

□

Example 2.6.5. A circular helix with axis the z-axis is a curve of the form

$$\gamma(\theta)=(a\cos\theta,a\sin\theta,b\theta),\quad \theta\in\mathbb{R}$$

where a and b are constants.

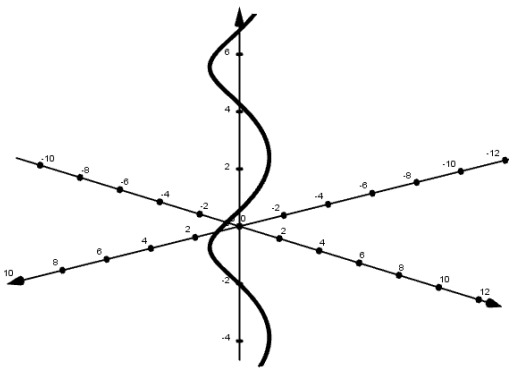


Figure 2.4: Circular helix

If (x,y,z) is a point on the helix, so that

$$x=a\cos\theta, y=a\sin\theta, z=b\theta$$

,
for some value of θ, then $x^2+y^2=a^2$, showing that the helix lies on the cylinder with axis the z-axis and radius $|a|$; the positive number $|a|$ is called the radius of the helix. As θ increases by 2π, the point $(a\cos\theta,a\sin\theta,b\theta)$ rotates once round the z-axis and moves parallel to the z-axis by $2\pi b$; the positive number $2\pi|b|$ is called the pitch of the helix.

Let us compute the curvature of the helix using the formula 2.6 . Denoting $d/d\theta$ by a dot, we have $\dot\gamma(\theta)=(-a\sin\theta,a\cos\theta,b)$ so

$$\|\dot\gamma(\theta)\|=\sqrt{a^2+b^2}$$

This shows that $\dot\gamma(\theta)$ is never zero, so γ is regular (unless $a=b=0$, in which case the image of the helix is a single point).Hence, by the formula 2.6,

we have $\ddot{\gamma} = (-a\cos\theta, -a\sin\theta, 0)$ so $\ddot{\gamma} \times \dot{\gamma} = (-ab\sin\theta, ab\cos\theta, -a^2)$ and hence

$$\kappa = \frac{\|(-ab\sin\theta, ab\cos\theta, -a^2)\|}{\|(-a\sin\theta, a\cos\theta, b)\|^3} = \frac{(a^2b^2 + a^4)^{1/2}}{(a^2+b^2)^{3/2}} = \frac{|a|}{a^2+b^2}. \tag{2.8}$$

Thus, the curvature of the helix is constant.

Let us examine some limiting cases to see if this result agrees with what we already know. First, suppose that $b = 0$ (but $a \neq 0$). Then, the helix is simply a circle in the xy-plane of radius $|a|$, so by example 2.6.3 its curvature is $\dfrac{1}{|a|}$. On the other hand, the formula 2.8 gives the curvature as

$$\frac{|a|}{a^2+0^2} = \frac{|a|}{a^2} = \frac{|a|}{|a|^2} = \frac{1}{|a|}$$

Next, suppose that $a = 0$ (but $b \neq 0$). Then, the image of the helix is just the z-axis, a straight line, so by example 2.6.2 the curvature is zero. And formula 2.8 gives zero when $a = 0$ too.

Exercise 2.6.6. Compute the curvature of the following curves:

(i) $\gamma(t) = \left(\frac{1}{3}(1+t)^{3/2}, \frac{1}{3}(1-t)^{3/2}, \frac{t}{\sqrt{2}}\right)$.

(ii) $\gamma(t) = \left(\frac{4}{5}\cos t, 1-\sin t, -\frac{3}{5}\cos t\right)$.

(iii) $\gamma(t) = (t, \cosh t)$.

(iv) $\gamma(t) = \left(\cos^3 t, \sin^3 t\right)$.

Solution:

(i) γ is unit-speed (Exercise 2.2.6 (i)) so $\kappa = \|\ddot{\gamma}\| = \| \left(\frac{1}{4}(1+t)^{-1/2}, \frac{1}{4}(1-t)^{-1/2}, 0\right) \| = \frac{1}{\sqrt{8(1-t^2)}}$.

(ii) γ is unit-speed (Exercise 2.2.6 (ii)) so $\kappa = \|\ddot{\gamma}\| = \left\|\left(-\frac{4}{5}\cos t, \sin t, \frac{3}{5}\cos t\right)\right\| = 1$.

(iii) $\kappa = \frac{\|(1,\sinh t,0)\times(0,\cosh t,0)\|}{\|(1,\sinh t,0)\|^3} = \frac{\cosh t}{\cosh^3 t} = \operatorname{sech}^2 t$ using Proposition 2.6.4.

(iv) $\left(-3\cos^2 t\sin t, 3\sin^2 t\cos t, 0\right)\times\left(-3\cos^3 t + 6\cos t\sin^2 t, 6\sin t\cos^2 t - 3\sin^3 t, 0\right) = \left(0, 0, -9\sin^2 t\cos^2 t\right)$, so $\kappa = \dfrac{\left\|\left(0,0,-9\sin^2 t\cos^2 t\right)\right\|}{\left\|\left(-3\cos^2 t\sin t, 3\sin^2 t\cos t, 0\right)\right\|^3} = \frac{1}{3|\sin t\cos t|}$. This becomes infinite when t is an integer multiple of $\pi/2$, i.e., at the four cusps $(\pm 1, 0)$ and $(0, \pm 1)$ of the astroid.

2.7 Plane curves

Definition 2.7.1. Suppose that $\gamma(s)$ is a unit-speed curve in $\mathbb{R}^2$. Denoting d/ds by a dot, let

$$\mathbf{t} = \dot{\gamma}$$

be the tangent vector of γ; note that t is a unit vector. Let $\mathbf{n}_s$ be the unit vector obtained by rotating $\mathbf{t}$ anticlockwise by $\frac{\pi}{2}$ called, the signed unit normal of γ.

Definition 2.7.2. By Proposition 2.2.4, $\dot{\mathbf{t}} = \ddot{\gamma}$ is perpendicular to $\mathbf{t}$, and hence parallel to $\mathbf{n}_s$. Thus, there is a scalar κ_s such that

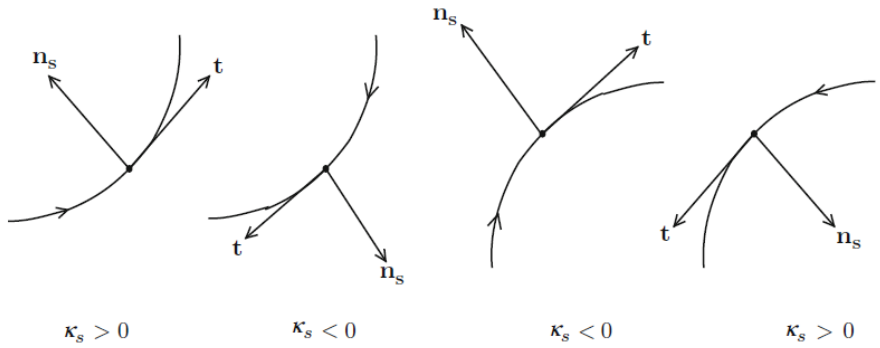
$$\ddot{\gamma} = \kappa_s \mathbf{n}_s;$$

κ_s is called the signed curvature of γ (it can be positive, negative or zero). Note that, since $\|\mathbf{n}_s\| = 1$, we have

$$\kappa = \|\ddot{\gamma}\| = \|\kappa_s \mathbf{n}_s\| = |\kappa_s|\,, \tag{2.9}$$

so the curvature of γ is the absolute value of its signed curvature.

The following diagrams show how the sign of the signed curvature is determined (in each case, the arrow on the curve indicates the direction of increasing s)



If $\gamma(t)$ is a regular, but not necessarily unit-speed, we define the unit tangent vector $\mathbf{t}$, signed unit normal $\mathbf{n}_s$ and signed curvature κ_s of γ to be those of its unit-speed parametrization $\tilde{\gamma}(s)$, where s is the arc-length of γ. Thus,

$$\mathbf{t} = \frac{d\gamma/dt}{ds/dt} = \frac{d\gamma/dt}{\|d\gamma/dt\|}$$

$\mathbf{n}_s$ is obtained by rotating $\mathbf{t}$ anticlockwise by $\pi/2$, and

$$\frac{d\mathbf{t}}{dt} = \frac{d\mathbf{t}}{ds}\frac{ds}{dt} = \kappa_s \frac{ds}{dt}\mathbf{n}_s = \kappa_s \left\| \frac{d\gamma}{dt} \right\| \mathbf{n}_s$$

Proposition 2.7.3. Let $\gamma : (\alpha, \beta) \to \mathbb{R}^2$ be a unit-speed curve, let $s_0 \in (\alpha, \beta)$ and let φ_0 be such that

$$\dot{\gamma}(s_0) = (\cos\varphi_0, \sin\varphi_0)$$

Then there is a unique smooth function $\varphi : (\alpha, \beta) \to \mathbb{R}$ such that $\varphi(s_0) = \varphi_0$ and

$$\dot{\gamma}(s) = (\cos\varphi(s), \sin\varphi(s)) \tag{2.10}$$

for all $s \in (\alpha, \beta)$.

Proof. Let

$$\dot{\gamma}(s) = (f(s), g(s))$$

Since γ is unit speed,

$$f(s)^2 + g(s)^2 = 1 \quad \text{for all } s \tag{2.11}$$

Define

$$\varphi(s) = \varphi_0 + \int_{s_0}^{s} (f\dot{g} - g\dot{f})\, dt.$$

Obviously $\varphi(s_0) = \varphi_0$. Moreover, since the functions f and g are smooth, so is $\dot{\varphi} = f\dot{g} - g\dot{f}$, and hence so is φ.
Let

$$F = f\cos\varphi + g\sin\varphi, \quad G = f\sin\varphi - g\cos\varphi.$$

Then,

$$\dot{F} = (\dot{f} + g\dot{\varphi})\cos\varphi + (\dot{g} - f\dot{\varphi})\sin\varphi.$$

But

$$\dot{f} + g\dot{\varphi} = \dot{f}\left(1 - g^2\right) + fg\dot{g} = f(f\dot{f} + g\dot{g}) = 0$$

where the second equality used Equation 2.11 and the last equality used its consequence

$$f\dot{f} + g\dot{g} = 0.$$

Similarly, $\dot{g} - f\dot{\varphi} = 0$. Hence, $\dot{F} = 0$ and F is constant. A similar argument shows that G is constant. But

$$F(s_0) = f(s_0)\cos\varphi_0 + g(s_0)\sin\varphi_0 = \cos^2\varphi_0 + \sin^2\varphi_0 = 1,$$

and similarly $G(s_0) = 0$. It follows that

$$f\cos\varphi + g\sin\varphi = 1, \quad f\sin\varphi - g\cos\varphi = 0$$

for all s. These equations imply that $f = \cos\varphi, g = \sin\varphi$, and hence that the smooth function φ satisfies Equation 2.10.

As to the uniqueness, if ψ is another smooth function such that $\psi(s_0) = \varphi_0$ and $\dot{\gamma}(s) = (\cos\psi(s), \sin\psi(s))$ for $s \in (\alpha, \beta)$, there is an integer $n(s)$ such that

$$\psi(s) - \varphi(s) = 2\pi n(s) \quad \text{for all } s \in (\alpha, \beta)$$

Because φ and ψ are smooth, n is a smooth, hence continuous, function of s. This implies that n is a constant: otherwise we would have $n(s_0) \neq n(s_1)$ for some $s_1 \in (\alpha, \beta)$, and then by the intermediate value theorem the continuous function $n(s)$ would have to take all values between $n(s_0)$ and $n(s_1)$ when s is between s_0 and s_1. But most real numbers between $n(s_0)$ and $n(s_1)$ are not integers. Thus, n is actually independent of s, and since $\psi(s_0) = \varphi(s_0) = \varphi_0$, we must have $n = 0$ and hence $\psi(s_0) = \varphi(s_0)$ for all $s \in (\alpha, \beta)$. □

Definition 2.7.4. The smooth function φ in Proposition 2.7.3 is called the *turning angle* of γ determined by the condition $\varphi(s_0) = \varphi_0$.

Proposition 2.7.5. Let $\gamma(s)$ be a unit-speed plane curve, and let $\varphi(s)$ be a turning angle for γ. Then,

$$\kappa_{\bar{s}} = \frac{d\varphi}{ds}.$$

Thus, *the signed curvature is the rate at which the tangent vector of the curve rotates*

Proof. By Equation 2.10 the tangent vector $\mathbf{t} = (\cos\varphi, \sin\varphi)$, so

$$\dot{\mathbf{t}} = \dot{\varphi}(-\sin\varphi, \cos\varphi)$$

Since $\mathbf{n}_s = (-\sin\varphi, \cos\varphi)$, the equation $\dot{t} = \kappa_s \mathbf{n}_s$ gives the stated result. □

Example 2.7.6. Let us find the signed curvature of the curve $\gamma(t) = (t, \cosh t)$ (catenary) . We have $\dot{\gamma} = (1, \sinh t)$ and hence

$$s = \int_0^t \sqrt{1 + \sinh^2 t}\, dt = \sinh t,$$

so if φ is the angle between $\dot{\gamma}$ and the x-axis,

$$\begin{aligned} &\tan\varphi = \sinh t = s, \\ \therefore \quad &\sec^2\varphi \frac{d\varphi}{ds} = 1, \\ \therefore \quad &\kappa_s = \frac{d\varphi}{ds} = \frac{1}{\sec^2\varphi} = \frac{1}{1 + \tan^2\varphi} = \frac{1}{1 + s^2}. \end{aligned}$$

Definition 2.7.7. The *total signed curvature* of a unit-speed closed curve γ of length ℓ is

$$\int_0^l \kappa_s(s)\ ds.$$

Corollary 2.7.8. The total signed curvature of a closed plane curve is an integer multiple of 2π.

Proof. Let γ be a unit-speed closed plane curve and let ℓ be its length. By Proposition 2.7.5, the total signed curvature of γ is

$$\int_0^\ell \frac{d\varphi}{ds} ds = \varphi(\ell) - \varphi(0),$$

where φ is a turning angle for γ. Now, γ is l-periodic .Therefore

$$\gamma(s + \ell) = \gamma(s)$$

Differentiating both sides gives

$$\dot{\gamma}(s + \ell) = \dot{\gamma}(s)$$

and in particular $\dot{\gamma}(\ell) = \dot{\gamma}(0)$. Hence, by Equation. 2.10,

$$(\cos\varphi(\ell), \sin\varphi(\ell)) = (\cos\varphi(0), \sin\varphi(0)),$$

which implies that $\varphi(\ell) - \varphi(0)$ is an integer multiple of 2π. □

Definition 2.7.9. Let $n \geq 1$. An isometry of $\mathbb{R}^n$ is a map $F : \mathbb{R}^n \to \mathbb{R}^n$ that preserves the distance between any two points of $\mathbb{R}^n$:

$$\|F(\mathbf{v}) - F(\mathbf{w})\| = \|\mathbf{v} - \mathbf{w}\| \quad \text{for all } \mathbf{v}, \mathbf{w} \in \mathbb{R}^n$$

It is obvious that any composite of isometries is an isometry.

Result 2.7.10. Let P be an $n \times n$ orthogonal matrix and let $\mathbf{a} \in \mathbb{R}^n$. Then, the map $F : \mathbb{R}^n \to \mathbb{R}^n$ defined by

$$F(\mathbf{v}) = P\mathbf{v} + \mathbf{a} \tag{2.12}$$

is an isometry of $\mathbb{R}^n$.The isometries in 2.12 for which $\det(P) = 1$ are said to be direct isometries those for which $\det(P) = -1$ are said to opposite isometries. In particular , if

$$P = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$$

the corresponding isometry $\rho_\theta(\mathbf{v}) = P_\mathbf{v}$ is the anticlockwise rotation of $\mathbf{v}$ through an angle θ about the origin.

Theorem 2.7.11. Let $k : (\alpha, \beta) \to \mathbb{R}$ be any smooth function. Then, there is a unit-speed curve $\gamma : (\alpha, \beta) \to \mathbb{R}^2$ whose signed curvature is k.

Further, if $\tilde{\gamma} : (\alpha, \beta) \to \mathbb{R}^2$ is any other unit-speed curve whose signed curvature is k, there is a direct isometry M of $\mathbb{R}^2$ such that

$$\tilde{\gamma}(s) = M(\gamma(s)) \quad \text{for all } s \in (\alpha, \beta).$$

Proof. For the first part, fix $s_0 \in (\alpha, \beta)$ and define, for any $s \in (\alpha, \beta)$,

$$\begin{aligned} \varphi(s) &= \int_{s_0}^{s} k(u)du, \\ \gamma(s) &= \left(\int_{s_0}^{s} \cos\varphi(t)dt, \int_{s_0}^{s} \sin\varphi(t)dt \right) \end{aligned}$$

Then, the tangent vector of γ is

$$\dot{\gamma}(s) = (\cos\varphi(s), \sin\varphi(s)),$$

which is a unit vector making an angle $\varphi(s)$ with the x-axis. Thus, γ is unit-speed and, by Proposition 2.7.5, its signed curvature is

$$\frac{d\varphi}{ds} = \frac{d}{ds}\int_{s_0}^{s} k(u)du = k(s).$$

For the second part, let $\tilde{\varphi}(s)$ be a smooth turning angle for $\tilde{\gamma}$. Thus,

$$\begin{aligned} &\dot{\tilde{\gamma}}(s) = (\cos\tilde{\varphi}(s), \sin\tilde{\varphi}(s)), \\ \therefore \quad &\tilde{\gamma}(s) = \left(\int_{s_0}^{s} \cos\tilde{\varphi}(t)dt, \int_{s_0}^{s} \sin\tilde{\varphi}(t)dt \right) + \tilde{\gamma}(s_0) \end{aligned} \tag{2.13}$$

By Proposition 2.7.5, $k(s) = \dfrac{d\tilde{\varphi}}{ds}$ so

$$\tilde{\varphi}(s) = \int_{s_0}^{s} k(u)du + \tilde{\varphi}(s_0).$$

Inserting this into Equation 2.13, and writing $\mathbf{a}$ for the constant vector $\tilde{\gamma}(s_0)$

and θ for the constant scalar $\tilde{\varphi}(s_0)$, we get

$$\begin{aligned}
\tilde{\gamma}(s) &= T_{\mathbf{a}}\left(\int_{s_0}^{s}\cos(\varphi(t)+\theta)dt, \int_{s_0}^{s}\sin(\varphi(t)+\theta)dt\right)\\
&= T_{\mathbf{a}}\left(\cos\theta\int_{s_0}^{s}\cos\varphi(t)dt - \sin\theta\int_{s_0}^{s}\sin\varphi(t)dt\ ,\right.\\
&\qquad\left.\sin\theta\int_{s_0}^{s}\cos\varphi(t)dt + \cos\theta\int_{s_0}^{s}\sin\varphi(t)dt\right)\\
&= T_{\mathbf{a}}\rho_\theta\left(\int_{s_0}^{s}\cos\varphi(t)dt, \int_{s_0}^{s}\sin\varphi(t)dt\right)\\
&= T_{\mathbf{a}}\rho_\theta(\gamma(s))
\end{aligned}$$

□

Example 2.7.12.

Any regular plane curve γ whose curvature is a positive constant is part of a circle. To see this, let κ be the curvature of γ, and let κ_s be its signed curvature. Then, by Equation 2.9,

$$\kappa_s = \pm\kappa.$$

A priori, we could have $\kappa_s = \kappa$ at some points of the curve and $\kappa_s = -\kappa$ at others, but in fact this cannot happen since κ_s is a continuous function of s, so the intermediate value theorem tells us that, if κ_s takes both the value κ and the value $-\kappa$, it must take all values between. Thus, either $\kappa_s = \kappa$ at all points of the curve, or $\kappa_s = -\kappa$ at all points of the curve.In particular, κ_s is constant.

The idea now is to show that, whatever the value of κ_s, we can find a parametrized circle whose signed curvature is κ_s. The theorem 2.7.11 then tells us that every curve whose signed curvature is κ_s can be obtained by applying a direct isometry to this circle. Since rotations and translations obviously take circles to circles, it follows that every curve whose signed curvature is constant is (part of) a circle.

A unit-speed parametrization of the circle with centre the origin and radius R is

$$\gamma(s) = \left(R\cos\frac{s}{R}, R\sin\frac{s}{R}\right).$$

Its tangent vector

$$\mathbf{t} = \dot{\gamma} = \left(-\sin\frac{s}{R}, \cos\frac{s}{R}\right)$$

is the unit vector making an angle $\frac{\pi}{2} + \frac{s}{R}$ with the positive x-axis:

Hence, the signed curvature of γ is

$$\frac{d}{ds}\left(\frac{\pi}{2}+\frac{s}{R}\right)=\frac{1}{R}$$

Thus, if $\kappa_s > 0$, the circle of radius $1/\kappa_s$ has signed curvature κ_s.
If $\kappa_s < 0$, it is easy to check that the curve

$$\tilde{\gamma}(s)=\left(R\cos\frac{s}{R}, -R\sin\frac{s}{R}\right)$$

(which is just another parametrization of the circle with centre the origin and radius R) has signed curvature $-1/R$. Thus, if $R=-1/\kappa_s$ we again get a circle with signed curvature κ_s.

Example 2.7.13. Theorem 2.7.11 shows that we can find a plane curve with any given smooth function as its signed curvature. But simple curvatures can lead to complicated curves. For example, let the signed curvature be $\kappa_s(s)=s$. Following the proof of Theorem 2.7.11, and taking $s_0=0$, we get $\varphi(s)=\int_0^s u du=\frac{s^2}{2}$ so

$$\gamma(s)=\left(\int_0^s \cos\left(\frac{t^2}{2}\right)dt, \int_0^s \sin\left(\frac{t^2}{2}\right)dt\right).$$

These integrals cannot be evaluated in terms of 'elementary' functions.

Exercise 2.7.14. Show that, if γ is a unit-speed plane curve,

$$\dot{\mathbf{n}}_s=-\kappa_s\mathbf{t}$$

Solution:
Differentiate $\mathbf{t}\cdot\mathbf{n}_s=0$ and use $\dot{\mathbf{t}}=\kappa_s\mathbf{n}_s$.

Exercise 2.7.15. Show that the signed curvature of any regular plane curve $\gamma(t)$ is a smooth function of t.
Solution:
If γ is smooth, $\mathbf{t}=\dot{\gamma}$ is smooth and hence so are $\dot{\mathbf{t}}$ and $\mathbf{n}_s$ (since $\mathbf{n}_s$ is obtained by applying a rotation to $\mathbf{t}$). So $\kappa_s=\dot{\mathbf{t}}\cdot\mathbf{n}_s$ is smooth.

2.8 Space curves

Definition 2.8.1. Let $\gamma(s)$ be a unit-speed curve in $\mathbb{R}^3$, and let $\mathbf{t} = \dot{\gamma}$ be its unit tangent vector. If the curvature $\kappa(s)$ is non-zero, we define the *principal normal* of γ at the point $\gamma(s)$ to be the vector

$$\mathbf{n}(s) = \frac{1}{\kappa(s)}\dot{\mathbf{t}}(s) \tag{2.14}$$

Since $\|\dot{\mathbf{t}}\| = \kappa, \mathbf{n}$ is a unit vector. Further, by Proposition 2.2.4, $\mathbf{t} \cdot \dot{\mathbf{t}} = 0$, so $\mathbf{t}$ and $\mathbf{n}$ are actually perpendicular unit vectors. It follows that

$$\mathbf{b} = \mathbf{t} \times \mathbf{n} \tag{2.15}$$

is a unit vector perpendicular to both $\mathbf{t}$ and $\mathbf{n}$. The vector $\mathbf{b}(s)$ is called the *binormal* vector of γ at the point $\gamma(s)$.

Thus, $\{\mathbf{t}, \mathbf{n}, \mathbf{b}\}$ is an orthonormal basis of $\mathbb{R}^3$, and is *right-handed*, i.e.,

$$\mathbf{b} = \mathbf{t} \times \mathbf{n}, \quad \mathbf{n} = \mathbf{b} \times \mathbf{t}, \quad \mathbf{t} = \mathbf{n} \times \mathbf{b}$$

Since $\mathbf{b}(s)$ is a unit vector for all $s, \dot{\mathbf{b}}$ is perpendicular to $\mathbf{b}$. Now we use the 'product rule' for differentiating the vector product of vector-valued functions $\mathbf{u}$ and $\mathbf{v}$ of a parameter s :

$$\frac{d}{ds}(\mathbf{u} \times \mathbf{v}) = \frac{d\mathbf{u}}{ds} \times \mathbf{v} + \mathbf{u} \times \frac{d\mathbf{v}}{ds}$$

Applying this to $\mathbf{b} = \mathbf{t} \times \mathbf{n}$ gives

$$\dot{\mathbf{b}} = \dot{\mathbf{t}} \times \mathbf{n} + \mathbf{t} \times \dot{\mathbf{n}} = \mathbf{t} \times \dot{\mathbf{n}} \tag{2.16}$$

since by the definition 2.14 of $\mathbf{n}, \dot{\mathbf{t}} \times \mathbf{n} = \kappa\mathbf{n} \times \mathbf{n} = \mathbf{0}$. Equation 2.16 shows that $\dot{\mathbf{b}}$ is perpendicular to $\mathbf{t}$. Being perpendicular to both $\mathbf{t}$ and $\mathbf{b}, \dot{\mathbf{b}}$ must be parallel to $\mathbf{n}$, so

$$\dot{\mathbf{b}} = -\tau\mathbf{n} \tag{2.17}$$

for some scalar τ, which is called the torsion of γ . Note that the torsion is only defined if the curvature is non-zero.

