THE CHINESE UNIVERSITY OF HONG KONG
Department of Mathematics
Enrichment Programme for Young Mathematics Talents
Towards Differential Geometry

Contents

1 Preliminary

In this chapter, we review some basic knowledge in linear algebra and all materials here can be found in any introductory book on linear algebra.

1.1 Matrices

Definition 1.1.1 (Matrix)**.** *Let m and n be positive