We define the torsion of an arbitrary regular curve γ to be that of a unit-speed reparametrization of γ. As in the case of the curvature,torsion is same for any unit-speed reparametrization of γ.Thus the curvature and torsion are well defined for any regular curve.

Proposition 2.8.2. Let $\gamma(t)$ be a regular curve in $\mathbb{R}^3$ with nowhere-vanishing curvature. Then, denoting d/dt by a dot, its torsion is given by

$$\tau = \frac{(\dot{\gamma} \times \ddot{\gamma}) \cdot \dddot{\gamma}}{||\dot{\gamma} \times \ddot{\gamma}||^2}. \tag{2.18}$$

Proof. □

We first consider the case in which γ is unit-speed. Using Eqs. 2.15 and 2.17,

$$\tau = -\mathbf{n} \cdot \dot{\mathbf{b}} = -\mathbf{n} \cdot (\mathbf{t} \times \mathbf{n}) = -\mathbf{n} \cdot (\dot{\mathbf{t}} \times \mathbf{n} + \mathbf{t} \times \dot{\mathbf{n}}) = -\mathbf{n} \cdot (\mathbf{t} \times \dot{\mathbf{n}})$$

Now, $\mathbf{n} = \frac{1}{\kappa}\dot{\mathbf{t}} = \frac{1}{\kappa}\ddot{\gamma}$, so

$$\tau = -\frac{1}{\kappa}\ddot{\gamma} \cdot \left(\dot{\gamma} \times \frac{d}{dt}\left(\frac{1}{\kappa}\ddot{\gamma}\right)\right) = -\frac{1}{\kappa}\ddot{\gamma} \cdot \left(\dot{\gamma} \times \left(\frac{1}{\kappa}\dddot{\gamma} - \frac{\dot{\kappa}}{\kappa^2}\ddot{\gamma}\right)\right) = \frac{1}{\kappa^2}\dddot{\gamma} \cdot (\dot{\gamma} \times \ddot{\gamma}),$$

since $\ddot{\gamma} \cdot (\dot{\gamma} \times \ddot{\gamma}) = 0$ and $\ddot{\gamma} \cdot (\dot{\gamma} \times \dddot{\gamma}) = -\dddot{\gamma} \cdot (\dot{\gamma} \times \ddot{\gamma})$. This agrees with Equation 2.18, for, since γ is unit-speed, $\dot{\gamma}$ and $\ddot{\gamma}$ are perpendicular, so

$$\|\dot{\gamma} \times \ddot{\gamma}\| = \|\dot{\gamma}\|\|\ddot{\gamma}\| = \|\ddot{\gamma}\| = \kappa$$

In the general case, let s be arc-length along γ. Then,

$$\frac{d\gamma}{dt} = \frac{ds}{dt}\frac{d\gamma}{ds}, \quad \frac{d^2\gamma}{dt^2} = \left(\frac{ds}{dt}\right)^2 \frac{d^2\gamma}{ds^2} + \frac{d^2s}{dt^2}\frac{d\gamma}{ds}$$

$$\frac{d^3\gamma}{dt^3} = \left(\frac{ds}{dt}\right)^3 \frac{d^3\gamma}{ds^3} + 3\frac{ds}{dt}\frac{d^2s}{dt^2}\frac{d^2\gamma}{ds^2} + \frac{d^3s}{dt^3}\frac{d\gamma}{ds}$$

Hence,

$$\dot{\gamma} \times \ddot{\gamma} = \left(\frac{ds}{dt}\right)^3 \left(\frac{d\gamma}{ds} \times \frac{d^2\gamma}{ds^2}\right),$$

$$\dddot{\gamma} \cdot (\dot{\gamma} \times \ddot{\gamma}) = \left(\frac{ds}{dt}\right)^6 \left(\frac{d^3\gamma}{ds^3} \cdot \left(\frac{d\gamma}{ds} \times \frac{d^2\gamma}{ds^2}\right)\right)$$

So the torsion of γ is

$$\tau = \frac{\left(\frac{d^3\gamma}{ds^2} \cdot \left(\frac{d\gamma}{ds} \times \frac{d^2\gamma}{ds^2}\right)\right)}{\left\|\frac{d\gamma}{ds} \times \frac{d^2\gamma}{ds^2}\right\|^2} = \frac{\dddot{\gamma} \cdot (\dot{\gamma} \times \ddot{\gamma})}{\|\dot{\gamma} \times \ddot{\gamma}\|^2}.$$

Example 2.8.3. We compute the torsion of the circular helix $\gamma(\theta) = (a\cos\theta, a\sin\theta, b\theta)$ studied in Example 2.6.5 We have

$$\dot{\gamma}(\theta) = (-a\sin\theta, a\cos\theta, b), \quad \ddot{\gamma}(\theta) = (-a\cos\theta, -a\sin\theta, 0),$$
$$\dddot{\gamma}(\theta) = (a\sin\theta, -a\cos\theta, 0), \quad \dot{\gamma}\times\ddot{\gamma} = \left(ab\sin\theta, -ab\cos\theta, a^2\right)$$

$$\|\dot{\gamma}\times\ddot{\gamma}\|^2 = a^2\left(a^2+b^2\right), \quad (\dot{\gamma}\times\ddot{\gamma})\cdot\dddot{\gamma} = a^2b,$$

so the torsion

$$\tau = \frac{(\dot{\gamma}\times\ddot{\gamma})\cdot\dddot{\gamma}}{\|\dot{\gamma}\times\ddot{\gamma}\|^2} = \frac{a^2b}{a^2\left(a^2+b^2\right)} = \frac{b}{a^2+b^2}.$$

Note that the torsion of the circular helix in becomes zero when $b = 0$, in which case the helix is just a circle in the xy-plane.

Proposition 2.8.4. Let γ be a regular curve in $\mathbb{R}^3$ with nowhere vanishing curvature (so that the torsion τ of γ is defined). Then, the image of γ is contained in a plane if and only if τ is zero at every point of the curve.

Proof. We can assume that γ is unit-speed (for this can be achieved by reparametrizing γ, and reparametrizing changes neither the torsion nor the fact that γ is, or is not, contained in a plane). We denote the parameter of γ by s and d/ds by a dot as usual.

Suppose first that the image of γ is contained in the plane $\mathbf{v}\cdot\mathbf{N} = d$, where $\mathbf{N}$ is a constant vector and d is a constant scalar and $\mathbf{v}\in\mathbb{R}^3$. We can assume that $\mathbf{N}$ is a unit vector. Differentiating $\gamma\cdot\mathbf{N} = d$ with respect to s, we get

$$\mathbf{t}\cdot\mathbf{N} = 0 \tag{2.19}$$

$$\begin{aligned} \therefore \quad \dot{\mathbf{t}}\cdot\mathbf{N} &= 0, \quad (\text{since } \dot{\mathbf{N}} = 0) \\ \therefore \quad \kappa\mathbf{n}\cdot\mathbf{N} &= 0 \quad (\text{since } \dot{\mathbf{t}} = \kappa\mathbf{n} \\ \therefore \quad \mathbf{n}\cdot\mathbf{N} &= 0 \quad (\text{since } \kappa\neq 0). \end{aligned} \tag{2.20}$$

Equations 2.19 and 2.20 show that $\mathbf{t}$ and $\mathbf{n}$ are perpendicular to $\mathbf{N}$. It follows that $\mathbf{b} = \mathbf{t}\times\mathbf{n}$ is parallel to $\mathbf{N}$. Since $\mathbf{N}$ and $\mathbf{b}$ are both unit vectors, and $\mathbf{b}(s)$ is a smooth (hence continuous) function of s, we must have $\mathbf{b}(s) = \mathbf{N}$ for all s or $\mathbf{b}(s) = -\mathbf{N}$ for all s. In both cases, $\mathbf{b}$ is a constant vector. But then $\dot{\mathbf{b}} = 0$, so $\tau = 0$.

Conversely, suppose that $\tau = 0$ everywhere. By Equation 2.17, $\dot{\mathbf{b}} = 0$, so $\mathbf{b}$ is a constant vector. The first part of the proof suggests that γ should be contained in a plane $\mathbf{v}\cdot\mathbf{b} =$ constant. We therefore consider

$$\frac{d}{ds}(\gamma\cdot\mathbf{b}) = \dot{\gamma}\cdot\mathbf{b} = \mathbf{t}\cdot\mathbf{b} = 0$$

so $\gamma \cdot \mathbf{b}$ is a constant (scalar), say d. This means that γ is indeed contained in the plane $\mathbf{v} \cdot \mathbf{b} = d$. □

Theorem 2.8.5. Let γ be a unit-speed curve in $\mathbb{R}^3$ with nowhere vanishing curvature. Then,

$$\begin{aligned} \dot{\mathbf{t}} &= \kappa \mathbf{n} \\ \dot{\mathbf{n}} &= -\kappa \mathbf{t} + \tau \mathbf{b} \\ \dot{\mathbf{b}} &= -\tau \mathbf{n} \end{aligned} \tag{2.21}$$

Equations 2.21 are called the *Frenet-Serret equations*. Notice that the matrix

$$\begin{pmatrix} 0 & \kappa & 0 \\ -\kappa & 0 & \tau \\ 0 & -\tau & 0 \end{pmatrix}$$

which expresses $\dot{\mathbf{t}}, \dot{\mathbf{n}}, \dot{\mathbf{b}}$ in terms of $\mathbf{t}, \mathbf{n}$, and $\mathbf{b}$ is *skew-symmetric*, i.e., it is equal to the negative of its transpose. This helps when trying to remember the equations.

Proof. We have

$$\dot{\mathbf{t}} = \kappa \mathbf{n} \text{ and } \dot{\mathbf{b}} = -\tau \mathbf{n}$$

.Since $\{\mathbf{t}, \mathbf{n}, \mathbf{b}\}$ is a right-handed orthonormal basis of $\mathbb{R}^3$,

$$\mathbf{t} \times \mathbf{n} = \mathbf{b}, \quad \mathbf{n} \times \mathbf{b} = \mathbf{t}, \quad \mathbf{b} \times \mathbf{t} = \mathbf{n}.$$

Hence,

$$\dot{\mathbf{n}} = \dot{\mathbf{b}} \times \mathbf{t} + \mathbf{b} \times \dot{\mathbf{t}} = -\tau \mathbf{n} \times \mathbf{t} + \kappa \mathbf{b} \times \mathbf{n} = -\kappa \mathbf{t} + \tau \mathbf{b}$$

□

Proposition 2.8.6. Let γ be a unit-speed curve in $\mathbb{R}^3$ with constant curvature and zero torsion. Then, γ is a parametrization of (part of) a circle.

Proof. By the proof of Proposition 2.8.4, the binormal $\mathbf{b}$ is a constant vector and γ is contained in a plane Π, say, perpendicular to $\mathbf{b}$. Now

$$\frac{d}{ds}\left(\gamma + \frac{1}{\kappa}\mathbf{n}\right) = \mathbf{t} + \frac{1}{\kappa}\dot{\mathbf{n}} = \mathbf{0}$$

using the fact that the curvature κ is constant and the Frenet-Serret equation

$$\dot{\mathbf{n}} = -\kappa \mathbf{t} + \tau \mathbf{b} = -\kappa \mathbf{t} \quad (\text{since } \tau = 0)$$

Hence, $\gamma + \frac{1}{\kappa}\mathbf{n}$ is a constant vector, say $\mathbf{a}$, and we have

$$\|\gamma - \mathbf{a}\| = \left\| -\frac{1}{\kappa}\mathbf{n} \right\| = \frac{1}{\kappa}.$$

This shows that γ lies on the sphere $\mathcal{S}$, say, with centre $\mathbf{a}$ and radius $\frac{1}{\kappa}$. The intersection of Π and $\mathcal{S}$ is a circle, say $\mathcal{C}$, and we have shown that γ is a parametrization of part of $\mathcal{C}$. If r is the radius of $\mathcal{C}$, we have $\kappa = \frac{1}{r}$ so $r = \frac{1}{\kappa}$ is also the radius of $\mathcal{S}$. It follows that $\mathcal{C}$ is a *great circle* on $\mathcal{S}$, i.e., that Π passes through the centre $\mathbf{a}$ of $\mathcal{S}$. Thus, $\mathbf{a}$ is the centre of $\mathcal{C}$ and the equation of Π is $\mathbf{v} \cdot \mathbf{b} = \mathbf{a} \cdot \mathbf{b}$. □

Exercise 2.8.7. Let P be an $n \times n$ orthogonal matrix and let $\mathbf{a} \in \mathbb{R}^n$, so that $M(\mathbf{v}) = P\mathbf{v} + \mathbf{a}$ is an isometry of $\mathbb{R}^3$. Show that, if γ is a unit-speed curve in $\mathbb{R}^n$, the curve $\Gamma = M(\gamma)$ is also unit-speed. Show also that, if $\mathbf{t}, \mathbf{n}, \mathbf{b}$ and $\mathbf{T}, \mathbf{N}, \mathbf{B}$ are the tangent vector, principal normal and binormal of γ and Γ, respectively, then $\mathbf{T} = P\mathbf{t}, \mathbf{N} = P\mathbf{n}$ and $\mathbf{B} = P\mathbf{b}$.

Answer: $\dot{\Gamma} = P\dot{\gamma}$ so $\mathbf{T} = P\mathbf{t}$ and $\|\dot{\Gamma}\|^2 = (P\dot{\gamma}) \cdot (P\dot{\gamma}) = \dot{\gamma} \cdot \dot{\gamma}$ since P is orthogonal. Then, $\ddot{\Gamma} = P\ddot{\gamma}$, taking lengths shows that γ and Γ have the same curvature κ, and then dividing by κ gives $\mathbf{N} = P\mathbf{n}$. Then $\mathbf{B} = P\mathbf{t} \times P\mathbf{n}$. If P corresponds to a direct isometry, this is equal to $P(\mathbf{t} \times \mathbf{n}) = P\mathbf{b}$, but if P corresponds to an opposite isometry, $P\mathbf{t} \times P\mathbf{n} = -P\mathbf{b}$.

Theorem 2.8.8. Let $\gamma(s)$ and $\tilde{\gamma}(s)$ be two unit-speed curves in $\mathbb{R}^3$ with the same curvature $\kappa(s) > 0$ and the same torsion $\tau(s)$ for all s. Then, there is a direct isometry M of $\mathbb{R}^3$ such that

$$\tilde{\gamma}(s) = M(\gamma(s)) \text{ for all } s$$

Further, if k and t are smooth functions with $k > 0$ everywhere, there is a unit-speed curve in $\mathbb{R}^3$ whose curvature is k and whose torsion is t.

Proof. Let $\mathbf{t}, \mathbf{n}$ and $\mathbf{b}$ be the tangent vector, principal normal and binormal of γ, and let $\tilde{\mathbf{t}}, \tilde{\mathbf{n}}$ and $\tilde{\mathbf{b}}$ be those of $\tilde{\gamma}$. Let s_0 be a fixed value of the parameter s, let θ be the angle between $\mathbf{t}(s_0)$ and $\tilde{\mathbf{t}}(s_0)$ and let ρ be the rotation through an angle θ around the axis passing through the origin and perpendicular to both of these vectors. Then, $\rho(\mathbf{t}(s_0)) = \tilde{\mathbf{t}}(s_0)$; let $\hat{\mathbf{n}} = \rho(\mathbf{n}(s_0)), \hat{\mathbf{b}} = \rho(\mathbf{b}(s_0))$. If φ is the angle between $\hat{\mathbf{n}}$ and $\tilde{\mathbf{n}}(s_0)$, let ρ' be the rotation through an angle φ around the axis passing through the origin parallel to $\tilde{\mathbf{t}}(s_0)$. Then, ρ' fixes $\tilde{\mathbf{t}}(s_0)$ and takes $\hat{\mathbf{n}}$ to $\tilde{\mathbf{n}}(s_0)$. Since $\{\mathbf{t}(s_0), \mathbf{n}(s_0), \mathbf{b}(s_0)\}$ and $\left\{\tilde{\mathbf{t}}(s_0), \tilde{\mathbf{n}}(s_0), \tilde{\mathbf{b}}(s_0)\right\}$ are both right- handed orthonormal bases of $\mathbb{R}^3$,

$\rho' \circ \rho$ takes the vectors $\mathbf{t}(s_0), \mathbf{n}(s_0), \mathbf{b}(s_0)$ to the vectors $\tilde{\mathbf{t}}(s_0), \tilde{\mathbf{n}}(s_0), \tilde{\mathbf{b}}(s_0)$, respectively. Now let M be the direct isometry $M = T_{\tilde{\gamma}(s_0)-\gamma(s_0)} \circ \rho' \circ \rho$. By Exercise 2.8.7, the curve $\Gamma = M(\gamma)$ is unit- speed, and if $\mathbf{T}, \mathbf{N}$ and $\mathbf{B}$ denote its unit tangent vector, principal normal and binormal, we have

$$\Gamma(s_0) = \tilde{\gamma}(s_0), \mathbf{T}(s_0) = \tilde{\mathbf{t}}(s_0), \mathbf{N}(s_0) = \tilde{\mathbf{n}}(s_0), \mathbf{B}(s_0) = \tilde{\mathbf{b}}(s_0) \tag{2.22}$$

Consider the expression

$$A(s) = \tilde{\mathbf{t}} \cdot \mathbf{T} + \tilde{\mathbf{n}} \cdot \mathbf{N} + \tilde{\mathbf{b}} \cdot \mathbf{B}$$

By Equation 2.22, we have $A(s_0) = 3$. On the other hand, since $\tilde{\mathbf{t}}$ and $\mathbf{T}$ are unit vectors, $\tilde{\mathbf{t}} \cdot \mathbf{T} \leq 1$, with equality holding if and only if $\tilde{\mathbf{t}} = \mathbf{T}$; and similarly for $\tilde{\mathbf{n}} \cdot \mathbf{N}$ and $\tilde{\mathbf{b}} \cdot \mathbf{B}$. It follows that $A(s) \leq 3$, with equality holding if and only if $\tilde{\mathbf{t}} = \mathbf{T}, \tilde{\mathbf{n}} = \mathbf{N}$ and $\tilde{\mathbf{b}} = \mathbf{B}$. Thus, if we can prove that A is constant, it will follow in particular that $\tilde{\mathbf{t}} = \mathbf{T}$, i.e., that $\dot{\tilde{\gamma}} = \dot{\Gamma}$, and hence that $\tilde{\gamma}(s) - \Gamma(s)$ is a constant. But by Equation 2.22 again, this constant vector must be zero, so $\tilde{\gamma} = \Gamma$.

For the first part of the theorem, we are therefore reduced to proving that A is constant. But, using the Frenet-Serret equations,

$$\begin{aligned}\dot{A} =& \dot{\tilde{\mathbf{t}}} \cdot \mathbf{T} + \dot{\tilde{\mathbf{n}}} \cdot \mathbf{N} + \dot{\tilde{\mathbf{b}}} \cdot \mathbf{B} + \tilde{\mathbf{t}} \cdot \dot{\mathbf{T}} + \tilde{\mathbf{n}} \cdot \dot{\mathbf{N}} + \tilde{\mathbf{b}} \cdot \dot{\mathbf{B}} \\ =& \kappa \tilde{\mathbf{n}} \cdot \mathbf{T} + (-\kappa \tilde{\mathbf{t}} + \tau \tilde{\mathbf{b}}) \cdot \mathbf{N} + (-\tau \tilde{\mathbf{n}}) \cdot \mathbf{B} + \tilde{\mathbf{t}} \cdot \kappa \mathbf{N} \\ & + \tilde{\mathbf{n}} \cdot (-\kappa \mathbf{T} + \tau \mathbf{B}) + \tilde{\mathbf{b}} \cdot (-\tau \mathbf{N}),\end{aligned}$$

and this vanishes since the terms cancel in pairs.

For the second part of the theorem, we observe first that it follows from the theory of ordinary differential equations that the equations

$$\dot{\mathbf{T}} = k\,\mathbf{N}, \tag{2.23}$$

$$\dot{\mathbf{N}} = -k\mathbf{T} + t\,\mathbf{B}, \tag{2.24}$$

$$\dot{\mathbf{B}} = -t\,\mathbf{N} \tag{2.25}$$

have a unique solution $\mathbf{T}(s), \mathbf{N}(s), \mathbf{B}(s)$ such that $\mathbf{T}(s_0), \mathbf{N}(s_0), \mathbf{B}(s_0)$ are the standard orthonormal vectors $\mathbf{i} = (1,0,0), \mathbf{j} = (0,1,0), \mathbf{k} = (0,0,1)$, respectively. Since the matrix

$$\begin{pmatrix} 0 & k & 0 \\ -k & 0 & t \\ 0 & -t & 0 \end{pmatrix}$$

expressing $\dot{\mathbf{T}}, \dot{\mathbf{N}}$ and $\dot{\mathbf{B}}$ in terms of $\mathbf{T}, \mathbf{N}$ and $\mathbf{B}$ is skew-symmetric, it follows that the vectors $\mathbf{T}, \mathbf{N}$ and $\mathbf{B}$ are orthonormal for all values of s.

Now define

$$\gamma(s) = \int_{s_0}^{s} \mathbf{T}(u)du.$$

Then, $\dot{\gamma} = \mathbf{T}$, so since $\mathbf{T}$ is a unit vector, γ is unit-speed. Next, $\dot{\mathbf{T}} = k\mathbf{N}$ by Equation. 2.23, so since $\mathbf{N}$ is a unit vector, k is the curvature of γ and $\mathbf{N}$ is its principal normal. Next, since $\mathbf{B}$ is a unit vector perpendicular to $\mathbf{T}$ and $\mathbf{N}$, $\mathbf{B} = \lambda\mathbf{T} \times \mathbf{N}$ where λ is a smooth function of s that is equal to ± 1 for all s. Since $\mathbf{k} = \mathbf{i} \times \mathbf{j}$, we have $\lambda(s_0) = 1$, so it follows that $\lambda(s) = 1$ for all s. Hence, $\mathbf{B}$ is the binormal of γ and by Equation 2.25, t is its torsion. □

Exercise 2.8.9. Compute $\kappa, \tau, \mathbf{t}, \mathbf{n}$ and $\mathbf{b}$ for each of the following curves, and verify that the Frenet-Serret equations are satisfied:

(i) $\gamma(t) = \left(\frac{1}{3}(1+t)^{3/2}, \frac{1}{3}(1-t)^{3/2}, \frac{t}{\sqrt{2}}\right)$.

(ii) $\gamma(t) = \left(\frac{4}{5}\cos t, 1 - \sin t, -\frac{3}{5}\cos t\right)$.

Show that the curve in (ii) is a circle, and find its centre, radius and the plane in which it lies.

Solution:

(i) $\mathbf{t} = \left(\frac{1}{2}(1+t)^{1/2}, -\frac{1}{2}(1-t)^{1/2}, \frac{1}{\sqrt{2}}\right)$ is a unit vector so γ is unit-speed; $\dot{\mathbf{t}} = \left(\frac{1}{4}(1+t)^{-1/2}, \frac{1}{4}(1-t)^{-1/2}, 0\right)$, so $\kappa = \|\dot{\mathbf{t}}\| = 1/\sqrt{8(1-t^2)}$; $\mathbf{n} = \frac{1}{\kappa}\dot{\mathbf{t}} = \frac{1}{\sqrt{2}}\left((1-t)^{1/2}, (1+t)^{1/2}, 0\right)$; $\mathbf{b} = \mathbf{t} \times \mathbf{n} = \left(-\frac{1}{2}(1+t)^{1/2}, \frac{1}{2}(1-t)^{1/2}, \frac{1}{\sqrt{2}}\right)$ $\dot{\mathbf{b}} = \left(-\frac{1}{4}(1+t)^{-1/2}, -\frac{1}{4}(1-t)^{-1/2}, 0\right)$ so the torsion $\tau = 1/\sqrt{8(1-t^2)}$ The equation $\dot{\mathbf{n}} = -\kappa\mathbf{t} + \tau\mathbf{n}$ is easily checked.

(ii) $\mathbf{t} = \left(-\frac{4}{5}\sin t, -\cos t, \frac{3}{5}\sin t\right)$ is a unit vector so γ is unit-speed; $\dot{\mathbf{t}} = \left(-\frac{4}{5}\cos t, \sin t\ \frac{3}{5}\cos t\right)$; $\mathbf{b} = \mathbf{t} \times \mathbf{n} = \left(-\frac{3}{5}, 0, -\frac{4}{5}\right)$, so $\dot{\mathbf{b}} = \mathbf{0}$ and $\tau = 0$. By the proof of Proposition 2.8.6,γ is a circle of radius $1/\kappa = 1$ with centre $\gamma + \frac{1}{\kappa}\mathbf{n} = (0, 1, 0)$ in the plane passing through $(0, 1, 0)$ perpendicular to $\mathbf{b} = \left(-\frac{3}{5}, 0, -\frac{4}{5}\right)$, i.e., the plane $3x + 4z = 0$.

2.9 What is a surface?

Definition 2.9.1. First, a subset U of $\mathbb{R}^n$ is called *open* if, whenever $\mathbf{a}$ is a point in U, there is a positive number ϵ such that every point $\mathbf{u} \in \mathbb{R}^n$ within a distance ϵ of $\mathbf{a}$ is also in U:

$$\mathbf{a} \in U \text{ and } \|\mathbf{u}-\mathbf{a}\| < \epsilon \implies u \in U.$$

For example, the whole of $\mathbb{R}^n$ is an open set, as is

$$\mathcal{D}_r(\mathbf{a}) = \{\mathbf{u} \in \mathbb{R}^n \mid \|\mathbf{u}-\mathbf{a}\| < r\},$$

the *open ball* with centre $\mathbf{a}$ and radius $r > 0$. (If $n = 1$, an open ball is called an *open interval*; if $n = 2$ it is called an *open disc*.) However,

$$\overline{\mathcal{D}}_r(\mathbf{a}) = \{\mathbf{u} \in \mathbb{R}^n \mid \|\mathbf{u}-\mathbf{a}\| \leq r\}$$

is *not* open because however small the positive number ϵ is, there is a point within a distance ϵ of the point $(a_1 + r, a_2, \ldots, a_n) \in \overline{\mathcal{D}}_r(\mathbf{a})$ that is not in $\overline{\mathcal{D}}_r(\mathbf{a})$ (for example, the point $\left(a_1 + r + \frac{\epsilon}{2}, a_2, \ldots, a_n\right)$).
Next, if X and Y are subsets of $\mathbb{R}^m$ and $\mathbb{R}^n$, respectively, a map $f : X \to Y$ is said to be continuous at a point $\mathbf{a} \in X$ if points in X near $\mathbf{a}$ are mapped by f to points in Y near $f(\mathbf{a})$. More precisely, f is *continuous* at $\mathbf{a}$ if, given any number $\epsilon > 0$, there is a number $\delta > 0$ such that

$$\mathbf{u} \in X \text{ and } \|\mathbf{u}-\mathbf{a}\| < \delta \implies \|f(\mathbf{u}) - f(\mathbf{a})\| < \epsilon$$

Then f is said to be *continuous* if it is continuous at every point of X. Composites of continuous maps are continuous.

In view of the definition of an open set, this is equivalent to the following: f is continuous if and only if, for any open set V of $\mathbb{R}^n$, there is an open set U of $\mathbb{R}^m$ such that $U \cap X = \{x \in X \mid f(x) \in V\}$.

If $f : X \to Y$ is continuous and bijective, and if its inverse map $f^{-1} : Y \to X$ is also continuous, then f is called a *homeomorphism* and X and Y are said to be *homeomorphic*.

Definition 2.9.2. A subset $\mathcal{S}$ of $\mathbb{R}^3$ is a *surface* if, for every point $\mathbf{p} \in \mathcal{S}$, there is an open set U in $\mathbb{R}^2$ and an open set W in $\mathbb{R}^3$ containing $\mathbf{p}$ such that $\mathcal{S} \cap W$ is homeomorphic to U. A subset of a surface $\mathcal{S}$ of the form $\mathcal{S} \cap W$, where W is an open subset of $\mathbb{R}^3$, is called an *open subset* of $\mathcal{S}$. A homeomorphism $\sigma : \mathbf{U} \to \mathbf{S} \cap \mathbf{W}$ as in this definition is called a *surface patch* or *parametrization* of the open subset $\mathcal{S} \cap W$ of $\mathcal{S}$. A collection of such surface patches whose images cover the whole of $\mathcal{S}$ is called an *atlas* of $\mathcal{S}$.

Example 2.9.3. Every plane in $\mathbb{R}^3$ is a surface with an atlas consisting of a single surface patch. In fact, let $\mathbf{a}$ be a point on the plane, and let $\mathbf{p}$ and $\mathbf{q}$ be two unit vectors that are parallel to the plane and perpendicular to each other. If $\mathbf{v}$ is any point of the plane, $\mathbf{v} - \mathbf{a}$ is parallel to the plane, and so

$$\mathbf{v} - \mathbf{a} = u\mathbf{p} + v\mathbf{q}$$

for some scalars u and v. Thus, the desired surface patch is

$$\sigma(u, v) = \mathbf{a} + u\mathbf{p} + v\mathbf{q}$$

and its inverse map is

$$\sigma^{-1}(\mathbf{v}) = ((\mathbf{v} - \mathbf{a}) \cdot \mathbf{p}, (\mathbf{v} - \mathbf{a}) \cdot \mathbf{q}).$$

These formulas make it clear that σ and σ^{-1} are continuous, and hence that σ is a homeomorphism.

Example 2.9.4. A *circular cylinder* is the set of points of $\mathbb{R}^3$ that are at a fixed distance (the *radius* of the cylinder) from a fixed straight line (its *axis*). For example, the circular cylinder of radius 1 and axis the z-axis, which we shall call the *unit cylinder*, is

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

The simplest parametrization of $\mathcal{S}$ is

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

Clearly, $\sigma(u, v) \in \mathcal{S}$ for all $(u, v) \in \mathbb{R}^2$, and every point of $\mathcal{S}$ is of this form. Moreover, σ is continuous. However, σ is not injective, and so is not a homeomorphism, because $\sigma(u, v) = \sigma(u + 2\pi, v)$ for all (u, v). To get an injective map we can restrict u to lie in an interval of length $\leq 2\pi$, say $0 \leq u < 2\pi$. However, although the restriction $\sigma|_V$ of σ to

$$V = \left\{(u, v) \in \mathbb{R}^2 \mid 0 \leq u < 2\pi\right\}$$

is injective, V is not an open subset of $\mathbb{R}^2$ and so $\sigma|_V$ is not a surface patch. The largest open subset of $\mathbb{R}^2$ contained in V is

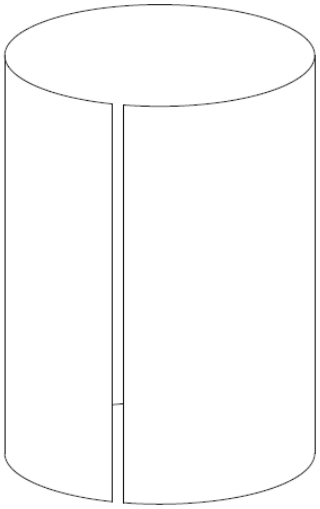
$$U = \left\{(u, v) \in \mathbb{R}^2 \mid 0 < u < 2\pi\right\}$$

and the restriction $\sigma|_U$ of σ to U is a surface patch. However, $\sigma|_U$ does not cover the whole of $\mathcal{S}$, but only the open subset obtained by removing the line $x = 1, y = 0$ from $\mathcal{S}$.

To get an atlas for $\mathcal{S}$ we therefore need at least one more surface patch. We can take $\sigma|_{\tilde{U}}$, where

$$\tilde{U} = \left\{(u, v) \in \mathbb{R}^2 \mid -\pi < u < \pi\right\};$$

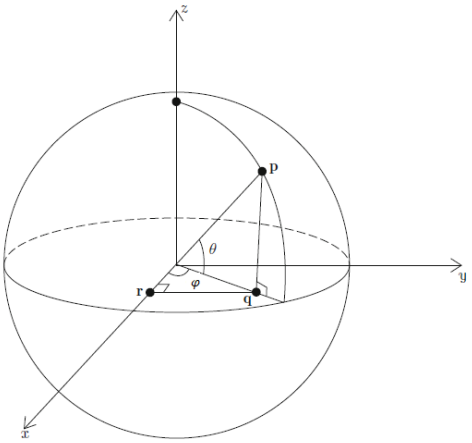
this covers the open subset of $\mathcal{S}$ obtained by removing the line $x = -1, y = 0$. Every point of $\mathcal{S}$ is in the image of at least one of the surface patches $\sigma|_U$, $\sigma|_{\tilde{U}}$, so $\{\sigma|_U, \sigma|_{\tilde{U}}\}$ is an atlas for $\mathcal{S}$, and $\mathcal{S}$ is a surface.



Example 2.9.5. A sphere is the set of points of $\mathbb{R}^3$ that are a fixed distance (the *radius* of the sphere) from a fixed point (its *centre*). For example, the sphere of radius 1 and centre the origin, called the *unit sphere*, is

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

A parametrization of S^2 is that given by latitude θ and longitude φ : if $\mathbf{p}$ is a point of the sphere, the line through $\mathbf{p}$ parallel to the z-axis intersects the xy-plane at a point $\mathbf{q}$.Let θ is the angle between $\mathbf{q}$ and $\mathbf{p}$ and φ is the angle between $\mathbf{q}$ and the positive x-axis. Then the coordinate of $\mathbf{p}$ can be express in



terms of the angles θ and φ. From the right-angled triangle with vertices $\mathbf{0}$, $\mathbf{p}$ and $\mathbf{q}$, we see that the z-component of $\mathbf{p}$ is $\sin\theta$. The x - and y-components

can be found from the right-angled triangle in the xy-plane with vertices $\mathbf{0}, \mathbf{q}$ and $\mathbf{r}$, where $\mathbf{r}$ is the foot of the perpendicular from $\mathbf{q}$ to the x-axis. The length of the hypotenuse of this triangle is $\|\mathbf{q}\| = \cos\theta$, so the x - and y-components of $\mathbf{p}$ are

$$\|\mathbf{q}\| \cos\varphi = \cos\theta\cos\varphi \text{ and } \|\mathbf{q}\| \sin\varphi = \cos\theta\sin\varphi,$$

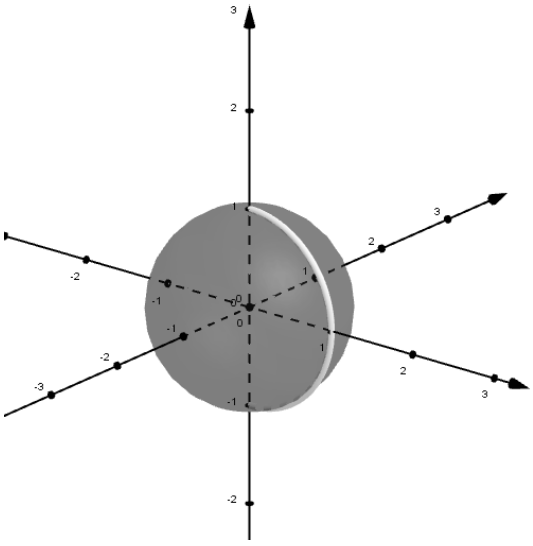
respectively. Putting all these together gives

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

Denote the right-hand side of this equation by $\sigma(\theta, \varphi)$; this is the *latitude - longitude parametrization* of S^2.

As in the case of the cylinder, σ is not injective since (for example) $\sigma(\theta, \varphi) = \sigma(\theta, \varphi + 2\pi)$. In fact, a little thought shows that to cover the whole sphere, it is sufficient to take

$$-\frac{\pi}{2} \leq \theta \leq \frac{\pi}{2}, \quad 0 \leq \varphi \leq 2\pi.$$



However, the set of points (θ, φ) satisfying these inequalities is not an open subset of $\mathbb{R}^2$. The largest open set consistent with the above inequalities is

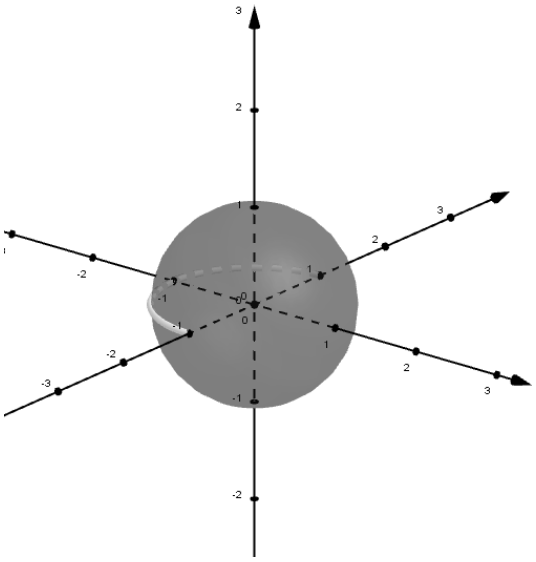
$$U = \left\{(\theta, \varphi) \mid -\frac{\pi}{2} < \theta < \frac{\pi}{2}, 0 < \varphi < 2\pi\right\};$$

however, the image of $\sigma|_U$ is not the whole of the sphere, but the open subset obtained by removing the semicircle $\mathcal{C}$ consisting of the points of the sphere of the form $(x, 0, z)$ with $x \geq 0$.

To show that the sphere is a surface, we must therefore produce at least one more surface patch covering the part of the sphere omitted by σ. One possibility is the patch $\tilde{\sigma}$ obtained by first rotating σ by π about the z-axis and then by $\pi/2$ about the x-axis. Explicitly, $\tilde{\sigma} : U \to \mathbb{R}^3$ is given by

$$\tilde{\boldsymbol{\sigma}}(\theta, \varphi) = (-\cos\theta\cos\varphi, -\sin\theta, -\cos\theta\sin\varphi)$$

(the open set U is the same as for $\boldsymbol{\sigma}$). The image of $\tilde{\sigma}$ is the open subset of S^2 obtained by removing the semicircle $\tilde{\mathcal{C}}$ consisting of the points of the sphere of the form $(x, y, 0)$ with $x \leq 0$. It is clear that $\mathcal{C}$ and $\tilde{\mathcal{C}}$ do not intersect, so the union of the images of $\sigma|_U$ and $\tilde{\boldsymbol{\sigma}}|_U$ is the whole sphere.



Note:To rotate a point by an angle θ about the $x-$, $y-$, or z-axis in $\mathbb{R}^3$,use the following rotation matrices

$$R_x(\theta) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\theta & -\sin\theta \\ 0 & \sin\theta & \cos\theta \end{bmatrix}$$

$$R_y(\theta) = \begin{bmatrix} \cos\theta & 0 & \sin\theta \\ 0 & 1 & 0 \\ -\sin\theta & 0 & \cos\theta \end{bmatrix}$$

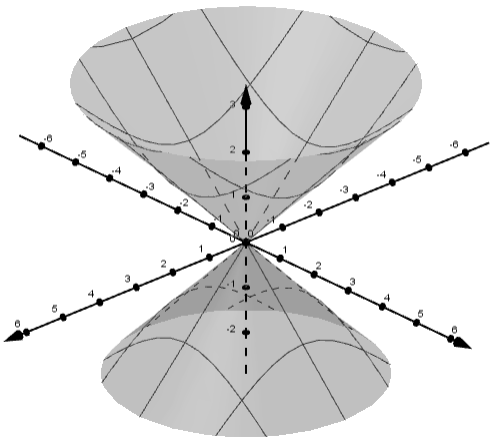
$$R_z(\theta) = \begin{bmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Example 2.9.6.

The *circular cone* with *vertex* a point $\mathbf{v}$, *axis* a straight line ℓ passing through $\mathbf{v}$, and *angle* α, where $0 < \alpha < \pi/2$, is the set of points $\mathbf{p}$ in $\mathbb{R}^3$ such that the straight line through $\mathbf{v}$ and $\mathbf{p}$ makes an angle α with the line ℓ. For example, if $\mathbf{v}$ is the origin, ℓ is the z-axis and $\alpha = \pi/4$, the circular cone is

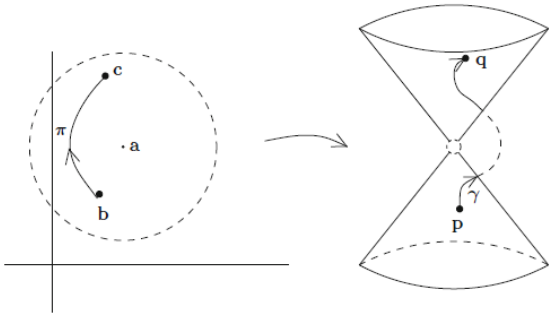
$$\mathcal{S} = \left\{(x, y, z) \in \mathbb{R}^3 \mid x^2 + y^2 = z^2\right\}.$$

To see that this is *not* a surface, suppose that $\sigma : U \to \mathcal{S} \cap W$ is a surface



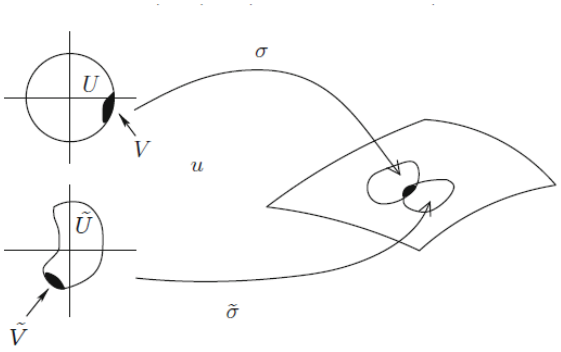
patch containing the vertex $(0, 0, 0)$ of the cone, and let $\mathbf{a} \in U$ correspond to the vertex. We can assume that U is an open ball with centre $\mathbf{a}$, since any open set U containing $\mathbf{a}$ must contain such an open ball. The open set W must obviously contain a point $\mathbf{p}$ in the lower half $\mathcal{S}_-$of $\mathcal{S}$ where $z < 0$ and

a point $\mathbf{q}$ in the upper half $\mathcal{S}_+$where $z > 0$; let $\mathbf{b}$ and $\mathbf{c}$ be the corresponding points in U. It is clear that there is a curve π in U passing through $\mathbf{b}$ and $\mathbf{c}$, but not passing through $\mathbf{a}$. This is mapped by σ into the curve $\gamma = \sigma \circ \pi$ lying entirely in $\mathcal{S}$, passing through $\mathbf{p}$ and $\mathbf{q}$, and not passing through the vertex. (It is true that γ will, in general, only be continuous, and not smooth, but this does not affect the argument.) This is clearly impossible.



If we remove the vertex, however, we do get a surface $\mathcal{S}_- \cup \mathcal{S}_+$. It has an atlas consisting of the two surface patches $\sigma_\pm : U \to \mathbb{R}^3$, where $U = \mathbb{R}^2 \backslash \{(0,0)\}$, given by the inverse of projection onto the xy plane:

$$\sigma_\pm(u,v) = \left(u, v, \pm\sqrt{u^2+v^2}\right)$$



As the example of the sphere shows, a point $\mathbf{a}$ of a surface $\mathcal{S}$ will generally lie in the image of more than one surface patch. In general, suppose then that $\sigma : U \to \mathcal{S} \cap W$ and $\tilde{\sigma} : \tilde{U} \to \mathcal{S} \cap \tilde{W}$ are two patches such that $\mathbf{a} \in \mathcal{S} \cap W \cap \tilde{W}$. Since σ and $\tilde{\sigma}$ are homeomorphisms, $\sigma^{-1}(\mathcal{S} \cap W \cap \tilde{W})$ and $\tilde{\sigma}^{-1}(\mathcal{S} \cap W \cap \tilde{W})$ are open sets $V \subseteq U$ and $\tilde{V} \subseteq \tilde{U}$, respectively. The composite homeomorphism

$\sigma^{-1} \circ \tilde{\sigma} : \tilde{V} \to V$ is called the *transition map* from σ to $\tilde{\sigma}$. If we denote this map by Φ, we have

$$\tilde{\sigma}(\tilde{u}, \tilde{v}) = \sigma(\Phi(\tilde{u}, \tilde{v}))$$

for all $(\tilde{u}, \tilde{v}) \in \tilde{V}$.

2.10 Smooth surfaces

Definition 2.10.1. If U is an open subset of $\mathbb{R}^m$, we say that a map $\mathbf{f} : U \to \mathbb{R}^n$ is *smooth* if each of the n components of $\mathbf{f}$, which are functions $U \to \mathbb{R}$, have continuous partial derivatives of all orders. The partial derivatives of $\mathbf{f}$ are then computed component wise. For example, if $m = 2$ and $n = 3$, and

$$\mathbf{f}(u, v) = (f_1(u, v), f_2(u, v), f_3(u, v)),$$

then

$$\frac{\partial \mathbf{f}}{\partial u} = \left(\frac{\partial f_1}{\partial u}, \frac{\partial f_2}{\partial u}, \frac{\partial f_3}{\partial u} \right), \quad \frac{\partial \mathbf{f}}{\partial v} = \left(\frac{\partial f_1}{\partial v}, \frac{\partial f_2}{\partial v}, \frac{\partial f_3}{\partial v} \right),$$

and similarly for higher derivatives. We often use the following abbreviations:

$$\frac{\partial \mathbf{f}}{\partial u} = \mathbf{f}_u, \quad \frac{\partial \mathbf{f}}{\partial v} = \mathbf{f}_v,$$

$$\frac{\partial^2 \mathbf{f}}{\partial u^2} = \mathbf{f}_{uu}, \quad \frac{\partial^2 \mathbf{f}}{\partial u \partial v} = \mathbf{f}_{uv}, \quad \frac{\partial^2 \mathbf{f}}{\partial v \partial u} = \mathbf{f}_{vu}, \quad \frac{\partial^2 \mathbf{f}}{\partial v^2} = \mathbf{f}_{vv},$$

and so on. From advanced calculus we know that $\mathbf{f}_{uv} = \mathbf{f}_{vu}$ if f is smooth.

Definition 2.10.2. A surface patch $\sigma : U \to \mathbb{R}^3$ is called *regular* if it is smooth and the vectors σ_u and σ_v are linearly independent at all points $(u, v) \in U$. Equivalently, σ should be smooth and the vector product $\sigma_u \times \sigma_v$ should be non-zero at every point of U.

Definition 2.10.3. If $\mathcal{S}$ is a surface, an *allowable surface patch* for $\mathcal{S}$ is a regular surface patch $\sigma : U \to \mathbb{R}^3$ such that σ is a homeomorphism from U to an open subset of $\mathcal{S}$. A *smooth surface* is a surface $\mathcal{S}$ such that, for any point $\mathbf{p} \in \mathcal{S}$, there is an allowable surface patch σ as above such that $\mathbf{p} \in \sigma(U)$. A collection $\mathcal{A}$ of allowable surface patches for a surface $\mathcal{S}$ such that every point of $\mathcal{S}$ is in the image of at least one patch in $\mathcal{A}$ is called an *atlas* for the smooth surface $\mathcal{S}$.

Example 2.10.4. The plane in Example 2.9.3 is a smooth surface. For

$$\sigma(u, v) = \mathbf{a} + u\mathbf{p} + v\mathbf{q}$$

is clearly smooth and $\sigma_u = \mathbf{p}$ and $\sigma_v = \mathbf{q}$ are linearly independent because $\mathbf{p}$ and $\mathbf{q}$ were chosen to be perpendicular unit vectors.

Example 2.10.5. The unit cylinder Example 2.9.4 is a smooth surface. Indeed,

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

is clearly smooth and

$$\sigma_u = (-\sin u, \cos u, 0), \quad \sigma_v = (0, 0, 1)$$

are obviously linearly independent for all (u, v), so $\sigma|_U$ and $\sigma|_{\tilde{U}}$ are regular surface patches.

Example 2.10.6. For the unit sphere S^2 in Example 2.9.5, it is clear that σ and $\tilde{\sigma}$ are smooth. As for regularity,

$$\sigma_\theta = (-\sin\theta\cos\varphi, -\sin\theta\sin\varphi, \cos\theta), \quad \sigma_\varphi = (-\cos\theta\sin\varphi, \cos\theta\cos\varphi, 0)$$
$$\sigma_\theta \times \sigma_\varphi = \big(-\cos^2\theta\cos\varphi, -\cos^2\theta\sin\varphi, -\sin\theta\cos\theta\big)$$

and hence $\|\sigma_\theta \times \sigma_\varphi\| = |\cos\theta|$. But if $(\theta, \varphi) \in U$, then $-\pi/2 < \theta < \pi/2$ so $\cos\theta \neq 0$. Similarly, we can check that $\tilde{\sigma}$ is regular.

Proposition 2.10.7. The transition maps of a smooth surface are smooth.

The proof of this will be given in Section 3.3. The next result is a kind of converse.

Proposition 2.10.8. Let U and $\tilde{U}$ be open subsets of $\mathbb{R}^2$ and let $\sigma : U \to \mathbb{R}^3$ be a regular surface patch. Let $\Phi : \tilde{U} \to U$ be a bijective smooth map with smooth inverse map $\Phi^{-1} : U \to \tilde{U}$. Then, $\tilde{\sigma} = \sigma \circ \Phi : \tilde{U} \to \mathbb{R}^3$ is a regular surface patch.

Proof. The patch $\tilde{\sigma}$ is smooth because any composite of smooth maps is smooth. As for regularity, let $(u, v) = \Phi(\tilde{u}, \tilde{v})$. By the chain rule,

$$\tilde{\sigma}_{\tilde{u}} = \frac{\partial u}{\partial \tilde{u}}\sigma_u + \frac{\partial v}{\partial \tilde{u}}\sigma_v, \quad \tilde{\sigma}_{\tilde{v}} = \frac{\partial u}{\partial \tilde{v}}\sigma_u + \frac{\partial v}{\partial \tilde{v}}\sigma_v$$

so

$$\tilde{\sigma}_{\tilde{u}} \times \tilde{\sigma}_{\tilde{v}} = \left(\frac{\partial u}{\partial \tilde{u}}\frac{\partial v}{\partial \tilde{v}} - \frac{\partial u}{\partial \tilde{v}}\frac{\partial v}{\partial \tilde{u}}\right)\sigma_u \times \sigma_v \tag{2.26}$$

The scalar on the right-hand side of this equation is the determinant of the *Jacobian matrix*

$$J(\Phi) = \begin{pmatrix} \frac{\partial u}{\partial \tilde{u}} & \frac{\partial v}{\partial \tilde{u}} \\ \\ \frac{\partial u}{\partial \tilde{v}} & \frac{\partial v}{\partial \tilde{v}} \end{pmatrix}$$

of Φ. We know that, if Ψ and $\tilde{\Psi}$ are two smooth maps between open sets of $\mathbb{R}^2$,

$$J(\tilde{\Psi} \circ \Psi) = J(\tilde{\Psi})J(\Psi)$$

Taking $\Psi = \Phi$ and $\tilde{\Psi} = \Phi^{-1}$, we see that $J(\Phi^{-1}) = J(\Phi)^{-1}$. In particular, $J(\Phi)$ is invertible, so its determinant is non-zero and Equation 2.26 shows that $\tilde{\sigma}$ is regular. □

If regular surface patches σ and $\tilde{\sigma}$ are related as in this proposition, we say that $\tilde{\sigma}$ is a *reparametrization* of σ, and that Φ is a *reparametrization map*. Note that σ is then a reparametrization of $\tilde{\sigma}$, since $\sigma = \tilde{\sigma} \circ \Phi^{-1}$.

Note also that, if $\sigma : U \to \mathcal{S} \cap W$ and $\tilde{\sigma} : \tilde{U} \to \mathcal{S} \cap \tilde{W}$ are two allowable surface patches of a smooth surface $\mathcal{S}$, and if $V \subseteq U$ and $\tilde{V} \subseteq \tilde{U}$ are the open subsets such that $\sigma(V) = \tilde{\sigma}(\tilde{V}) = \mathcal{S} \cap W \cap \tilde{W}$, then $\Phi = \sigma^{-1} \circ \tilde{\sigma} : \tilde{V} \to V$ is bijective, smooth and has a smooth inverse by Proposition 2.10.7. Thus, *$\tilde{\sigma}$ is a reparametrization of σ where they are both defined.*

In the rest of this book, by a surface we mean a smooth surface $\mathcal{S}$, and by a surface patch for $\mathcal{S}$ we mean an allowable surface patch for $\mathcal{S}$.

Exercise 2.10.9. Which of the following are regular surface patches (in each case, $u, v \in \mathbb{R}$)

(i) $\boldsymbol{\sigma}(u, v) = (u, v, uv)$.

(ii) $\sigma(u, v) = (u, v^2, v^3)$.

(iii) $\sigma(u, v) = (u + u^2, v, v^2)$

Solution:

(i) is clearly injective and is regular because σ is smooth and $\sigma_u \times \sigma_v = (-v, -u, 1)$ is never zero.

(ii) is injective but is not regular since $\boldsymbol{\sigma}_u \times \boldsymbol{\sigma}_v = (0, -3v^2, 2v)$ vanishes when $v = 0$.

(iii) is not injective because $\sigma(u, v) = \sigma(-u - 1, v)$ and is also not regular since $\sigma_u \times \sigma_v = 0, 2v(1 + 2u, 1 + 2u)$ vanishes when $u = -1/2$.

2.11 Smooth maps

Definition 2.11.1. Let $\mathcal{S}_1$ and $\mathcal{S}_2$ are smooth surfaces.Assume that $\mathcal{S}_1$ and $\mathcal{S}_2$ are covered by single surface patches $\sigma_1 : U_1 \to \mathbb{R}^3$ and $\sigma_2 : U_2 \to \mathbb{R}^3$.A function $f : \mathcal{S}_1 \to \mathcal{S}_2$ is smooth if $\sigma_2^{-1} \circ f \circ \sigma_1 : U_1 \to U_2$ is smooth.

Suppose that $\tilde{\sigma}_1 : \tilde{U}_1 \to \mathbb{R}^3$ and $\tilde{\sigma}_2 : \tilde{U}_2 \to \mathbb{R}^3$ are reparametrizations of σ_1 and σ_2, with reparametrization maps $\Phi_1 : \tilde{U}_1 \to U_1$ and $\Phi_2 : \tilde{U}_2 \to U_2$, respectively. Then if $\sigma_2^{-1} \circ f \circ \sigma_1 : U_1 \to U_2$ is smooth then the map $\tilde{\sigma}_2^{-1} \circ f \circ \tilde{\sigma}_1 : \tilde{U}_1 \to \tilde{U}_2$ is smooth , since

$$\begin{aligned}\tilde{\sigma}_2^{-1} \circ f \circ \tilde{\sigma}_1 &= \tilde{\sigma}_2^{-1} \circ \left(\sigma_2 \circ \sigma_2^{-1}\right) \circ f \circ \left(\sigma_1 \circ \sigma_1^{-1}\right) \circ \tilde{\sigma}_1 \\ &= \left(\tilde{\sigma}_2^{-1} \circ \sigma_2\right) \circ \left(\sigma_2^{-1} \circ f \circ \sigma_1\right) \circ \left(\sigma_1^{-1} \circ \tilde{\sigma}_1\right) \\ &= \Phi_2^{-1} \circ \left(\sigma_2^{-1} \circ f \circ \sigma_1\right) \circ \Phi_1\end{aligned}$$

and Φ_1 and Φ_2^{-1} are smooth maps (between open subsets of $\mathbb{R}^2$).

Definition 2.11.2. Let $\mathcal{S}_1$ and $\mathcal{S}_2$ are smooth surfaces.A function $f : \mathcal{S}_1 \to \mathcal{S}_2$ is bijective , smooth and whose inverse map $f^{-1} : \mathcal{S}_2 \to \mathcal{S}_1$ is smooth. Such maps are called *diffeomorphisms*, and $\mathcal{S}_1$ and $\mathcal{S}_2$ are said to be *diffeomorphic* if there is a diffeomorphism between them.

Proposition 2.11.3. Let $f : \mathcal{S}_1 \to \mathcal{S}_2$ be a diffeomorphism. If σ_1 is an allowable surface patch on $\mathcal{S}_1$, then $f \circ \sigma_1$ is an allowable surface patch on $\mathcal{S}_2$.

Proof. We can assume that $\mathcal{S}_1$ and $\mathcal{S}_2$ are covered by single allowable patches $\sigma_1 : U_1 \to \mathbb{R}^3$ and $\sigma_2 : U_2 \to \mathbb{R}^3$, respectively. Since f is a diffeomorphism, $f(\sigma_1(u, v)) = \sigma_2(F(u, v))$, where $F : U_1 \to U_2$ is bijective, smooth and F^{-1} is smooth. The result now follows from Proposition 2.10.7. □

Definition 2.11.4. A smooth map $f : \mathcal{S}_1 \to \mathcal{S}_2$ between smooth surfaces is called a *local diffeomorphism* if, for any point $\mathbf{p} \in \mathcal{S}_1$, there is an open subset $\mathcal{O}$ of $\mathcal{S}_1$ such that $f(\mathcal{O})$ is an open subset of $\mathcal{S}_2$ and $f|_{\mathcal{O}} : \mathcal{O} \to f(\mathcal{O})$ is a diffeomorphism. It is obvious that every diffeomorphism is a local diffeomorphism (take $\mathcal{O} = \mathcal{S}_1$).

Example 2.11.5. Consider the map from the yz-plane onto the unit cylinder $\mathcal{S}$, which wraps each line parallel to the y-axis around the 'waist' of the cylinder at height z above the xy-plane. This map is given by

$$f(0, y, z) = (\cos y, \sin y, z).$$

Clearly, f is not a diffeomorphism because it is not injective − the plane wraps around $\mathcal{S}$ infinitely many times. To see that f is a local diffeomorphism, we parametrize the yz-plane in the obvious way by the single surface patch $\pi(u, v) = (0, u, v)$, and use the atlas $\{\sigma|_U , \sigma|_{\tilde{U}}\}$ of $\mathcal{S}$ in Example 2.9.4. Let $\mathbf{p}$ be any point in the yz-plane, say $(0, a, b)$. If a is not an even multiple of 2π, there is an integer n such that $2n\pi < a < 2(n + 1)\pi$ and we have

$$f(\boldsymbol{\pi}(u, v)) = \sigma(u - 2n\pi, v) \quad \text{if} \quad 2n\pi < u < 2(n + 1)\pi$$

showing that f is a diffeomorphism from the open subset

$$\mathcal{O} = \{(0, y, z) \mid 2n\pi < y < 2(n+1)\pi\}$$

of the plane to the open subset

$$f(\mathcal{O}) = \{(x, y, z) \in \mathcal{S} \mid x \neq 1\}$$

of $\mathcal{S}$. If a is not an odd multiple of π, a similar argument works with $\sigma|_U$ replaced by $\sigma|_{\tilde{U}}$. Since a cannot both be an even and an odd multiple of π, we have shown that for all points $\mathbf{p}$ of the plane, there is an open subset $\mathcal{O}$ of the plane containing p that f maps diffeomorphically onto an open subset of $\mathcal{S}$.

2.12 Tangents and derivatives

Definition 2.12.1. A *tangent vector* to a surface $\mathcal{S}$ at a point $\mathbf{p} \in \mathcal{S}$ is the tangent vector at $\mathbf{p}$ of a curve in $\mathcal{S}$ passing through $\mathbf{p}$. The *tangent space* $T_{\mathbf{p}}\mathcal{S}$ of $\mathcal{S}$ at $\mathbf{p}$ is the set of all tangent vectors to $\mathcal{S}$ at $\mathbf{p}$.

To understand the tangent space $T_{\mathbf{p}}\mathcal{S}$, choose a surface patch $\sigma : U \to \mathbb{R}^3$ of $\mathcal{S}$ such that $\mathbf{p}$ is in the image of $\boldsymbol{\sigma}$, say $\boldsymbol{\sigma}(u_0, v_0) = \mathbf{p}$. If a curve $\boldsymbol{\gamma}$ lies in $\mathcal{S}$ and passes through $\mathbf{p}$ when $t = t_0$, say, there are functions $u(t)$ and $v(t)$ such that

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

for all values of t close to t_0, and $u(t_0) = u_0, v(t_0) = v_0$.

Proposition 2.12.2. Let $\sigma : U \to \mathbb{R}^3$ be a patch of a surface $\mathcal{S}$ containing a point $\mathbf{p} \in \mathcal{S}$, and let (u, v) be coordinates in U. The tangent space to $\mathcal{S}$ at $\mathbf{p}$ is the vector subspace of $\mathbb{R}^3$ spanned by the vectors σ_u and σ_v (the derivatives are evaluated at the point $(u_0, v_0) \in U$ such that $\sigma(u_0, v_0) = \mathbf{p}$).

Proof. Let γ be a smooth curve in $\mathcal{S}$, say

$$\gamma(t) = \sigma(u(t), v(t))$$

Denoting d/dt by a dot, we have, by the chain rule,

$$\dot{\gamma} = \sigma_u \dot{u} + \sigma_v \dot{v}.$$

Thus, $\dot{\gamma}$ is a linear combination of σ_u and σ_v.

Conversely, any vector in the vector subspace of $\mathbb{R}^3$ spanned by σ_u and σ_v is of the form $\lambda\sigma_u + \mu\sigma_v$ for some scalars λ and μ. Define

$$\gamma(t) = \sigma\left(u_0 + \lambda t, v_0 + \mu t\right).$$

Then, γ is a smooth curve in $\mathcal{S}$ and at $t = 0$, i.e., at the point $\mathbf{p} \in \mathcal{S}$, we have

$$\dot{\gamma} = \lambda\sigma_u + \mu\sigma_v$$

This shows that every vector in the span of σ_u and σ_v is the tangent vector at $\mathbf{p}$ of some curve in $\mathcal{S}$. □

Since σ is assumed to be regular, σ_u and σ_v are linearly independent so the tangent space is two-dimensional, and will be called the *tangent plane* from now on. Definition 2.12.1 shows that the tangent plane is independent of the choice of patch containing $\mathbf{p}$. Also , the vectors σ_u and σ_v that form a basis of the tangent plane at some point $\sigma\left(u_0, v_0\right)$ of the surface are the tangent vectors of the *parameter curves* on the surface, i.e., the curves $u \mapsto \sigma\left(u, v_0\right)$ and $v \mapsto \sigma\left(u_0, v\right)$.

Definition 2.12.3. Let $\mathbf{w} \in T_{\mathbf{p}}\mathcal{S}$ be a tangent vector to $\mathcal{S}$ at $\mathbf{p}$. By definition, $\mathbf{w}$ is the tangent vector at $\mathbf{p}$ of a curve γ in $\mathcal{S}$ passing through $\mathbf{p}$, say $\mathbf{w} = \dot{\gamma}\left(t_0\right)$. Then, $\tilde{\gamma} = f \circ \gamma$ is a curve in $\tilde{\mathcal{S}}$ passing through $f(\mathbf{p})$ when $t = t_0$, so $\tilde{\mathbf{w}} = \dot{\tilde{\gamma}}\left(t_0\right) \in T_{f(\mathbf{p})}\tilde{\mathcal{S}}$. Then we define derivative $D_{\mathbf{p}}f$ of f at the point $\mathbf{p} \in \mathcal{S}$ as the map $D_{\mathbf{p}}f : T_{\mathbf{p}}\mathcal{S} \to T_{f(\mathbf{p})}\tilde{\mathcal{S}}$ such that $D_{\mathbf{p}}f(\mathbf{w}) = \tilde{\mathbf{w}}$ for any tangent vector $\mathbf{w} \in T_{\mathbf{p}}\mathcal{S}$.

Note that $D_{\mathbf{p}}f(\mathbf{w})$ depends only on $f, \mathbf{p}$, and $\mathbf{w}$, For that Let $\sigma : U \to \mathbb{R}^3$ be a surface patch of $\mathcal{S}$ containing $\mathbf{p}$, say $\mathbf{p} = \sigma\left(u_0, v_0\right)$, and let α, β be the smooth functions on U such that

$$f(\sigma(u, v)) = \tilde{\sigma}(\alpha(u, v), \beta(u, v))$$

Let $\mathbf{w} = \lambda\boldsymbol{\sigma}_u + \mu\sigma_v$ be the tangent vector at $\mathbf{p}$ of a curve $\gamma(t) = \sigma(u(t), v(t))$, where u and v are smooth functions such that $\dot{u}\left(t_0\right) = \lambda, \dot{v}\left(t_0\right) = \mu$. Since the corresponding curve on $\tilde{\mathcal{S}}$ is $\tilde{\gamma}(t) = \tilde{\sigma}(\tilde{u}(t), \tilde{v}(t))$, where $\tilde{u}(t) = \alpha(u(t), v(t))$ and $\tilde{v}(t) = \beta(u(t), v(t))$, we have

$$D_{\mathbf{p}}f(\mathbf{w}) = \dot{\tilde{u}}\tilde{\sigma}_{\tilde{u}} + \dot{\tilde{v}}\tilde{\sigma}_{\tilde{v}} = \left(\dot{u}\alpha_u + \dot{v}\alpha_v\right)\tilde{\sigma}_{\tilde{u}} + \left(\dot{u}\beta_u + \dot{v}\beta_v\right)\tilde{\sigma}_{\tilde{v}}$$

the derivatives of u and v being evaluated at t_0. Thus,

$$D_{\mathbf{p}}f(\mathbf{w}) = \left(\lambda\alpha_u + \mu\alpha_v\right)\tilde{\sigma}_{\tilde{u}} + \left(\lambda\beta_u + \mu\beta_v\right)\tilde{\sigma}_{\tilde{v}} \tag{2.27}$$

The right-hand side depends only on $\mathbf{p}, f, \lambda$ and μ, i.e., on $\mathbf{p}, f$ and $\mathbf{w}$, as we want.

Proposition 2.12.4. If $f : \mathcal{S} \to \tilde{\mathcal{S}}$ is a smooth map between surfaces and $\mathbf{p} \in \mathcal{S}$, the derivative $D_{\mathbf{p}}f : T_{\mathbf{p}}\mathcal{S} \to T_{f(\mathbf{p})}\tilde{\mathcal{S}}$ is a linear map.

Proof. Equation 2.27 shows that $D_{\mathbf{p}}f$ is a linear map and the matrix of the linear map $D_{\mathbf{p}}f$ with respect to the basis $\{\sigma_u, \sigma_v\}$ of $T_{\mathbf{p}}f$ and the basis $\{\tilde{\sigma}_{\tilde{u}}, \tilde{\sigma}_{\tilde{v}}\}$ of $T_{f(\mathbf{p})}\tilde{\mathcal{S}}$ is the Jacobian matrix

$$\begin{pmatrix} \alpha_u & \alpha_v \\ \beta_u & \beta_v \end{pmatrix}$$

of the smooth map $(u, v) \mapsto (\alpha(u, v), \beta(u, v))$. □

Proposition 2.12.5.

(i) If $\mathcal{S}$ is a surface and $\mathbf{p} \in \mathcal{S}$, the derivative at $\mathbf{p}$ of the identity map $\mathcal{S} \to \mathcal{S}$ is the identity map $T_{\mathbf{p}}\mathcal{S} \to T_{\mathbf{p}}\mathcal{S}$.

(ii) If $\mathcal{S}_1, \mathcal{S}_2$ and $\mathcal{S}_3$ are surfaces and $f_1 : \mathcal{S}_1 \to \mathcal{S}_2$ and $f_2 : \mathcal{S}_2 \to \mathcal{S}_3$ are smooth maps, then for all $\mathbf{p} \in \mathcal{S}_1$,

$$D_{\mathbf{p}}(f_2 \circ f_1) = D_{f_1(\mathbf{p})}f_2 \circ D_{\mathbf{p}}f_1.$$

(iii) If $f : \mathcal{S}_1 \to \mathcal{S}_2$ is a diffeomorphism, then for all $\mathbf{p} \in \mathcal{S}_1$ the linear map $D_{\mathbf{p}}f : T_{\mathbf{p}}\mathcal{S}_1 \to T_{f(\mathbf{p})}\mathcal{S}_2$ is invertible.

Proof.

(i) From the definition of identity map,we can show that $D_{\mathbf{p}}f(\mathbf{w}) = \mathbf{w}$ for all $\mathbf{w} \in T_{\mathrm{p}}\mathcal{S}$.

(ii) let $\mathbf{w} \in T_{\mathrm{p}}\mathcal{S}_1$ be the tangent vector at $\mathbf{p}$ of a curve γ_1 on $\mathcal{S}_1$. Then, $\gamma_2 = f_1 \circ \gamma_1$ is a curve on $\mathcal{S}_2$ with tangent vector $D_{\mathbf{p}}f_1(\mathbf{w})$ at $f_1(\mathbf{p})$, so $\gamma_3 = f_2 \circ \gamma_2 = (f_2 \circ f_1) \circ \gamma_1$ is a curve on $\mathcal{S}_3$ with tangent vector $D_{f_1(\mathbf{p})}f_2\left(D_{\mathbf{p}}f_1(\mathbf{w})\right)$ at $f_2\left(f_1(\mathbf{p})\right)$. But the tangent vector of γ_3 at $\mathbf{p}$ is also $D_{\mathbf{p}}\left(f_2 \circ f_1\right)(\mathbf{w})$.

(iii) let $g : \mathcal{S}_2 \to \mathcal{S}_1$ be the inverse map of f, so that $g \circ f$ and $f \circ g$ are the identity maps $\mathcal{S}_1 \to \mathcal{S}_1$ and $\mathcal{S}_2 \to \mathcal{S}_2$, respectively. Parts (i) and (ii) show that $D_{f(\mathbf{p})}g$ is the inverse of the linear map $D_{\mathbf{p}}f$.

□

Proposition 2.12.6. Let $\mathcal{S}$ and $\tilde{\mathcal{S}}$ be surfaces and let $f : \mathcal{S} \to \tilde{\mathcal{S}}$ be a smooth map. Then, f is a local diffeomorphism if and only if, for all $\mathbf{p} \in \mathcal{S}$, the linear map $D_{\mathbf{p}}f : T_{\mathbf{p}}\mathcal{S} \to T_{f(\mathbf{p})}\tilde{\mathcal{S}}$ is invertible.

Proof. Suppose first that f is a local diffeomorphism and let $\mathbf{p} \in \mathcal{S}$. Then, there is an open subset $\mathcal{O}$ of $\mathcal{S}$ containing $\mathbf{p}$ such that $f(\mathcal{O})$ is an open subset of $\tilde{\mathcal{S}}$ and $f|_{\mathcal{O}} : \mathcal{O} \to f(\mathcal{O})$ is a diffeomorphism. By Proposition 2.12.5 (iii), $D_{\mathbf{p}}f$ is invertible (note that it is obvious that $T_{\mathbf{p}}\mathcal{S} = T_{\mathbf{p}}\mathcal{O}$). The proof of the 'if' part requires the inverse function theorem and will be given in Section 3.3 □

2.13 Normals and orientability

Definition 2.13.1. Let $\sigma : U \to \mathbb{R}^3$ be a patch of a surface $\mathcal{S}$ containing a point $\mathbf{p} \in \mathcal{S}$ then ,

$$\mathbf{N}_\sigma = \frac{\sigma_u \times \sigma_v}{\|\sigma_u \times \sigma_v\|} \tag{2.28}$$

is clearly a unit vector perpendicular to every linear combination of σ_u and σ_v.(Hence perpendicular to tangent plane $T_{\mathbf{p}}\mathcal{S}$ by Proposition 2.12.2) is called the standard unit normal of the surface patch σ at $\mathbf{p}$

If $\tilde{\sigma} : \tilde{U} \to \mathbb{R}^3$ is another surface patch in the atlas of $\mathcal{S}$ containing $\mathbf{p}$, we showed in the proof of Proposition 2.10.8 that

$$\tilde{\sigma}_{\tilde{u}} \times \tilde{\sigma}_{\tilde{v}} = \det(J(\Phi))\sigma_u \times \sigma_v$$

where $J(\Phi)$ is the Jacobian matrix of the transition map Φ from σ to $\tilde{\sigma}$. So the standard unit normal of $\tilde{\sigma}$ is

$$\mathbf{N}_{\tilde{\sigma}} = \frac{\tilde{\sigma}_{\tilde{u}} \times \tilde{\sigma}_{\tilde{v}}}{\|\tilde{\sigma}_{\tilde{u}} \times \tilde{\sigma}_{\tilde{v}}\|} = \pm\frac{\sigma_u \times \sigma_v}{\|\sigma_u \times \sigma_v\|} = \pm\mathbf{N}_\sigma$$

where the sign is that of the determinant of $J(\Phi)$.

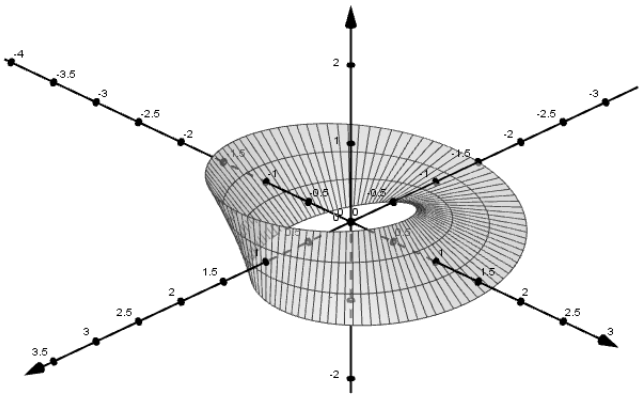
Definition 2.13.2. A surface $\mathcal{S}$ is orientable if there exists an atlas $\mathcal{A}$ for $\mathcal{S}$ with the property that, if Φ is the transition map between any two surface patches in $\mathcal{A}$, then $\det(J(\Phi)) > 0$ where Φ is defined.

Proposition 2.13.3. Let $\mathcal{S}$ be an orientable surface equipped with an atlas $\mathcal{A}$ as in Definition 2.13.2. Then, there is a smooth choice of unit normal at any point of $\mathcal{S}$: take the standard unit normal of any surface patch in $\mathcal{A}$.

Definition 2.13.4. An oriented surface is a surface $\mathcal{S}$ together with a smooth choice of unit normal $\mathbf{N}$ at each point, i.e., a smooth map $\mathbf{N} : \mathcal{S} \to \mathbb{R}^3$ such that, for all $\mathbf{p} \in \mathcal{S}, \mathbf{N}(\mathbf{p})$ is a unit vector perpendicular to $T_{\mathrm{p}}\mathcal{S}$. Any oriented surface is orientable.

From now onwards, *whenever we are dealing with an oriented surface $\mathcal{S}$, we shall only use surface patches for $\mathcal{S}$ whose standard unit normal is the same as the chosen normal of $\mathcal{S}$*

Example 2.13.5. The *Möbius band* is the surface obtained by rotating a straight line segment l around its midpoint $\mathbf{p}$ at the same time as $\mathbf{p}$ moves around a circle $\mathcal{C}$, in such a way that as $\mathbf{p}$ moves once around $\mathcal{C}, l$ makes a half-turn about $\mathbf{p}$. If we take $\mathcal{C}$ to be the circle $x^2 + y^2 = 1$ in the xy-plane, and l to be a segment of length 1 that is initially parallel to the z-axis with its midpoint $\mathbf{p}$ at $(1, 0, 0)$, then after $\mathbf{p}$ has rotated by an angle θ around the z-axis, l should have rotated by $\theta/2$ around $\mathbf{p}$ in the plane containing $\mathbf{p}$ and the z-axis. The point of l initially at $(1, 0, t)$ is then at the point



$$\sigma(t, \theta) = \left(\left(1 - t \sin\frac{\theta}{2}\right)\cos\theta, \left(1 - t \sin\frac{\theta}{2}\right)\sin\theta, t\cos\frac{\theta}{2}\right).$$

We take the domain of definition of σ to be

$$U = \left\{(t, \theta) \in \mathbb{R}^2 \mid -1/2 < t < 1/2, 0 < \theta < 2\pi\right\}$$

We can define a second patch $\tilde{\sigma}$ by the same formula as σ but with domain of definition $\tilde{U} = \{(t, \theta) \in \mathbb{R}^2 \mid -1/2 < t < 1/2, -\pi < \theta < \pi\}$. It can be checked that these two patches form an atlas for the Möbius band consisting of regular surface patches, making the Möbius band into a smooth surface $\mathcal{S}$.

We compute the standard unit normal $\mathbf{N}_\sigma$ at points on the median circle (where $t = 0$). At such points, we have

$$\sigma_t = \left(-\sin\frac{\theta}{2}\cos\theta, -\sin\frac{\theta}{2}\sin\theta, \cos\frac{\theta}{2}\right), \quad \sigma_\theta = (-\sin\theta, \cos\theta, 0)$$

so

$$\sigma_t \times \sigma_\theta = \left(-\cos\theta\cos\frac{\theta}{2}, -\sin\theta\cos\frac{\theta}{2}, -\sin\frac{\theta}{2} \right)$$

This is a unit vector, so it is equal to $\mathbf{N}_\sigma$.

If the Möbius band was orientable, there would be a well-defined unit normal N defined at every point of $\mathcal{S}$ and varying smoothly over $\mathcal{S}$. At a point $\sigma(0,\theta)$ on the median circle, we would have

$$\mathbf{N} = \lambda(\theta)\mathbf{N}_{\boldsymbol{\sigma}}$$

where $\lambda : (0, 2\pi) \to \mathbb{R}$ is smooth and $\lambda(\theta) = \pm 1$ for all θ. It follows that either $\lambda(\theta) = +1$ for all $\theta \in (0, 2\pi)$, or $\lambda(\theta) = -1$ for all $\theta \in (0, 2\pi)$. Replacing $\mathbf{N}$ by $-\mathbf{N}$ if necessary, we can assume that $\lambda = 1$. At the point $\sigma(0,0) = \sigma(0, 2\pi)$, we would have (since $\mathbf{N}$ is smooth)

$$\mathbf{N} = \lim_{\theta \downarrow 0} \mathbf{N}_\sigma = (-1, 0, 0) \text{ and also } \mathbf{N} = \lim_{\theta \uparrow 2\pi} \mathbf{N}_{\boldsymbol{\sigma}} = (1, 0, 0)$$

This contradiction shows that the Möbius band is not orientable.

If a surface $\mathcal{S}$ is oriented, it is possible to give a sign to the angle between two tangent vectors at a point of $\mathcal{S}$. Let $\mathbf{p} \in \mathcal{S}$ and let $\mathbf{N}$ be the chosen unit normal at $\mathbf{p}$. A rotation in the tangent plane $T_{\mathbf{p}}\mathcal{S}$ is said to be in the *positive sense*, or *anticlockwise*, if rotation in this sense of a right-handed screw held perpendicular to $T_{\mathrm{p}}\mathcal{S}$ would cause it to advance in the direction of $\mathbf{N}$. Put another way, the choice of $\mathbf{N}$ enables us to distinguish the two 'sides' of $T_{\mathrm{p}}\mathcal{S}$: the 'positive' side is the half-space into which $\mathbf{N}$ points. Then, if we view $T_{\mathrm{p}}\mathcal{S}$ from a point on the positive side, a positive rotation in $T_{\mathrm{p}}\mathcal{S}$ would be seen as anticlockwise in the usual sense.

If $\mathbf{v}$ and $\mathbf{w}$ are non-zero vectors in $T_{\mathbf{p}}\mathcal{S}$, the *oriented angle* (which we shall sometimes just call the angle) between $\mathbf{v}$ and $\mathbf{w}$ is the angle through which $\mathbf{v}$ must be rotated in the positive sense in order for the resulting vector to be a positive scalar multiple of $\mathbf{w}$. We shall denote this angle by $\widehat{\mathbf{vw}}$. Note that

$$\widehat{\mathbf{wv}} = -\widehat{\mathbf{vw}}$$

and that the sign of $\widehat{\mathbf{vw}}$ will change if we change the choice of unit normal to $T_{\mathbf{p}}\mathcal{S}$. Note also that $\widehat{\mathbf{vw}}$ is determined only up to the addition of an integer multiple of 2π.

Example 2.13.6. It is clear that at a point $\mathbf{p} \in S^2$, the tangent plane is perpendicular to $\mathbf{p}$. Since $\mathbf{p}$ is a unit vector, the two unit normals at $\mathbf{p}$ are $\mathbf{p}$, the 'outward' normal, and $-\mathbf{p}$, the 'inward' normal. There are two smooth choices of unit normal on S^2 : either always inwards or always outwards.

Module 3

Curvature of Surfaces

3.1 Level surfaces

Theorem 3.1.1. Let $\mathcal{S}$ be a subset of $\mathbb{R}^3$ with the following property: for each point $\mathbf{p} \in \mathcal{S}$, there is an open subset W of $\mathbb{R}^3$ containing $\mathbf{p}$ and a smooth function $f : W \to \mathbb{R}$ such that

(i) $\mathcal{S} \cap W = \{(x, y, z) \in W \mid f(x, y, z) = 0\}$;

(ii) The gradient $\boldsymbol{\nabla} f = (f_x, f_y, f_z)$ of f does not vanish at $\mathbf{p}$. Then, $\mathcal{S}$ is a smooth surface.

We postpone the proof to Section 3.3.

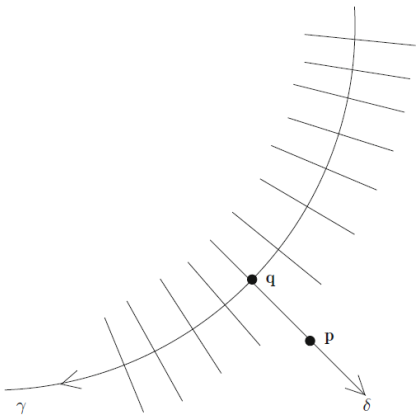
Example 3.1.2. For the unit sphere S^2, we can take $W = \mathbb{R}^3$ and use the single function $f(x, y, z) = x^2 + y^2 + z^2 - 1$. Then, $\boldsymbol{\nabla} f = (2x, 2y, 2z)$ so $\|\boldsymbol{\nabla} f\| = 2$ at all points of S^2. In particular, ∇f is non-zero everywhere on S^2. Hence, Theorem 3.1.1 tells us that S^2 is a smooth surface.

Example 3.1.3. For the circular cone of 2.9.6, $f(x, y, z) = x^2+y^2-z^2$. Hence, $\boldsymbol{\nabla} f = (2x, 2y, -2z)$, and this vanishes only at the vertex $(0, 0, 0)$. Theorem 3.1.1 applies with $W = \{(x, y, z) \in \mathbb{R}^3 \mid z \neq 0\}$, so the circular cone with the vertex removed is a smooth surface.

Exercise 3.1.4. Show that $x^2 + y^2 + z^4 = 1$ is a smooth surfaces:
Solution: $f_x = 2x, f_y = 2y$ and $f_z = 4z^3$ vanish simultaneously only when $x = y = z = 0$, but this does not satisfy $x^2 + y^2 + z^4 = 1$. So by Theorem 3.1.1 this is a smooth surface.

3.2 Ruled surfaces and surfaces of revolution

Example 3.2.1. A *ruled surface* is a surface that is a union of straight lines, called the *rulings* (or sometimes the *generators*) of the surface.



Suppose that $\mathcal{C}$ is a curve in $\mathbb{R}^3$ that meets each of these lines. Any point $\mathbf{p}$ of the surface lies on one of the given straight lines, which intersects $\mathcal{C}$ at $\mathbf{q}$, say. If $\gamma : (\alpha, \beta) \to \mathbb{R}^3$ is a parametrization of $\mathcal{C}$ with $\gamma(u) = \mathbf{q}$, and if $\delta(u)$ is a non-zero vector in the direction of the line passing through $\gamma(u)$, then

$$\mathbf{p} = \gamma(u) + v\delta(u)$$

for some scalar v. Denoting the right-hand side by $\sigma(u, v)$, it is clear that $\sigma : U \to \mathbb{R}^3$ is a smooth map, where $U = \{(u, v) \in \mathbb{R}^2 \mid \alpha < u < \beta\}$. Moreover, denoting d/du by a dot,

$$\sigma_u = \dot{\gamma} + v\dot{\delta}, \quad \sigma_v = \delta$$

Thus, σ is regular if $\dot{\gamma} + v\dot{\delta}$ and δ are linearly independent. This will be true, for example, if $\dot{\gamma}$ and δ are linearly independent and v is sufficiently small. Thus, to get a surface, the curve $\mathcal{C}$ must never be tangent to the rulings.

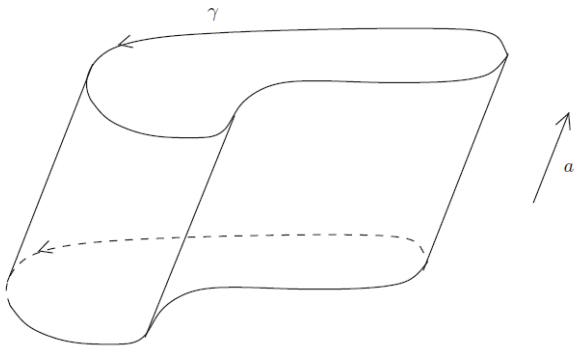
An important special case is that in which the rulings are all parallel to each other; the ruled surface $\mathcal{S}$ is then called a *generalized cylinder*. In the above notation, we can take δ to be a constant unit vector, say $\mathbf{a}$, parallel to the rulings, and the parametrization becomes

$$\sigma(u, v) = \gamma(u) + v\mathbf{a}$$

Since

$$\sigma(u, v) = \sigma(u', v') \iff \gamma(u) - \gamma(u') = (v' - v)\,\mathbf{a},$$

for σ to be a injective (and hence a surface patch), no straight line parallel to $\mathbf{a}$ should meet γ in more than one point. Finally, $\sigma_u = \dot{\gamma}, \sigma_v = \mathbf{a}$, so σ is regular if and only if γ is never tangent to the rulings.



The parametrization is simplest when γ lies in a plane perpendicular to $\mathbf{a}$. The regularity condition is then clearly satisfied provided $\dot{\gamma}$ is never zero, i.e., provided γ is regular. We might as well take the plane to be the xy-plane and $\mathbf{a} = (0, 0, 1)$ to be parallel to the z-axis. Then, $\gamma(u) = (f(u), g(u), 0)$ for some smooth functions f and g, and the parametrization becomes

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

For example, starting with a circle, we get a circular cylinder. Taking the circle to have centre the origin, radius 1 and to lie in the xy-plane, it can be parametrized by

$$\gamma(u) = (\cos u, \sin u, 0)$$

defined for $0 < u < 2\pi$ and $-\pi < u < \pi$, say.

The second special case we shall consider is that in which the rulings all pass through a certain fixed point, say $\mathbf{v}$; then $\mathcal{S}$ is called a *generalized cone with vertex* $\mathbf{v}$.

We can take $\delta(u) = \gamma(u) - \mathbf{v}$, giving

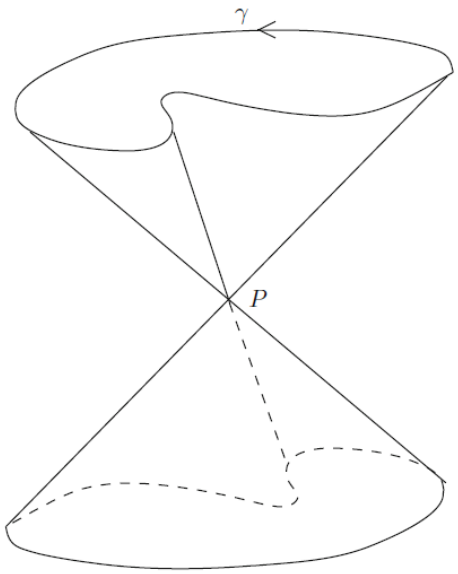
$$\sigma(u, v) = (1 + v)\gamma(u) - v\mathbf{v}.$$

Now,

$$\sigma(u, v) = \sigma(u', v') \iff (1 + v)\gamma(u) - (1 + v')\gamma(u') + (v' - v)\mathbf{v} = \mathbf{0}$$

since $(1+v)-(1+v')+(v'-v) = 0$, the equation on the right-hand side means that the points $\mathbf{v}, \gamma(u)$ and $\gamma(u')$ are collinear. So, for σ to be a surface patch, no straight line passing through $\mathbf{v}$ should pass through more than one point of

γ (in particular, γ should not pass through $\mathbf{v}$). Finally, we have $\sigma_u = (1+v)\dot{\gamma}$, $\sigma_v = \gamma - \mathrm{v}$, so σ is regular provided $v \neq -1$, i.e., the vertex of the cone is omitted (Example 2.9.6), and none of the straight lines forming the cone is tangent to γ.



The parametrization is simplest when γ lies in a plane. If this plane contains $\mathbf{v}$, the cone is simply part of that plane. Otherwise, we can take $\mathbf{v}$ to be the origin and the plane to be $z = 1$. Then, $\gamma(u) = (f(u), g(u), 1)$ for some smooth functions f and g, and the parametrization takes the form

$$\sigma(u, v) = v(f(u), g(u), 1)$$

after making the reparametrization $v \mapsto v - 1$.

Example 3.2.2. A *surface of revolution* is the surface obtained by rotating a plane curve, called the *profile curve*, around a straight line in the plane. The circles obtained by rotating a fixed point on the profile curve around the axis of rotation are called the *parallels* of the surface, and the curves on the surface obtained by rotating the profile curve through a fixed angle are called its *meridians*.

Let us take the axis of rotation to be the z-axis and the plane to be the xz-plane. Any point $\mathbf{p}$ of the surface is obtained by rotating some point $\mathbf{q}$ of the profile curve through an angle v (say) around the z-axis. If

$$\gamma(u) = (f(u), 0, g(u))$$

is a parametrization of the profile curve containing $\mathbf{q}, \mathbf{p}$ is of the form

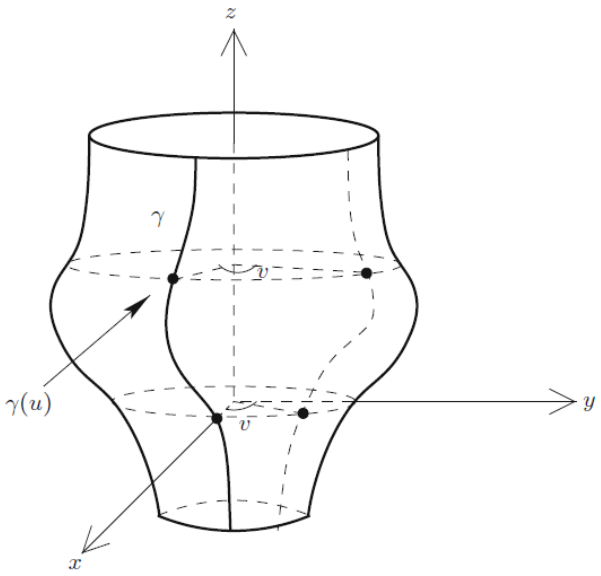
$$\sigma(u, v) = (f(u)\cos v, f(u)\sin v, g(u)).$$

To check regularity, we compute (with a dot denoting d/du):

$$\sigma_u = (\dot{f}\cos v, \dot{f}\sin v, \dot{g}), \quad \sigma_v = (-f\sin v, f\cos v, 0)$$

$$\therefore \quad \sigma_u \times \sigma_v = (f\dot{g}\cos v, -f\dot{g}\sin v, f\dot{f})$$

$$\therefore \quad \|\sigma_u \times \sigma_v\|^2 = f^2\left(\dot{f}^2 + \dot{g}^2\right)$$



Thus, $\sigma_u \times \sigma_v$ will be non-vanishing if $f(u)$ is never zero, i.e., if γ does not intersect the z-axis, and if $\dot{f}$ and $\dot{g}$ are never zero simultaneously, i.e., if γ is regular. In this case, we might as well assume that $f(u) > 0$, so that $f(u)$ is the distance of $\sigma(u, v)$ from the axis of rotation. Then, σ is injective provided that γ does not self-intersect and the angle of rotation v is restricted to lie in an open interval of length $\leq 2\pi$. Under these conditions, surface patches of the form σ give the surface of revolution the structure of a surface.

3.3 Applications of the inverse function theorem

In this section we give the proofs of Propositions 2.10.7 and 2.12.6 and Theorem 3.1.1.

Suppose first that $f : U \to \mathbb{R}^n$ is a smooth map, where U is an open subset of $\mathbb{R}^m$. If we write $(\tilde{u}_1, \ldots, \tilde{u}_n) = f(u_1, \ldots, u_m)$, the J*acobian matrix* of f is

$$J(f) = \begin{pmatrix} \frac{\partial \tilde{u}_1}{\partial u_1} & \frac{\partial \tilde{u}_1}{\partial u_2} & \cdots & \frac{\partial \tilde{u}_1}{\partial u_m} \\ \frac{\partial \tilde{u}_2}{\partial u_1} & \frac{\partial \tilde{u}_2}{\partial u_2} & \cdots & \frac{\partial \tilde{u}_2}{\partial u_m} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial \tilde{u}_n}{\partial u_1} & \frac{\partial \tilde{u}_n}{\partial u_2} & \cdots & \frac{\partial \tilde{u}_n}{\partial u_m} \end{pmatrix}$$

Theorem 3.3.1. (Inverse Function Theorem) Let $f : U \to \mathbb{R}^n$ be a smooth map defined on an open subset U of $\mathbb{R}^n (n \geq 1)$. Assume that, at some point $x_0 \in U$, the Jacobian matrix $J(f)$ is invertible. Then, there is an open subset V of $\mathbb{R}^n$ and a smooth map $g : V \to \mathbb{R}^n$ such that

(i) $y_0 = f(x_0) \in V$

(ii) $g(y_0) = x_0$

(iii) $g(V) \subseteq U$

(iv) $g(V)$ is an open subset of $\mathbb{R}^n$

(v) $f(g(y)) = y$ for all $y \in V$

In particular, $g : V \to g(V)$ and $f : g(V) \to V$ are inverse bijections.

Thus, the inverse function theorem says that, if $J(f)$ is invertible at some point, then f is bijective near that point and its inverse map is smooth.

As our first application of Theorem 3.3.1, we complete the proof of Proposition 2.12.6. Suppose then that $f : \mathcal{S} \to \tilde{\mathcal{S}}$ is a smooth map between surfaces $\mathcal{S}$ and $\tilde{\mathcal{S}}$, let $\mathbf{p} \in \mathcal{S}$ and assume that the linear map $D_{\mathbf{p}} f : T_{\mathbf{p}}\mathcal{S} \to T_{f(\mathbf{p})}\tilde{\mathcal{S}}$ is invertible. Let $\sigma : U \to \mathbb{R}^3$ be a surface patch of $\mathcal{S}$ containing $\mathbf{p}$, say $\sigma(u_0, v_0) = \mathbf{p}$, and let $\tilde{\sigma} : \tilde{U} \to \mathbb{R}^3$ be a surface patch of $\tilde{\mathcal{S}}$ containing $f(\mathbf{p})$.By shrinking U if necessary, we can assume that f maps $\sigma(U)$ into $\tilde{\sigma}(\tilde{U})$. Since f is smooth, there are smooth functions $\alpha : U \to \mathbb{R}$ and $\beta : U \to \mathbb{R}$ such that

$$f(\sigma(u, v)) = \tilde{\sigma}(\alpha(u, v), \beta(u, v)).$$

From the remarks following Proposition 2.12.4, the matrix of $D_{\mathbf{p}}f$ with respect to the bases $\{\sigma_u, \sigma_v\}$ of $T_{\mathbf{p}}\mathcal{S}$ and $\{\tilde{\sigma}_{\tilde{u}}, \tilde{\sigma}_{\tilde{v}}\}$ of $T_{f(\mathbf{p})}\tilde{\mathcal{S}}$ is the Jacobian matrix

$$\begin{pmatrix} \alpha_u & \alpha_v \\ \beta_u & \beta_v \end{pmatrix}.$$

Since $D_{\mathrm{p}}f$ is invertible, so is this matrix. By the inverse function theorem, the smooth map $U \to \mathbb{R}^2$ given by $(u, v) \mapsto (\alpha(u, v), \beta(u, v))$ is a diffeomorphism from an open subset V (say) of U containing (u_0, v_0) to an open subset $\tilde{V}$ (say) of $\tilde{U}$. Then $\mathcal{O} = \sigma(V)$ and $\tilde{\mathcal{O}} = \tilde{\sigma}(\tilde{V})$ are open subsets of $\mathcal{S}$ and $\tilde{\mathcal{S}}$, respectively, and f is a diffeomorphism from $\mathcal{O}$ to $\tilde{\mathcal{O}}$. This proves that f is a local diffeomorphism.

We now give the proof of Proposition 2.10.7. We want to show that, if $\sigma : U \to \mathbb{R}^3$ and $\tilde{\sigma} : \tilde{U} \to \mathbb{R}^3$ are two regular patches in the atlas of a surface $\mathcal{S}$, the transition map from σ to $\tilde{\sigma}$ is smooth where it is defined.

Suppose that a point $\mathbf{p}$ lies in both patches, say $\sigma(u_0, v_0) = \tilde{\sigma}(\tilde{u}_0, \tilde{v}_0) = \mathbf{p}$. Write

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

Since σ_u and σ_v are linearly independent, the Jacobian matrix

$$\begin{pmatrix} f_u & f_v \\ g_u & g_v \\ h_u & h_v \end{pmatrix}$$

of σ has rank 2 everywhere. Hence, at least one of its three 2×2 submatrices is invertible at each point. Suppose that the submatrix

$$\begin{pmatrix} f_u & f_v \\ g_u & g_v \end{pmatrix}$$

is invertible at $\mathbf{p}$. (The proof is similar in the other two cases.) By the inverse function theorem applied to the map $F : U \to \mathbb{R}^2$ given by

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

there is an open subset V of $\mathbb{R}^2$ containing $F(u_0, v_0)$ and an open subset W of U containing (u_0, v_0) such that $F : W \to V$ is bijective with a smooth inverse $F^{-1} : V \to W$. Since $\sigma : W \to \sigma(W)$ is bijective, the projection $\pi : \sigma(W) \to V$ given by $\pi(x, y, z) = (x, y)$ is also bijective, since $\pi = F \circ \sigma^{-1}$ on $\sigma(W)$. It follows that $\tilde{W} = \tilde{\sigma}^{-1}(\sigma(W))$ is an open subset of $\tilde{U}$ and that

$$\sigma^{-1} \circ \tilde{\sigma} = F^{-1} \circ \tilde{F}$$

on $\tilde{W}$, where $\tilde{F} = \pi \circ \tilde{\boldsymbol{\sigma}}$. Since F^{-1} and $\tilde{F}$ are smooth on $\tilde{W}$, so is the transition map $\sigma^{-1} \circ \tilde{\sigma}$. Since $\sigma^{-1} \circ \tilde{\sigma}$ is smooth on an open set containing any point (u_0, v_0) where it is defined, it is smooth.

Finally, we give the proof of Theorem 3.1.1. Let $\mathbf{p}, W$ and f be as in the statement of the theorem, and suppose that $\mathbf{p} = (x_0, y_0, z_0)$ and that $f_z \neq 0$ at $\mathbf{p}$. (The proof is similar in the other two cases.) Consider the map $F : W \to \mathbb{R}^3$ defined by

$$F(x, y, z) = (x, y, f(x, y, z)).$$

The Jacobian matrix of F is

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ f_x & f_y & f_z \end{pmatrix}$$

and is clearly invertible at $\mathbf{p}$ since $f_z \neq 0$. By the inverse function theorem, there is an open subset V of $\mathbb{R}^3$ containing $F(x_0, y_0, z_0) = (x_0, y_0, 0)$ and a smooth map $G : V \to W$ such that $\tilde{W} = G(V)$ is open and $F : \tilde{W} \to V$ and $G : V \to \tilde{W}$ are inverse bijections.

Since V is open, there are open subsets U_1 of $\mathbb{R}^2$ containing (x_0, y_0) and U_2 of $\mathbb{R}$ containing 0 such that V contains the open set $U_1 \times U_2$ of all points (x, y, w) with $(x, y) \in U_1$ and $w \in U_2$. Hence, we might as well assume that $V = U_1 \times U_2$. The fact that F and G are inverse bijections means that

$$G(x, y, w) = (x, y, g(x, y, w))$$

for some smooth map $g : U_1 \times U_2 \to \mathbb{R}$, and

$$f(x, y, g(x, y, w)) = w$$

for all $(x, y) \in U_1, w \in U_2$.

Define $\sigma : U_1 \to \mathbb{R}^3$ by

$$\sigma(x, y) = (x, y, g(x, y, 0)).$$

Then σ is a homeomorphism from U_1 to $\mathcal{S} \cap \tilde{W}$ (whose inverse is the restriction to $\mathcal{S} \cap \tilde{W}$ of the projection $\pi(x, y, z) = (x, y)$). It is obvious that σ is smooth, and it is regular because

$$\sigma_x \times \sigma_y = (-g_x, -g_y, 1)$$

is nowhere zero. So σ is a regular surface patch on $\mathcal{S}$ containing the given point $\mathbf{p}$. Since $\mathbf{p}$ was an arbitrary point of $\mathcal{S}$, we have constructed an atlas for $\mathcal{S}$ making it into a (smooth) surface.

3.4 The first fundamental form

Definition 3.4.1. Let V denotes a vector space of finite dimension n over $\mathbb{R}$. A map $V \times V \to \mathbb{R}$, denoted $(\mathbf{v}, \mathbf{w}) \mapsto \langle \mathbf{v}, \mathbf{w} \rangle$, is called a bilinear form if, for all $\lambda_1, \lambda_2 \in \mathbb{R}, \mathbf{v}_1, \mathbf{v}_2, \mathbf{w} \in V$, we have

$$\langle \lambda_1 \mathbf{v}_1 + \lambda_2 \mathbf{v}_2, \mathbf{w} \rangle = \lambda_1 \langle \mathbf{w}, \mathbf{v}_1 \rangle + \lambda_2 \langle \mathbf{w}, \mathbf{v}_2 \rangle$$
$$\langle \mathbf{w}, \lambda_1 \mathbf{v}_1 + \lambda_2 \mathbf{v}_2 \rangle = \lambda_1 \langle \mathbf{w}, \mathbf{v}_1 \rangle + \lambda_2 \langle \mathbf{w}, \mathbf{v}_2 \rangle .$$

Thus, $\langle \mathbf{v}, \mathbf{w} \rangle$ is a linear function of $\mathbf{v}$ for each fixed $\mathbf{w}$, and a linear function of $\mathbf{w}$ for each fixed $\mathbf{v}$.

A bilinear form $\langle , \rangle$ is called symmetric if

$$\langle \mathbf{v}, \mathbf{w} \rangle = \langle \mathbf{w}, \mathbf{v} \rangle \quad \text{for all } \mathbf{v}, \mathbf{w} \in V.$$

An inner product on V is a symmetric bilinear form $\langle , \rangle$ on V such that $\langle \mathbf{v}, \mathbf{v} \rangle > 0$ for all non-zero $\mathbf{v} \in V$.

If V has an inner product $\langle , \rangle$, a linear map $L : V \to V$ is called self-adjoint if

$$\langle L(\mathbf{v}), \mathbf{w} \rangle = \langle \mathbf{v}, L(\mathbf{w}) \rangle \text{ for all } \mathbf{v}, \mathbf{w} \in V.$$

Result 3.4.2. Let $L : V \to V$ be a self-adjoint linear map. Then, V has a basis $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ consisting of eigenvectors of L. Moreover, if $\mathbf{v}_i$ and $\mathbf{v}_j$ are eigenvectors corresponding to distinct eigenvalues, then $\langle \mathbf{v}_i, \mathbf{v}_j \rangle = 0$.

Definition 3.4.3. Let $\mathbf{p}$ be a point of a surface $\mathcal{S}$. The first fundamental form of $\mathcal{S}$ at $\mathbf{p}$ associates to tangent vectors $\mathbf{v}, \mathbf{w} \in T_{\mathrm{p}}\mathcal{S}$ the scalar

$$\langle \mathbf{v}, \mathbf{w} \rangle_{\mathbf{p},\mathcal{S}} = \mathbf{v} \cdot \mathbf{w}.$$

Thus, $\langle \mathbf{v}, \mathbf{w} \rangle_{\mathbf{p},\mathcal{S}}$ is just the dot product, but restricted to tangent vectors to $\mathcal{S}$ at $\mathbf{p}$.

The first fundamental form $\langle , \rangle$ is an example of an inner product.

Suppose that $\sigma(u, v)$ is a surface patch of $\mathcal{S}$. Then, any tangent vector to $\mathcal{S}$ at a point $\mathbf{p}$ in the image of σ can be expressed uniquely as a linear combination of σ_u and σ_v. Define maps $du : T_{\mathbf{p}}\mathcal{S} \to \mathbb{R}$ and $dv : T_{\mathbf{p}}\mathcal{S} \to \mathbb{R}$ by

$$du(\mathbf{v}) = \lambda, \quad dv(\mathbf{v}) = \mu \quad \text{if } \mathbf{v} = \lambda\sigma_u + \mu\sigma_v$$

for some $\lambda, \mu \in \mathbb{R}$. It is easy to see that du and dv are *linear maps*. Then, using the fact that $\langle , \rangle$ is a symmetric bilinear form, we have

$$\langle \mathbf{v}, \mathbf{v} \rangle = \lambda^2 \langle \sigma_u, \sigma_u \rangle + 2\lambda\mu \langle \sigma_u, \sigma_v \rangle + \mu^2 \langle \sigma_v, \sigma_v \rangle .$$

Writing

$$E = \|\sigma_u\|^2 , \quad F = \sigma_u \cdot \sigma_v, \quad G = \|\sigma_v\|^2 ,$$

this becomes

$$\langle \mathbf{v}, \mathbf{v} \rangle = E\lambda^2 + 2F\lambda\mu + G\mu^2 = Edu(\mathbf{v})^2 + 2Fdu(\mathbf{v})dv(\mathbf{v}) + Gdv(\mathbf{v})^2.$$

Traditionally, the expression

$$Edu^2 + 2Fdudv + Gdv^2$$

is called the first fundamental form of the surface patch $\sigma(u, v)$. Note that the coefficients E, F, G and the linear maps du, dv depend on the choice of surface patch for $\mathcal{S}$, but the first fundamental form itself depends only on $\mathcal{S}$ and $\mathbf{p}$. If γ is a curve lying in the image of a surface patch σ, we have

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

for some smooth functions $u(t)$ and $v(t)$. Then, denoting d/dt by a dot, we have $\dot{\gamma} = \dot{u}\sigma_u + \dot{v}\sigma_v$ by the chain rule, so

$$\langle \dot{\gamma}, \dot{\gamma} \rangle = E\dot{u}^2 + 2F\dot{u}\dot{v} + G\dot{v}^2,$$

and the length of γ is given by

$$\int \left(E\dot{u}^2 + 2F\dot{u}\dot{v} + G\dot{v}^2\right)^{1/2} dt.$$

Example 3.4.4. For the plane

$$\sigma(u, v) = \mathbf{a} + u\mathbf{p} + v\mathbf{q}$$

with $\mathbf{p}$ and $\mathbf{q}$ being perpendicular unit vectors, we have $\sigma_u = \mathbf{p}, \sigma_v = \mathbf{q}$, so $E = \|\sigma_u\|^2 = \|\mathbf{p}\|^2 = 1, F = \sigma_u \cdot \sigma_v = \mathbf{p} \cdot \mathbf{q} = 0$ $G = \|\sigma_v\|^2 = \|\mathbf{q}\|^2 = 1$, and the first fundamental form is simply

$$du^2 + dv^2.$$

Example 3.4.5. Consider a surface of revolution (Example 3.2.2)

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

We can assume that $f(u) > 0$ for all values of u and that the profile curve $u \mapsto (f(u), 0, g(u))$ is unit-speed, i.e., $\dot{f}^2 + \dot{g}^2 = 1$ (a dot denoting d/du). Then:

$$\sigma_u = (\dot{f}\cos v, \dot{f}\sin v, \dot{g}), \quad \sigma_v = (-f\sin v, f\cos v, 0),$$

$$\therefore \quad E = \|\sigma_u\|^2 = \dot{f}^2 + \dot{g}^2 = 1, F = \sigma_u \cdot \sigma_v = 0, G = \|\sigma_v\|^2 = f^2.$$

So the first fundamental form is

$$du^2 + f(u)^2 dv^2.$$

A special case is the unit sphere S^2 in latitude-longitude coordinates. We take $u = \theta, v = \varphi, f(\theta) = \cos\theta, g(\theta) = \sin\theta$, giving the first fundamental form of S^2 as

$$d\theta^2 + \cos^2\theta d\varphi^2.$$

Example 3.4.6. We consider a generalized cylinder

$$\sigma(u,v) = \gamma(u) + v\mathbf{a}$$

defined in Example 3.2.1. We can assume that γ is unit-speed, $\mathbf{a}$ is a unit vector, and γ is contained in a plane perpendicular to $\mathbf{a}$. Then, denoting d/du by a dot, $\sigma_u = \dot{\gamma}, \sigma_v = \mathbf{a}$, so $E = \|\sigma_u\|^2 = \|\dot{\gamma}\|^2 = 1$, $F = \sigma_u \cdot \sigma_v = \dot{\gamma} \cdot \mathbf{a} = 0, G = \|\sigma_v\|^2 = \|\mathbf{a}\|^2 = 1$, and the first fundamental form of σ is

$$du^2 + dv^2.$$

Note that this is the *same* as the first fundamental form of the plane (Example 3.4.4).

Example 3.4.7. We consider a generalized cone

$$\boldsymbol{\sigma}(u,v) = (1+v)\gamma(u) - v\mathbf{v}$$

(Example 3.2.1). Before computing its first fundamental form, we make some simplifications to σ.

First, translating the surface by $\mathbf{v}$ (which does not change its first fundamental form), we get the surface patch $\sigma_1 = \sigma - \mathbf{v} = (1+v)(\gamma - \mathbf{v})$, so if we replace γ by $\gamma_1 = \gamma - \mathbf{v}$ we get $\sigma_1 = (1+v)\gamma_1$. This means that we might as well assume that $\mathbf{v} = \mathbf{0}$ to begin with. Next, we saw in Example 3.2.1 that for σ to be a regular surface patch, γ must not pass through the origin, so we can define

a new curve $\tilde{\gamma}$ by $\tilde{\gamma}(u) = \gamma(u)/\|\gamma(u)\|$. Setting $\tilde{u} = u, \tilde{v} = (1+v)/\|\gamma(u)\|$, we get a reparametrization $\tilde{\sigma}(\tilde{u}, \tilde{v}) = \tilde{v}\tilde{\gamma}(\tilde{u})$ of σ with $\|\tilde{\gamma}\| = 1$. We can therefore assume to begin with that $\sigma(u, v) = v\gamma(u)$ with $\|\gamma(u)\| = 1$ for all values of u (geometrically, this means that we can replace γ by the intersection of the cone with S^2). Finally, reparametrizing again, we can assume that γ is unit-speed, for we saw in Example 3.2.1 that for σ to be regular, γ must be regular.

With these assumptions, and with a dot denoting d/du, we have $\sigma_u = v\dot{\gamma}$, $\sigma_v = \gamma$, so $E = \|v\dot{\gamma}\|^2 = v^2\|\dot{\gamma}\|^2 = v^2, F = v\dot{\gamma} \cdot \gamma = 0$ (since $\|\gamma\| = 1$), $G = \|\gamma\|^2 = 1$, and the first fundamental form is

$$v^2 du^2 + dv^2.$$

Note that, as for the generalized cylinder in Example 3.4.6 , the first fundamental form of the generalized cone does not depend on the curve γ.

Exercise 3.4.8. Calculate the first fundamental forms of the following surfaces:

(i) $\sigma(u, v) = (\sinh u \sinh v, \sinh u \cosh v, \sinh u)$.

(ii) $\sigma(u, v) = (u - v, u + v, u^2 + v^2)$.

(iii) $\sigma(u, v) = (\cosh u, \sinh u, v)$.

(iv) $\sigma(u, v) = (u, v, u^2 + v^2)$.

Solution:

(i) $\boldsymbol{\sigma}_u = (\cosh u \sinh v, \cosh u \cosh v, \cosh u)$, $\boldsymbol{\sigma}_v = (\sinh u \cosh v, \sinh u \sinh v, 0)$,
$\|\boldsymbol{\sigma}_u\|^2 = 2\cosh^2 u \cosh^2 v$,
$\boldsymbol{\sigma}_u \cdot \boldsymbol{\sigma}_v = 2\sinh u \cosh u \sinh v \cosh v$, $\|\sigma_v\|^2 = \sinh^2 u \cosh 2v$, and the first fundamental form is
$2\cosh^2 u \cosh^2 v du^2 + \sinh 2u \sinh 2v du dv + \sinh^2 u \cosh 2v dv^2$.

(ii) $(2 + 4u^2)\, du^2 + 8uv du dv + (2 + 4v^2)\, dv^2$.

(iii) $\left(\cosh^2 u + \sinh^2 u\right) du^2 + dv^2$.

(iv) $(1 + 4u^2)\, du^2 + 8uv du dv + (1 + 4v^2)\, dv^2$.

3.5 Area of a surface patch

Definition 3.5.1. The area $\mathcal{A}_\sigma(R)$ of the part $\sigma(R)$ of a surface patch $\sigma : U \to \mathbb{R}^3$ corresponding to a region $R \subseteq U$ is

$$\mathcal{A}_\sigma(R) = \int_R \|\sigma_u \times \sigma_v\| \, dudv.$$

The quantity $\|\sigma_u \times \sigma_v\|$ that appears in the definition of area is easily computed in terms of the first fundamental form $Edu^2 + 2Fdudv + Gdv^2$ of σ.

Proposition 3.5.2. $||\sigma_u \times \sigma_v\| = (EG - F^2)^{1/2}$.

Proof. We use a result from vector algebra: if $\mathbf{a}, \mathbf{b}, \mathbf{c}$ and $\mathbf{d}$ are vectors in $\mathbb{R}^3$,

$$(\mathbf{a} \times \mathbf{b}) \cdot (\mathbf{c} \times \mathbf{d}) = (\mathbf{a} \cdot \mathbf{c})(\mathbf{b} \cdot \mathbf{d}) - (\mathbf{a} \cdot \mathbf{d})(\mathbf{b} \cdot \mathbf{c}).$$

Applying this to $\|\sigma_u \times \sigma_v\|^2 = (\sigma_u \times \sigma_v) \cdot (\sigma_u \times \sigma_v)$, we get

$$||\sigma_u \times \sigma_v\|^2 = (\sigma_u \cdot \sigma_u)(\sigma_v \cdot \sigma_v) - (\sigma_u \cdot \sigma_v)^2 = EG - F^2.$$

□

Proposition 3.5.3. The area of a surface patch is unchanged by reparametrization.

Proof. Let $\sigma : U \to \mathbb{R}^3$ be a surface patch and let $\tilde{\sigma} : \tilde{U} \to \mathbb{R}^3$ be a reparametrization of σ, with reparametrization map $\Phi : \tilde{U} \to U$. Thus, if $\Phi(\tilde{u}, \tilde{v}) = (u, v)$, we have

$$\tilde{\sigma}(\tilde{u}, \tilde{v}) = \sigma(u, v).$$

Let $\tilde{R} \subseteq \tilde{U}$ be a region, and let $R = \Phi(\tilde{R}) \subseteq U$. We have to prove that

$$\int_R \|\sigma_u \times \sigma_v\| \, dudv = \int_{\tilde{R}} \|\tilde{\sigma}_{\tilde{u}} \times \tilde{\sigma}_{\tilde{v}}\| \, d\tilde{u}d\tilde{v}.$$

We showed in the proof of Proposition 2.10.8 that

$$\tilde{\sigma}_{\tilde{u}} \times \tilde{\sigma}_{\tilde{v}} = \det(J(\Phi))\sigma_u \times \sigma_v,$$

where $J(\Phi)$ is the Jacobian matrix of Φ. Hence,

$$\int_{\tilde{R}} \|\tilde{\sigma}_{\tilde{u}} \times \tilde{\sigma}_{\tilde{v}}\| \, d\tilde{u}d\tilde{v} = \int_{\tilde{R}} |\det(J(\Phi))| \, \|\sigma_u \times \sigma_v\| \, d\tilde{u}d\tilde{v}.$$

By the change of variables formula for double integrals, the right-hand side of this equation is exactly

$$\int_R \|\sigma_u \times \sigma_v\| \, dudv.$$

□

3.6 The second fundamental form

Suppose then that σ is a surface patch in $\mathbb{R}^3$ with standard unit normal $\mathbf{N}$. Let

$$L = \sigma_{uu} \cdot \mathbf{N}, \quad M = \sigma_{uv} \cdot \mathbf{N}, \quad N = \sigma_{vv} \cdot \mathbf{N}. \tag{3.1}$$

then the expression

$$Ldu^2 + 2Mdudv + Ndv^2$$

is called the *second fundamental form* of the surface patch σ. Using this we define a symmetric bilinear form on the tangent plane by

$$\langle\langle \mathbf{v}, \mathbf{w} \rangle\rangle = Ldu(\mathbf{v})du(\mathbf{w}) + M(du(\mathbf{v})dv(\mathbf{w}) + du(\mathbf{w})dv(\mathbf{v})) + Ndv(\mathbf{v})dv(\mathbf{w}).$$

Example 3.6.1. Consider the plane

$$\sigma(u, v) = \mathbf{a} + u\mathbf{p} + v\mathbf{q}$$

(see Example 2.9.3). Since $\sigma_u = \mathbf{p}$ and $\sigma_v = \mathbf{q}$ are constant vectors, we have $\sigma_{uu} = \sigma_{uv} = \sigma_{vv} = 0$. Hence, the second fundamental form of a plane is zero.

Example 3.6.2. Consider a surface of revolution (Example 3.2.2)

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

we assume as usual that $f(u) > 0$ for all values of u and that the profile curve $u \mapsto (f(u), 0, g(u))$ is unit-speed, i.e., $\dot{f}^2 + \dot{g}^2 = 1$ (a dot denoting d/du). Then:

$$\begin{aligned}
&\sigma_u = (\dot{f}\cos v, \dot{f}\sin v, \dot{g}), \quad \sigma_v = (-f\sin v, f\cos v, 0),\\
&\therefore \quad \sigma_u \times \sigma_v = (-f\dot{g}\cos v, -f\dot{g}\sin v, f\dot{f}),\\
&\therefore \quad \|\sigma_u \times \sigma_v\| = f \quad \Big(\text{since } \dot{f}^2 + \dot{g}^2 = 1\Big),\\
&\therefore \quad \mathrm{N} = \frac{\sigma_u \times \sigma_v}{\|\sigma_u \times \sigma_v\|} = (-\dot{g}\cos v, -\dot{g}\sin v, \dot{f}),\\
&\sigma_{uu} = (\ddot{f}\cos v, \ddot{f}\sin v, \ddot{g})\\
&\sigma_{uv} = (-\dot{f}\sin v, \dot{f}\cos v, 0)\\
&\sigma_{vv} = (-f\cos v, -f\sin v, 0),
\end{aligned}$$

$$\therefore \quad L = \sigma_{uu} \cdot \mathbf{N} = \dot{f}\ddot{g} - \ddot{f}\dot{g}, M = \sigma_{uv} \cdot \mathbf{N} = 0, N = \sigma_{vv} \cdot \mathbf{N} = f\dot{g}$$

so the second fundamental form is

$$(\dot{f}\ddot{g} - \ddot{f}\dot{g})du^2 + f\dot{g}dv^2.$$

For the unit sphere S^2 in latitude-longitude coordinates (Example 2.9.5), $u = \theta, v = \varphi, f(\theta) = \cos\theta, g(\theta) = \sin\theta$, giving the second fundamental form of S^2 as

$$d\theta^2 + \cos^2\theta d\varphi^2.$$

Note that this is the same as the first fundamental form of S^2 (see Example 3.4.5).

If the surface is the unit cylinder, we can take $f(u) = 1, g(u) = u$ (again, the conditions $f > 0$ and $\dot{f}^2 + \dot{g}^2 = 1$ are satisfied). This gives $L = M = 0, \quad N = 1$, so the second fundamental form of the cylinder is dv^2.

Exercise 3.6.3. Compute the second fundamental form of the elliptic paraboloid

$$\sigma(u, v) = \left(u, v, u^2 + v^2\right).$$

Solution:
$\sigma_u = (1, 0, 2u), \sigma_v = (0, 1, 2v),$ so $\mathbf{N} = \lambda(-2u, -2v, 1)$, where $\lambda = (1 + 4u^2 + 4v^2)^{-1/2}$; $\sigma_{uu} = (0, 0, 2), \sigma_{uv} = 0, \sigma_{vv} = (0, 0, 2)$, so $L = 2\lambda, M = 0, N = 2\lambda$, and the second fundamental form is $2\lambda\left(du^2 + dv^2\right)$.

3.7 The Gauss and Weingarten maps

Let $\mathcal{S}$ be an oriented surface with unit normal $\mathbf{N}$. The values of $\mathbf{N}$ at the points of $\mathcal{S}$ are recorded by *Gauss map* $\mathcal{G}_{\mathcal{S}}$ (or $\mathcal{G}$) This is the map from $\mathcal{S}$ to the unit sphere S^2 that assigns to any point $\mathbf{p} \in \mathcal{S}$ the point $\mathbf{N_p} \in S^2$, where $\mathbf{N_p}$ is the unit normal of $\mathcal{S}$ at $\mathbf{p}$. The rate at which $\mathbf{N}$ varies across $\mathcal{S}$ is measured by the derivative of $\mathcal{G}$:

$$D_{\mathbf{p}}\mathcal{G} : T_{\mathbf{p}}\mathcal{S} \to T_{\mathcal{G}}(\mathbf{p})S^2.$$

Now, the tangent plane at a point $\mathbf{q} \in S^2$ is the plane passing through the origin perpendicular to $\mathbf{q}$. Thus, $T_{\mathcal{G}}(\mathbf{p})S^2$ is the plane through the origin perpendicular to $\mathbf{N_p}$, in other words, $T_{\mathbf{p}}\mathcal{S}$. Thus, the derivative of $\mathcal{G}$ is a linear map from the tangent plane of $\mathcal{S}$ to itself:

$$D_{\mathbf{p}}\mathcal{G} : T_{\mathrm{p}}\mathcal{S} \to T_{\mathrm{p}}\mathcal{S}.$$

Definition 3.7.1. Let $\mathbf{p}$ be a point of a surface $\mathcal{S}$. *The Weingarten map* $\mathcal{W}_{\mathbf{p},\mathcal{S}}$ of $\mathcal{S}$ at $\mathbf{p}$ is defined by

$$\mathcal{W}_{\mathbf{p},\mathcal{S}} = -D_{\mathbf{p}}\mathcal{G}.$$

The *second fundamental form* of $\mathcal{S}$ at $\mathbf{p} \in \mathcal{S}$ is the bilinear form on $T_{\mathrm{p}}\mathcal{S}$ given by

$$\langle\langle \mathbf{v}, \mathbf{w} \rangle\rangle_{\mathbf{p},\mathcal{S}} = \langle \mathcal{W}_{\mathbf{p},\mathcal{S}}(\mathbf{v}), \mathbf{w} \rangle_{\mathbf{p},\mathcal{S}}, \quad \mathbf{v}, \mathbf{w} \in T_{\mathbf{p}}\mathcal{S}.$$

The bilinearity in this definition is easy to check. Let $\mathbf{v}_1, \mathbf{v}_2, \mathbf{w} \in T_{\mathbf{p}}\mathcal{S}, \lambda_1, \lambda_2 \in \mathbb{R}$. Then,

$$\begin{aligned}
\langle\langle \lambda_1\mathbf{v}_1 + \lambda_2\mathbf{v}_2, \mathbf{w} \rangle\rangle &= \langle \mathcal{W}(\lambda_1\mathbf{v}_1 + \lambda_2\mathbf{v}_2), \mathbf{w} \rangle \\
&= \langle \lambda_1\mathcal{W}(\mathbf{v}_1) + \lambda_2\mathcal{W}(\mathbf{v}_2), \mathbf{w} \rangle \\
&= \lambda_1 \langle \mathcal{W}(\mathbf{v}_1), \mathbf{w} \rangle + \lambda_2 \langle \mathcal{W}(\mathbf{v}_2), \mathbf{w} \rangle \\
&= \lambda_1 \langle\langle \mathbf{v}_1, \mathbf{w} \rangle\rangle + \lambda_2 \langle\langle \mathbf{v}_2, \mathbf{w} \rangle\rangle
\end{aligned}$$

where we used the linearity of $\mathcal{W}$ in passing from the first line to the second and the bilinearity of $\langle , \rangle$, in passing from the second line to the third. This proves that $\langle\langle \mathbf{v}, \mathbf{w} \rangle\rangle$ is a linear function of $\mathbf{v}$ for each fixed $\mathbf{w}$. The proof that it is linear in $\mathbf{w}$ for each fixed $\mathbf{v}$ is similar.

Lemma 3.7.2. Let $\sigma(u, v)$ be a surface patch with standard unit normal $\mathbf{N}(u, v)$. Then,

$$\mathbf{N}_u \cdot \sigma_u = -L, \quad \mathbf{N}_u \cdot \sigma_v = \mathbf{N}_v \cdot \sigma_u = -M, \quad \mathbf{N}_v \cdot \sigma_v = -N$$

where L, M and N are as defined in Equation 3.1.

Proof. Since σ_u and σ_v are tangent vectors to the surface patch,

$$\mathbf{N} \cdot \sigma_u = \mathbf{N} \cdot \sigma_v = 0$$

Differentiating these equations with respect to u and v gives

$$\begin{aligned}
\mathbf{N}_u \cdot \sigma_u = -\mathbf{N} \cdot \sigma_{uu} = -L, \quad \mathbf{N}_v \cdot \sigma_u = -\mathbf{N} \cdot \sigma_{uv} = -M \\
\mathbf{N}_u \cdot \sigma_v = -\mathbf{N} \cdot \sigma_{uv} = -M, \quad \mathbf{N}_v \cdot \sigma_v = -\mathbf{N} \cdot \sigma_{vv} = -N.
\end{aligned}$$

□

Proposition 3.7.3. Let $\mathbf{p}$ be a point of a surface $\mathcal{S}$, let $\sigma(u, v)$ be a surface patch of $\mathcal{S}$ with $\mathbf{p}$ in its image, and let $Ldu^2 + 2Mdudv + Ndv^2$ be the second fundamental form of σ defined in Section 3.6. Then, for any $\mathbf{v}, \mathbf{w} \in T_{\mathbf{p}}\mathcal{S}$,

$$\langle\langle \mathbf{v}, \mathbf{w} \rangle\rangle = Ldu(\mathbf{v})du(\mathbf{w}) + M(du(\mathbf{v})dv(\mathbf{w}) + du(\mathbf{w})dv(\mathbf{v})) + Ndv(\mathbf{v})dv(\mathbf{w}). \tag{3.2}$$

Proof. Since both sides of the equation in the statement of the proposition define bilinear forms on $T_{\mathrm{p}}\mathcal{S}$, it suffices to verify that they agree when v and $\mathbf{w}$ are σ_u or σ_v. Recalling that $du(\sigma_u) = dv(\sigma_v) = 1, du(\sigma_v) = dv(\sigma_u) = 0$, we have to prove that

$$\langle\langle \sigma_u, \sigma_u \rangle\rangle = L, \quad \langle\langle \sigma_u, \sigma_v \rangle\rangle = \langle\langle \sigma_v, \sigma_u \rangle\rangle = M, \quad \langle\langle \sigma_v, \sigma_v \rangle\rangle = N. \tag{3.3}$$

Let $\sigma(u_0, v_0) = \mathbf{p}$. Then, with the derivatives evaluated at (u_0, v_0),

$$\mathcal{W}(\sigma_u) = -\frac{d}{du}\bigg|_{u=u_0} \mathcal{G}(\sigma(u, v_0)) = -\frac{d}{du}\bigg|_{u=u_0} \mathbf{N}(u, v_0) = -\mathbf{N}_u,$$

where $\mathbf{N}$ is the standard unit normal of σ. Similarly, $\mathcal{W}(\sigma_v) = -\mathbf{N}_v$. Hence,

$$\langle\langle \sigma_u, \sigma_u \rangle\rangle = \langle \mathcal{W}(\sigma_u), \sigma_u \rangle = -\mathbf{N}_u \cdot \sigma_u,$$

which is equal to L by Lemma 3.7.2. The other equations in 3.3 are proved similarly.

□

Corollary 3.7.4. The second fundamental form is a symmetric bilinear form. Equivalently, the Weingarten map is self-adjoint.

Proof. The formula 3.2 for $\langle\langle \mathbf{v}, \mathbf{w} \rangle\rangle$ is obviously unchanged when $\mathbf{v}$ and $\mathbf{w}$ are interchanged. □

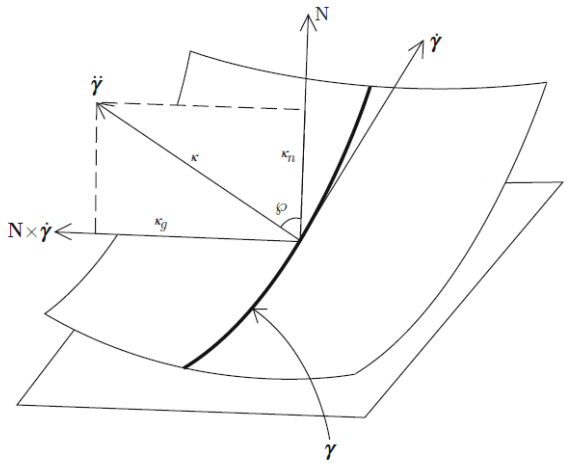
Exercise 3.7.5. Calculate the Gauss map $\mathcal{G}$ of the paraboloid $\mathcal{S}$ with equation $z = x^2 + y^2$. What is the image of $\mathcal{G}$?

Solution:
The paraboloid is the level surface $f = 0$ where $f(x, y, z) = z - x^2 - y^2$ and $\mathbf{N} = \frac{(f_x, f_y, f_z)}{\|(x_x, f_y, f_z)\|}$ is the corresponding unit normal. So $\mathcal{G}(x, y, z) = \frac{(-2x, -2y, 1)}{(4x^2+4y^2+1)^{1/2}}$.

3.8 Normal and geodesic curvatures

Definition 3.8.1. If γ is a unit-speed curve on an oriented surface $\mathcal{S}$, then $\dot{\gamma}$ is a unit vector and is, by definition, a tangent vector to $\mathcal{S}$. Hence, $\dot{\gamma}$ is perpendicular to the unit normal N of $\mathcal{S}$, so $\dot{\gamma}, \mathrm{N}$ and $\mathrm{N} \times \dot{\gamma}$ are mutually perpendicular unit vectors. Again since γ is unit-speed, $\ddot{\gamma}$ is perpendicular to $\dot{\gamma}$, and hence is a linear combination of N and $\mathrm{N} \times \dot{\gamma}$:

$$\ddot{\gamma} = \kappa_n \, \mathrm{N} + \kappa_g \, \mathrm{N} \times \dot{\gamma} \tag{3.4}$$



The scalars κ_n and κ_g in Equation 3.4 are called the *normal curvature* and the *geodesic curvature* of γ, respectively.

Note that κ_n and κ_s both change sign when $\mathbf{N}$ is replaced by $-\mathbf{N}$, so on a general (not necessarily orientable) surface only the magnitudes of κ_n and κ_s are well defined.

Proposition 3.8.2. If γ is a unit-speed curve on an oriented surface $\mathcal{S}$ with unit normal $\mathbf{N}$, κ_n and κ_g are the *normal curvature* and the *geodesic curvature* of γ respectively. Then

$$\kappa_n = \ddot{\gamma} \cdot \mathbf{N}, \quad \kappa_g = \ddot{\gamma} \cdot (\mathbf{N} \times \dot{\gamma}) \tag{3.5}$$

$$\kappa^2 = \kappa_n^2 + \kappa_g^2, \tag{3.6}$$

$$\kappa_n = \kappa \cos\psi, \quad \kappa_g = \pm\kappa \sin\psi \tag{3.7}$$

where κ is the curvature of γ and ψ is the angle between N and the principal normal $\mathbf{n}$ of γ.

Proof. Equations 3.5 and 3.6 follow from Equation 3.4 and the fact that $\mathbf{N}$ and $\mathbf{N} \times \dot{\gamma}$ are perpendicular unit vectors. The first equation in 3.7 follows from $\ddot{\gamma} = \kappa\mathbf{n}$, and the second then follows from Equation 3.6. □

If γ is regular, but not necessarily unit-speed, we define the geodesic and normal curvatures of γ to be those of a unit-speed reparametrization of γ. When a unit-speed parameter t is changed to another such parameter $\pm t + c$, where c is a constant, it is clear that $\kappa_n \mapsto \kappa_n$ and $\kappa_g \mapsto \pm\kappa_g$, so κ_n is well defined for any regular curve, while κ_g is well defined up to sign. Equations 3.6 and 3.7 continue to hold if γ is any regular curve.

Proposition 3.8.3. If γ is a unit-speed curve on an oriented surface $\mathcal{S}$, its normal curvature is given by

$$\kappa_n = \langle\langle \dot{\gamma}, \dot{\gamma} \rangle\rangle.$$

If σ is a surface patch of $\mathcal{S}$ and $\gamma(t) = \sigma(u(t), v(t))$ is a curve in σ

$$\kappa_n = L\dot{u}^2 + 2M\dot{u}\dot{v} + N\dot{v}^2$$

in the notation of Section 3.6.

Proof. Since $\dot{\gamma}$ is a tangent vector to $\mathcal{S}, \mathbf{N} \cdot \dot{\gamma} = 0$. Hence, $\mathbf{N} \cdot \ddot{\gamma} = -\dot{\mathbf{N}} \cdot \dot{\gamma}$ so

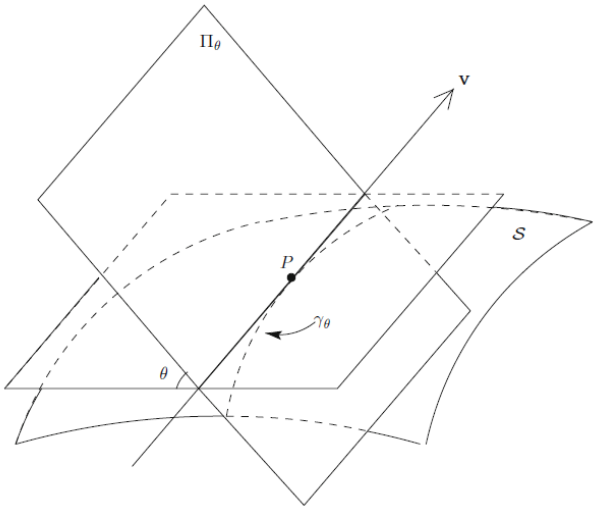
$$\kappa_n = \mathbf{N} \cdot \ddot{\gamma} = -\dot{\mathbf{N}} \cdot \dot{\gamma} = \langle \mathcal{W}(\dot{\gamma}), \dot{\gamma} \rangle = \langle\langle \dot{\gamma}, \dot{\gamma} \rangle\rangle$$

since

$$\dot{\mathbf{N}} = \frac{d}{dt}\mathcal{G}(\gamma(t)) = -\mathcal{W}(\dot{\gamma}).$$

The second part follows from the first and Proposition 3.7.3. □

Proposition 3.8.4. Meusnier's Theorem: Let $\mathbf{p}$ be a point of a surface $\mathcal{S}$ and let $\mathbf{v}$ be a unit tangent vector to $\mathcal{S}$ at $\mathbf{p}$. Let Π_θ be the plane containing the line through $\mathbf{p}$ parallel to $\mathbf{v}$ and making an angle θ with the tangent plane $T_{\mathrm{p}}\mathcal{S}$, and assume that Π_θ is not parallel to $T_{\mathrm{p}}\mathcal{S}$. Suppose that Π_θ intersects $\mathcal{S}$ in a curve with curvature κ_θ. Then, $\kappa_\theta \sin\theta$ is independent of θ.



Proof. Assume that γ_θ is a unit-speed parametrization of the curve of intersection of Π_θ and $\mathcal{S}$. Then, at $\mathbf{p}, \dot{\gamma}_\theta = \pm\mathbf{v}$, so $\ddot{\gamma}_\theta$ is perpendicular to $\mathbf{v}$ and is

parallel to Π_θ. Thus, in the notation of Proposition 3.8.2, $\psi = \pi/2 - \theta$ and so Equation 3.7 gives

$$\kappa_\theta \sin\theta = \kappa_n.$$

But κ_n depends only on $\mathbf{p}$ and $\mathbf{v}$, and not on θ. □

An important special case is that in which γ is a *normal section* of the surface, i.e., γ is the intersection of the surface with a plane Π that is perpendicular to the tangent plane of the surface at every point of γ.

Corollary 3.8.5. The curvature κ, normal curvature κ_n and geodesic curvature κ_g of a normal section of a surface are related by

$$\kappa_n = \pm\kappa, \quad \kappa_g = 0.$$

Proof. As in the proof of Proposition 3.8.3, $\kappa_n = \kappa\sin\theta$, where $\theta = \pm\pi/2$ for a normal section. This gives the first equation; the second follows from it and Equation 3.6. □

Exercise 3.8.6. Show that the normal curvature of any curve on a sphere of radius r is $\pm 1/r$.
Solution:

Let γ be a unit-speed curve on the sphere of centre $\mathbf{a}$ and radius r. Then, $(\gamma - \mathbf{a}) \cdot (\gamma - \mathbf{a}) = r^2$; differentiating gives $\dot\gamma \cdot (\gamma - \mathbf{a}) = 0,$ so $\ddot\gamma \cdot (\gamma - \mathbf{a}) = -\dot\gamma \cdot \dot\gamma = -1$. At the point $\gamma(t),$ the unit normal of the sphere is $\mathbf{N} = \pm\frac{1}{r}(\gamma(t) - \mathbf{a}),$ so $\kappa_n = \ddot\gamma \cdot \mathbf{N} = \pm\frac{1}{r}\ddot\gamma \cdot (\gamma - \mathbf{a}) = \mp\frac{1}{r}$.

3.9 Gaussian and mean curvatures

Definition 3.9.1. Let $\mathcal{W}$ be the Weingarten map of an oriented surface $\mathcal{S}$ at a point $\mathbf{p} \in \mathcal{S}$. The Gaussian curvature K and mean curvature H of $\mathcal{S}$ at $\mathbf{p}$ are defined by

$$K = \det(\mathcal{W}), \quad H = \frac{1}{2}\,\mathrm{trace}(\mathcal{W}).$$

Let $\sigma(u, v)$ be a surface patch with first and second fundamental forms

$$Edu^2 + 2Fdudv + Gdv^2 \quad \text{and} \quad Ldu^2 + 2Mdudv + Ndv^2,$$

respectively. Define symmetric 2×2 matrices $\mathcal{F}_I$ and $\mathcal{F}_{II}$ by

$$\mathcal{F}_I = \begin{pmatrix} E & F \\ F & G \end{pmatrix}, \quad \mathcal{F}_{II} = \begin{pmatrix} L & M \\ M & N \end{pmatrix}.$$

Proposition 3.9.2. Let σ be a surface patch of an oriented surface $\mathcal{S}$. Then, with the above notation, the matrix of $\mathcal{W}_{\mathbf{p},\mathcal{S}}$ with respect to the basis $\{\sigma_u, \sigma_v\}$ of $T_{\mathrm{p}}\mathcal{S}$ is $\mathcal{F}_I^{-1}\mathcal{F}_{II}$.

Proof. By the proof of Proposition 3.7.3, $\mathcal{W}(\sigma_u) = -\mathbf{N}_u$ and $\mathcal{W}(\sigma_v) = -\mathbf{N}_v$, so the matrix of $\mathcal{W}$ is $\begin{pmatrix} a & c \\ b & d \end{pmatrix}$, where

$$-\mathbf{N}_u = a\sigma_u + b\sigma_v, \quad -\mathbf{N}_v = c\sigma_u + d\sigma_v.$$

Take the dot product of each of these equations with σ_u and σ_v and use Lemma 3.7.2; this gives

$$\begin{aligned} L = aE + bF, \quad M = cE + dF \\ M = aF + bG, \quad N = cF + dG \end{aligned}.$$

These four scalar equations are equivalent to the single matrix equation

$$\begin{pmatrix} L & M \\ M & N \end{pmatrix} = \begin{pmatrix} E & F \\ F & G \end{pmatrix} \begin{pmatrix} a & c \\ b & d \end{pmatrix}$$

$$\text{i.e.,} \quad \mathcal{F}_{II} = \mathcal{F}_I \begin{pmatrix} a & c \\ b & d \end{pmatrix}$$

Hence, the matrix of $\mathcal{W}$ with respect to the basis $\{\sigma_u, \sigma_v\}$ is

$$\begin{pmatrix} a & c \\ b & d \end{pmatrix} = \mathcal{F}_I^{-1}\mathcal{F}_{II}.$$

□

Corollary 3.9.3. We have

$$H = \frac{LG - 2MF + NE}{2(EG - F^2)}, \quad K = \frac{LN - M^2}{EG - F^2}.$$

Proof. By Definition 3.9.1,

$$K = \det\left(\mathcal{F}_I^{-1}\mathcal{F}_{II}\right) = \frac{\det(\mathcal{F}_{II})}{\det(\mathcal{F}_I)} = \frac{LN - M^2}{EG - F^2}.$$

To compute H, we need the trace of the matrix

$$\begin{aligned} \mathcal{F}_I^{-1}\mathcal{F}_{II} &= \frac{1}{EG - F^2} \begin{pmatrix} G & -F \\ -F & E \end{pmatrix} \begin{pmatrix} L & M \\ M & N \end{pmatrix} \\ &= \frac{1}{EG - F^2} \begin{pmatrix} LG - MF & MG - NF \\ ME - LF & NE - MF \end{pmatrix}. \end{aligned}$$

Thus,

$$2H = \text{trace}\left(\mathcal{F}_I^{-1}\mathcal{F}_{II}\right) = \frac{LG - 2MF + NE}{EG - F^2}.$$

□

Example 3.9.4. In Examples 3.4.5 and 3.6.2 we considered the surface of revolution

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

where we can assume that $f > 0$ and $\dot{f}^2 + \dot{g}^2 = 1$ everywhere (a dot denoting d/du). We found that

$$E = 1, F = 0, G = f^2, \quad L = \dot{f}\ddot{g} - \ddot{f}\dot{g}, M = 0, N = f\dot{g}.$$

By Corollary 3.9.3, the Gaussian curvature is

$$K = \frac{LN - M^2}{EG - F^2} = \frac{(\dot{f}\ddot{g} - \ddot{f}\dot{g})f\dot{g}}{f^2}.$$

We can simplify this formula by noting that $\dot{f}^2 + \dot{g}^2 = 1$ implies (by differentiating with respect to u) that $\dot{f}\ddot{f} + \dot{g}\ddot{g} = 0$,

$$\therefore \quad (\dot{f}\ddot{g} - \ddot{f}\dot{g})\dot{g} = -\dot{f}^2\ddot{f} - \ddot{f}\dot{g}^2 = -\ddot{f}\left(\dot{f}^2 + \dot{g}^2\right) = -\ddot{f},$$

$$\therefore \quad K = -\frac{\ddot{f}f}{f^2} = -\frac{\ddot{f}}{f}.$$

We consider some special cases. If $\gamma(u) = (u, 0, 0)$ is the x-axis, the corresponding surface of revolution is the xy-plane; since $f(u) = u$, we have $\dot{f} = 1, \ddot{f} = 0$, so $K = 0$. If $\gamma(u) = (1, 0, u)$ is a straight line parallel to the z-axis, the corresponding surface is the unit cylinder; since $f(u) = 1, \ddot{f} = 0$, so $K = 0$. Finally, if $\gamma(u) = (\cos u, 0, \sin u)$ is a circle of radius 1, the corresponding surface is the unit sphere; since $f(u) = \cos u, \dot{f} = -\sin u, \ddot{f} = -\cos u$ so $K = -\ddot{f}/f = -(-\cos u)/\cos u = 1$. Note that in each of these examples the curve γ is unit-speed.

Example 3.9.5. For a ruled surface, take a patch

$$\sigma(u, v) = \gamma(u) + v\delta(u)$$

(see Example 3.2.1) Denoting d/du by a dot, we have $\sigma_u = \dot{\gamma} + v\dot{\delta}, \sigma_v = \delta$, so

$$\sigma_{uv} = \dot{\delta}, \quad \sigma_{vv} = 0.$$

Hence, if $\mathbf{N} = (\sigma_u \times \sigma_v) / \|\sigma_u \times \sigma_v\|$ is the standard unit normal of σ, then $M = \sigma_{uv} \cdot \mathbf{N} = \dot{\delta} \cdot \mathbf{N}$ and $N = 0$. So

$$K = \frac{LN - M^2}{EG - F^2} = \frac{-(\dot{\delta} \cdot \mathrm{N})^2}{EG - F^2} \leq 0,$$

i.e., *the Gaussian curvature of a ruled surface is negative or zero.*

Theorem 3.9.6. Let $\sigma : U \to \mathbb{R}^3$ be a surface patch, let $(u_0, v_0) \in U$, and let $\delta > 0$ be such that the closed disc

$$R_\delta = \left\{(u, v) \in \mathbb{R}^2 \mid (u - u_0)^2 + (v - v_0)^2 \leq \delta^2\right\}$$

with centre (u_0, v_0) and radius δ is contained in U. Then,

$$\lim_{\delta \to 0} \frac{\mathcal{A}_{\mathrm{N}}(R_\delta)}{\mathcal{A}_\sigma(R_\delta)} = |K|,$$

where K is the Gaussian curvature of σ at $\sigma(u_0, v_0)$.

Note that a δ with the properties in the statement of the theorem exists because U is open.

Proof. By Definition 3.5.1,

$$\frac{\mathcal{A}_{\mathrm{N}}(R_\delta)}{\mathcal{A}_\sigma(R_\delta)} = \frac{\int_{R_\delta} \|\mathbf{N}_u \times \mathbf{N}_v\| \, dudv}{\int_{R_\delta} \|\sigma_u \times \sigma_v\| \, dudv}. \tag{3.8}$$

In the notation of the proof of Proposition 3.9.2,

$$\begin{aligned}
\mathbf{N}_u \times \mathbf{N}_v &= (a\sigma_u + b\sigma_v) \times (c\sigma_u + d\sigma_v) \\
&= (ad - bc)\sigma_u \times \sigma_v \\
&= \det\left(\mathcal{F}_I^{-1}\mathcal{F}_{II}\right) \sigma_u \times \sigma_v \\
&= \frac{\det(\mathcal{F}_{II})}{\det(\mathcal{F}_I)} \boldsymbol{\sigma}_u \times \boldsymbol{\sigma}_v \\
&= \frac{LN - M^2}{EG - F^2} \boldsymbol{\sigma}_u \times \sigma_v \\
&= K\boldsymbol{\sigma}_u \times \sigma_v \quad \text{(by Corollary 3.9.3).}
\end{aligned}$$

Substituting in Equation 3.8 we get

$$\frac{\mathcal{A}_{\mathrm{N}}(R_\delta)}{\mathcal{A}_{\boldsymbol{\sigma}}(R_\delta)} = \frac{\int_{R_\delta} |K| \, \|\boldsymbol{\sigma}_u \times \boldsymbol{\sigma}_v\| \, dudv}{\int_{R_\delta} \|\boldsymbol{\sigma}_u \times \boldsymbol{\sigma}_v\| \, dudv}$$

Let ϵ be any positive number. Since $K(u,v)$ is a continuous function of (u,v), we can choose $\delta > 0$ so small that

$$|K(u,v) - K(u_0,v_0)| < \epsilon$$

if $(u,v) \in R_\delta$. Since, for any real numbers $a, b, |a-b| \geq \|a| - |b\|$, it follows that $\|K(u,v)| - |K(u_0,v_0)\| < \epsilon$ if $(u,v) \in R_\delta$, i.e.,

$$|K(u_0,v_0)| - \epsilon < |K(u,v)| < |K(u_0,v_0)| + \epsilon$$

if $(u,v) \in R_\delta$. Multiplying through by $\|\sigma_u \times \sigma_v\|$ and integrating over R_δ, we get

$$\begin{aligned}
(|K(u_0,v_0)| - \epsilon) \int \|\sigma_u \times \sigma_v\| \, dudv &< \int |K(u,v)| \, \|\sigma_u \times \sigma_v\| \, dudv \\
&< (|K(u_0,v_0)| + \epsilon) \int \|\sigma_u \times \sigma_v\| \, dudv, \\
\therefore \quad |K(u_0,v_0)| - \epsilon < \frac{\mathcal{A}_{\mathrm{N}}(R_\delta)}{\mathcal{A}_\sigma(R_\delta)} &< |K(u_0,v_0)| + \epsilon \quad \text{(using Equation 3.8)} \\
\therefore \left| \frac{\mathcal{A}_{\mathrm{N}}(R_\delta)}{\mathcal{A}_\sigma(R_\delta)} - |K(u_0,v_0)| \right| &< \epsilon
\end{aligned}$$

This proves the theorem.

□

Exercise 3.9.7. Show that the Gaussian and mean curvatures of the surface $z = f(x,y)$, where f is a smooth function, are

$$K = \frac{f_{xx}f_{yy} - f_{xy}^2}{\left(1 + f_x^2 + f_y^2\right)^2}, \quad H = \frac{\left(1+f_y^2\right) f_{xx} - 2f_x f_y f_{xy} + \left(1+f_x^2\right) f_{yy}}{2\left(1 + f_x^2 + f_y^2\right)^{3/2}}.$$

Solution:

Parametrize the surface by $\boldsymbol{\sigma}(x,y) = (x,y,f(x,y))$. Then, $\boldsymbol{\sigma}_x = (1,0,f_x)$,

$\sigma_y = (0,1,f_y), \mathbf{N} = \left(1 + f_x^2 + f_y^2\right)^{-1/2} (-f_x, -f_y, 1), \sigma_{xx} = (0,0,f_{xx}),$
$\sigma_{xy} = (0,0,f_{xy}), \sigma_{yy} = (0,0,f_{yy})$. This gives $E = 1 + f_x^2, F = f_x f_y$

$G = 1 + f_y^2$ and $L = \left(1 + f_x^2 + f_y^2\right)^{-1/2} f_{xx}, M = \left(1 + f_x^2 + f_y^2\right)^{-1/2} f_{xy}$
$N = \left(1 + f_x^2 + f_y^2\right)^{-1/2} f_{yy}$. By Corollary 8.1.3, $K = \dfrac{f_{xx}f_{yy} - f_{xy}^2}{\left(1 + f_x^2 + f_y^2\right)^2}, H =$
$\dfrac{\left(1+f_y^2\right) f_{xx} - 2f_x f_y f_{xy} + \left(1+f_x^2\right) f_{yy}}{2\left(1 + f_x^2 + f_y^2\right)^{3/2}}$

Exercise 3.9.8. Calculate the Gaussian curvature of the helicoid $\sigma(u,v) = (v\cos u, v\sin u, \lambda u$ and catenoid (The surface obtained by rotaing the curve $x = \cosh z$ in the $xz-$ plane around the $z-$ axis can be parametrized by $\sigma(u,v) = (\cosh u\cos v, \cosh u\sin v, u)$ with $u \in \mathbb{R}$ and $-\pi < v < \pi$ or $0 < v < 2\pi$.

Solution: For the helicoid

$\sigma(u,v) = (v\cos u, v\sin u, \lambda u), \boldsymbol{\sigma}_u = (-v\sin u, v\cos u, \lambda)$, $\sigma_v = (\cos u, \sin u, 0), \mathbf{N} = \left(\lambda^2+v^2\right)^{-1/2}(-\lambda\sin u, \lambda\cos u, -v), \sigma_{uu} = (-v\cos u, -v\sin u, 0), \sigma_{uv} = (-\sin u, \cos u, 0), \sigma_{vv} = \mathbf{0}$. This gives $E = \lambda^2+v^2, F = 0, G = 1$ and $L = N = 0, M = \lambda/\sqrt{\lambda^2+v^2}$. Hence, $K = \left(LN-M^2\right)/\left(EG-F^2\right) = -\lambda^2/\left(\lambda^2+v^2\right)^2$.

For the catenoid $\sigma(u,v) = (\cosh u\cos v, \cosh u\sin v, u)$, $\sigma_u = (\sinh u\cos v, \sinh u\sin v, 1)$, $\sigma_v = (-\cosh u\sin v, \cosh u\cos v, 0)$, $\mathbf{N} = \operatorname{sech} u(-\cos v, -\sin v, \sinh u), \boldsymbol{\sigma}_{uu} = (\cosh u\cos v, \cosh u\sin v, 0), \boldsymbol{\sigma}_{uv} = (-\sinh u\sin v, \sinh u\cos v, 0)$, $\sigma_{vv} = (-\cosh u\cos v, -\cosh u\sin v, 0)$

This gives $E = G = \cosh^2 u, F = 0$ and $L = -1, M = 0, N = 1$. Hence, $K = \left(LN-M^2\right)/\left(EG-F^2\right) = -\operatorname{sech}^4 u$.

3.10 Principal curvatures of a surface

Proposition 3.10.1. Let $\mathbf{p}$ be a point of a surface $\mathcal{S}$. There are scalars κ_1, κ_2 and a basis $\{\mathbf{t}_1, \mathbf{t}_2\}$ of the tangent plane $T_{\mathrm{p}}\mathcal{S}$ such that

$$\mathcal{W}(\mathbf{t}_1) = \kappa_1\mathbf{t}_1, \quad \mathcal{W}(\mathbf{t}_2) = \kappa_2\mathbf{t}_2.$$

Moreover, if $\kappa_1 \neq \kappa_2$, then $\langle \mathbf{t}_1, \mathbf{t}_2\rangle = 0$.

Proof. By Corollary 3.7.4 $\mathcal{W}$ is self-adjoint.Then apply the result 3.4.2 □

The real numbers κ_1 and κ_2 are the eigenvalues of $\mathcal{W}$, and $\mathbf{t}_1$ and $\mathbf{t}_2$ are corresponding eigenvectors. But in this situation, we adopt a special terminology: κ_1 and κ_2 are called the *principal curvatures* of $\mathcal{S}$, and $\mathbf{t}_1$ and $\mathbf{t}_2$ are called *principal vectors* corresponding to κ_1 and κ_2.

Points of the surface at which the two principal curvatures are equal (to κ, say) are called *umbilics*. At an umbilic, the equations $\mathcal{W}(\mathbf{t}_1) = \kappa\mathbf{t}_1$ and $\mathcal{W}(\mathbf{t}_2) = \kappa\mathbf{t}_2$ imply that $\mathcal{W}(\mathbf{t}) = \kappa\mathbf{t}$ if $\mathbf{t}$ is any linear combination of $\mathbf{t}_1$ and $\mathbf{t}_2$. Thus, $\mathbf{p}$ *is an umbilic if and only if* $\mathcal{W}_{p,\mathcal{S}}$ *is a scalar multiple of the identity map*, and in that case every tangent vector is principal. On the other

hand, if $\mathbf{p} \in \mathcal{S}$ is not an umbilic, Proposition 3.10.1 tells us that principal vectors corresponding to the two principal curvatures are necessarily orthogonal (Result:3.4.2). Thus, whether or not $\mathbf{p}$ is an umbilic we can always find two orthogonal principal vectors in $T_{\mathbf{p}}\mathcal{S}$, and we obtain:

Corollary 3.10.2. If $\mathbf{p}$ is a point of a surface $\mathcal{S}$, there is an orthonormal basis of the tangent plane $T_{\mathbf{p}}\mathcal{S}$ consisting of principal vectors.

Proposition 3.10.3. If κ_1 and κ_2 are the principal curvatures of a surface, the mean and Gaussian curvatures are given by

$$H = \frac{1}{2}(\kappa_1 + \kappa_2), \quad K = \kappa_1\kappa_2$$

Proof. The determinant and trace of the Weingarten map $\mathcal{W}$ can be computed using the matrix of $\mathcal{W}$ with respect to *any* basis of the tangent plane. Using the basis formed by the principal vectors, the matrix is

$$\begin{pmatrix} \kappa_1 & 0 \\ 0 & \kappa_2 \end{pmatrix}$$

The proposition now follows immediately from Definition 3.9.1. □

Theorem 3.10.4. Euler's Theorem: Let γ be a curve on an oriented surface $\mathcal{S}$, and let κ_1 and κ_2 be the principal curvatures of σ, with non-zero principal vectors $\mathbf{t}_1$ and $\mathbf{t}_2$. Then, the normal curvature of γ is

$$\kappa_n = \kappa_1 \cos^2\theta + \kappa_2 \sin^2\theta,$$

where θ is the oriented angle $\widehat{\mathbf{t}_1\dot{\gamma}}$.

Proof. Let $\mathbf{p} \in \mathcal{S}$, let κ_1 and κ_2 be the principal curvatures of $\mathcal{S}$ at $\mathbf{p}$, and let $\mathbf{t}_1$ and $\mathbf{t}_2$ be corresponding principal vectors. By Corollary 3.10.2, we can assume that $\{\mathbf{t}_1, \mathbf{t}_2\}$ is an orthonormal basis of $T_{\mathrm{p}}\mathcal{S}$. Moreover, by replacing $\mathbf{t}_2$ by $-\mathbf{t}_2$ if necessary, we can assume that the oriented angle $\widehat{\mathbf{t}_1\mathbf{t}_2} = +\pi/2$. With these assumptions, we have

$$\dot{\gamma} = \cos\theta\mathbf{t}_1 + \sin\theta\mathbf{t}_2.$$

By Proposition 3.8.3,

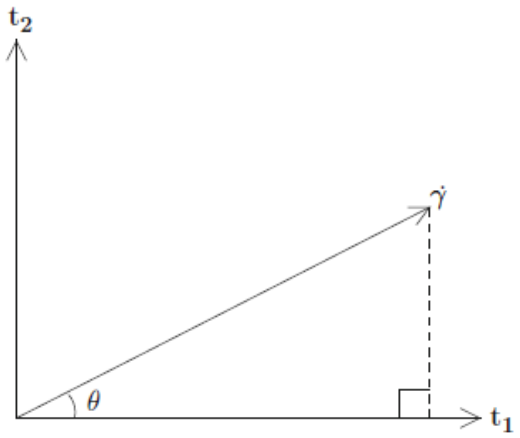
$$\kappa_n = \langle\langle\dot{\gamma}, \dot{\gamma}\rangle\rangle = \cos^2\theta\,\langle\langle\mathbf{t}_1, \mathbf{t}_1\rangle\rangle + 2\sin\theta\cos\theta\,\langle\langle\mathbf{t}_1, \mathbf{t}_2\rangle\rangle + \sin^2\theta\,\langle\langle\mathbf{t}_2, \mathbf{t}_2\rangle\rangle.$$

Now, for $i, j = 1, 2$,

$$\langle\langle\mathbf{t}_i, \mathbf{t}_j\rangle\rangle = \langle\mathcal{W}(\mathbf{t}_i), \mathbf{t}_j\rangle = \langle\kappa_i\mathbf{t}_i, \mathbf{t}_j\rangle = \begin{cases} \kappa_i & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}$$

Hence the result.

□



Corollary 3.10.5. The principal curvatures at a point of a surface are the maximum and minimum values of the normal curvature of all curves on the surface that pass through the point. Moreover, the principal vectors are the tangent vectors of the curves giving these maximum and minimum values.

Proof. If the principal curvatures κ_1 and κ_2 are different, we might as well suppose that $\kappa_1 > \kappa_2$. Let κ_n be the normal curvature of a curve γ on the surface. Then, since

$$\kappa_n = \kappa_1 \cos^2\theta + \kappa_2 \sin^2\theta = \kappa_1 - (\kappa_1 - \kappa_2)\sin^2\theta,$$

it is clear that $\kappa_n \leq \kappa_1$ with equality if and only if $\theta = 0$ or π, i.e., if and only if the tangent vector $\dot{\gamma}$ of γ is parallel to the principal vector $\mathbf{t}_1$. Similarly, one shows that $\kappa_n \geq \kappa_2$ with equality if and only if $\dot{\gamma}$ is parallel to $\mathbf{t}_2$.

If $\kappa_1 = \kappa_2$, the normal curvature of every curve is equal to κ_1 by Euler's Theorem and every tangent vector to the surface is a principal vector. □

To compute the principal curvatures, we work in a surface patch $\sigma(u, v)$; let

$$Edu^2 + 2Fdudv + Gdv^2 \quad \text{and } Ldu^2 + 2Mdudv + Ndv^2$$

be its first and second fundamental forms. In the notation of Section 3.9 , the matrix of the Weingarten map $\mathcal{W}$ with respect to the basis $\{\sigma_u, \sigma_v\}$ of the tangent plane is $\mathcal{F}_I^{-1}\mathcal{F}_{II}$. Hence, the principal curvatures are the roots κ of the equation

$$\det\left(\mathcal{F}_I^{-1}\mathcal{F}_{II} - \kappa I\right) = 0,$$

and a tangent vector $\mathbf{t} = \xi\sigma_u + \eta\sigma_v$ is a principal vector if

$$\left(\mathcal{F}_I^{-1}\mathcal{F}_{II} - \kappa I\right)\begin{pmatrix}\xi\\ \eta\end{pmatrix} = \begin{pmatrix}0\\ 0\end{pmatrix}.$$

Writing $\mathcal{F}_I^{-1}\mathcal{F}_{II} - \kappa I$ as $\mathcal{F}_I^{-1}(\mathcal{F}_{II} - \kappa\mathcal{F}_I)$, we obtain the following.

Proposition 3.10.6. In the above notation, the principal curvatures are the roots of the equation

$$\begin{vmatrix} L - \kappa E & M - \kappa F \\ M - \kappa F & N - \kappa G \end{vmatrix} = 0,$$

and the principal vectors corresponding to the principal curvature κ are the tangent vectors $\mathbf{t} = \xi\sigma_u + \eta\sigma_v$ such that

$$\begin{pmatrix} L - \kappa E & M - \kappa F \\ M - \kappa F & N - \kappa G \end{pmatrix} \begin{pmatrix} \xi \\ \eta \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

Example 3.10.7. It is intuitively clear that a sphere curves the same amount in every direction, and at every point of the sphere. Thus, we expect that the principal curvatures of a sphere are equal to each other at every point, and are constant over the sphere. To confirm this by calculation, we work with the unit sphere S^2 and use the latitude longitude parametrization as usual. We found in Example 3.4.5 that $E = 1, F = 0, G = \cos^2\theta$ and in Example 3.6.2 that $L = 1, M = 0, N = \cos^2\theta$. So the principal curvatures are the roots of

$$\begin{vmatrix} 1 - \kappa & 0 \\ 0 & \cos^2\theta - \kappa\cos^2\theta \end{vmatrix} = 0$$

i.e., $\kappa = 1$ (repeated root), as we expected. Every tangent vector is a principal vector.

Example 3.10.8. We consider the unit cylinder parametrized in the usual way:

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

We found in Example 3.4.6 that $E = 1, F = 0, G = 1$ and in Example 3.6.2 that $L = 0, M = 0, N = 1$. So the principal curvatures are the roots of

$$\begin{vmatrix} 0 - \kappa & 0 \\ 0 & 1 - \kappa \end{vmatrix} = 0$$

i.e., $\kappa = 0$ or 1 . Any principal vector $\mathbf{t}_1$ corresponding to $\kappa_1(= 1)$ satisfies

$$\begin{pmatrix} -1 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} \xi_1 \\ \eta_1 \end{pmatrix} = 0$$

so $\xi_1 = 0$ and $\mathbf{t}_1$ is a multiple of $\sigma_v = (-\sin v, \cos v, 0)$. Similarly, one finds that any principal vector corresponding to $\kappa_2(= 0)$ is a multiple of $\sigma_u = (0, 0, 1)$.

Proposition 3.10.9. Let $\mathcal{S}$ be a (connected) surface of which every point is an umbilic. Then, $\mathcal{S}$ is an open subset of a plane or a sphere.

Proof. For every tangent vector $\mathbf{t}$, we have $\mathcal{W}(\mathbf{t}) = \kappa\mathbf{t}$ where κ is the principal curvature. Let $\sigma : U \to \mathbb{R}^3$ be a surface patch of $\mathcal{S}$ with U a (connected) open subset of $\mathbb{R}^2$. Taking $\mathbf{t} = \sigma_u$ and σ_v and recalling from the proof of Proposition 3.7.3 that $\mathcal{W}(\sigma_u) = -\mathbf{N}_u, \mathcal{W}(\sigma_v) = -\mathbf{N}_v$, we get

$$\mathbf{N}_u = -\kappa\sigma_u, \quad \mathbf{N}_v = -\kappa\sigma_v \tag{3.9}$$

Hence,

$$(\kappa\sigma_u)_v = -(\mathbf{N}_u)_v = -(\mathbf{N}_v)_u = (\kappa\sigma_v)_u$$

so

$$\kappa_v\sigma_u = \kappa_u\sigma_v.$$

Since σ is regular, σ_u and σ_v are linearly independent, so the last equation implies that $\kappa_u = \kappa_v = 0$. Thus, κ is constant.

There are now two cases to consider. If $\kappa = 0$, Equation 3.9 show that N is constant. Then,

$$(\mathrm{N}\cdot\sigma)_u = \mathrm{N}\cdot\sigma_u = 0, \quad (\mathrm{N}\cdot\sigma)_v = \mathrm{N}\cdot\sigma_v = 0$$

so $\mathbf{N}\cdot\sigma$ is a constant, say c. Then $\sigma(U)$ is an open subset of the plane $\mathbf{v}\cdot\mathbf{N} = c$.

If $\kappa \neq 0$, Equation 3.9 shows that

$$\mathrm{N} = -\kappa\sigma + \mathrm{a},$$

where a is a constant vector. Hence,

$$\left\|\sigma - \frac{1}{\kappa}\mathbf{a}\right\|^2 = \left\|-\frac{1}{\kappa}\mathbf{N}\right\|^2 = \frac{1}{\kappa^2},$$

so $\sigma(U)$ is an open subset of the sphere with centre $\kappa^{-1}\mathbf{a}$ and radius κ^{-1}.

We have now proved the proposition when $\mathcal{S}$ is covered by a single surface patch. For an arbitrary surface $\mathcal{S}$, the preceding argument shows that each patch in the atlas of $\mathcal{S}$ is contained in a plane or a sphere. But if the images of two patches intersect they must clearly be part of the same plane or the same sphere. It follows that the whole of $\mathcal{S}$ is contained in a plane or a sphere.

□

Now we are going to show that how the values of the principal curvatures at a point $\mathbf{p}$ of a surface $\mathcal{S}$ provide information about the shape of $\mathcal{S}$ near $\mathbf{p}$. To simplify the situation, we assume that $\mathbf{p}$ is the origin and that $T_{\mathbf{p}}\mathcal{S}$ is

the xy-plane: this can be arranged by applying a suitable isometry of $\mathbb{R}^3$ to $\mathcal{S}$ (which does not change its shape). By a further rotation around the z-axis, we can also assume that the tangent vectors $\mathbf{t}_1 = (1, 0, 0)$ and $\mathbf{t}_2 = (0, 1, 0)$ are principal, and correspond to principal curvatures κ_1 and κ_2. Finally, by reflecting in the xy-plane if necessary, we can assume that the unit normal of $\mathcal{S}$ at $\mathbf{p}$ is $\mathbf{N} = (0, 0, 1)$.

Let σ be a surface patch of $\mathcal{S}$ with $\sigma(0, 0) = \mathbf{0}$. For any $x, y \in \mathbb{R}$, there are unique $s, t \in \mathbb{R}$ such that

$$(x, y, 0) = s\sigma_u + t\sigma_v$$

(here and below, the derivatives of σ are evaluated at $(0, 0)$). By Taylor's theorem,

$$\sigma(s, t) = \sigma(0, 0) + s\sigma_u + t\sigma_v + \frac{1}{2}\left(s^2\sigma_{uu} + 2st\sigma_{uv} + t^2\sigma_{vv}\right)$$

if we neglect terms involving higher powers of s and t. Hence, if x and y (and hence s and t) are small, we have $\sigma(s, t) = (x, y, z)$, where

$$z = \frac{1}{2}\left(s^2\sigma_{uu} + 2st\sigma_{uv} + t^2\sigma_{vv}\right) \cdot \mathbf{N} = \frac{1}{2}\left(Ls^2 + 2Mst + Nt^2\right)$$

approximately, where $Ldu^2 + 2Mdudv + Ndv^2$ is the second fundamental form of σ at the origin. If $\mathbf{t} = s\sigma_u + t\sigma_v$, then by Proposition 3.8.3,

$$Ls^2 + 2Mst + Nt^2 = \langle\langle \mathbf{t}, \mathbf{t} \rangle\rangle = \langle \mathcal{W}(\mathbf{t}), \mathbf{t} \rangle$$

Now, $\mathbf{t} = x\mathbf{t}_1 + y\mathbf{t}_2$ so

$$\mathcal{W}(\mathbf{t}) = x\mathcal{W}(\mathbf{t}_1) + y\mathcal{W}(\mathbf{t}_2) = \kappa_1 x\mathbf{t}_1 + \kappa_2 y\mathbf{t}_2 = (\kappa_1 x, \kappa_2 y, 0).$$

Hence,

$$Ls^2 + 2Mst + Nt^2 = (\kappa_1 x, \kappa_2 y, 0) \cdot (x, y, 0) = \kappa_1 x^2 + \kappa_2 y^2.$$

Hence, near the point $\mathbf{p}, \mathcal{S}$ is approximated by the quadric surface

$$z = \frac{1}{2}\left(\kappa_1 x^2 + \kappa_2 y^2\right). \tag{3.10}$$

We distinguish four cases:

(i) *κ_1 and κ_2 are both > 0 or both < 0.* Then, 3.10 is the equation of an elliptic paraboloid and one says that p is an *elliptic point* of the surface.

(ii) *κ_1 and κ_2 are of opposite sign (both non-zero).* Then, 3.10 is the equation of a hyperbolic paraboloid and one says that p is a *hyperbolic point* of the surface.

(iii) *One of κ_1 and κ_2 is zero, the other is non-zero.* Then, 3.10 is the equation of a parabolic cylinder and one says that p is a *parabolic point* of the surface.

(iv) *Both principal curvatures are zero at p.* Then, 3.10 is the equation of a plane, and one says that p is a *planar point* of the surface. In this case, one cannot determine the shape of the surface near p without examining derivatives of order higher than the second (in the non-planar case, these terms are small compared to $\kappa_1 x^2 + \kappa_2 y^2$ when x and y are small). For example, the surfaces above both have the origin as a planar point, but they have quite different shapes. (The surface on the right is called the *monkey saddle* as it is the right shape for the saddle on a bicycle ridden by a monkey: two wavs down for the two legs and a third for the tail.)

The classification of points of a surface as elliptic, hyperbolic, parabolic or planar is independent of the surface patch σ, since reparametrizing either leaves the principal curvatures unchanged or changes the sign of both of them.

Example 3.10.10. On $S^2, \kappa_1 = \kappa_2 = \pm 1$ (the sign depending on the parametrization) so all points are elliptic (and umbilics). On a circular cylinder, $\kappa_1 = \pm 1, \kappa_2 = 0$, so every point is parabolic (and there are no umbilics). On a plane, $\kappa_1 = \kappa_2 = 0$ so all points are planar (!) (and umbilics).

Example 3.10.11. For the torus $\sigma(\theta, \varphi) = ((a+b\cos\theta)\cos\varphi, (a+b\cos\theta)\sin\varphi, b\sin\theta)$, we find that the first and second fundamental forms are

$$b^2 d\theta^2 + (a + b\cos\theta)^2 d\varphi^2 \text{ and } b d\theta^2 + (a + b\cos\theta)\cos\theta d\varphi^2,$$

respectively, so the principal curvatures are

$$\kappa_1 = \frac{1}{b}, \quad \kappa_2 = \frac{\cos\theta}{a + b\cos\theta}.$$

Since $\kappa_1 > 0$ (everywhere), the point $\sigma(\theta, \varphi)$ of the torus is elliptic, parabolic or hyperbolic according to κ_2 is $> 0, = 0$ or < 0, respectively; from the formula for κ_2, these are the regions of the torus given by $-\pi/2 < \theta < \pi/2, \theta = \pm\pi/2$ and $\pi/2 < \theta < 3\pi/2$, respectively.

Exercise 3.10.12. Calculate the principal curvatures of the helicoid, and the catenoid, (Refer: Exercise 3.9.8).

Solution:

For the helicoid $\sigma(u,v) = (v\cos u, v\sin u, \lambda u)$, the first and second fundamental forms are $(\lambda^2+v^2)\,du^2 + dv^2$ and $2\lambda du dv/\sqrt{\lambda^2+v^2}$, respectively. Hence, the principal curvatures are the roots of $\begin{vmatrix} -\kappa\,(\lambda^2+v^2) & \frac{\lambda}{\sqrt{\lambda^2+v^2}} \\ \frac{\lambda}{\sqrt{\lambda^2+v^2}} & -\kappa \end{vmatrix} = 0$, i.e., $\kappa = \pm\lambda/\,(\lambda^2+v^2)$. For the catenoid $\sigma(u,v) = (\cosh u\cos v, \cosh u\sin v, u)$, the first and second fundamental forms are

$cosh^2 u\left(du^2+dv^2\right)$ and $-du^2+dv^2$, so the principal curvatures are the roots of $\begin{vmatrix} -1-\kappa\cosh^2 u & 0 \\ 0 & 1-\kappa\cosh^2 u \end{vmatrix} = 0$, i.e., $\kappa = \pm\,\mathrm{sech}^2\, u$.

Reference

(1) Rudin, W. (1986). Principles of mathematical analysis (Vol. 3). New York: McGraw-hill.

(2) Pressley, A. N. (2010). Elementary differential geometry. Springer Science and Business Media.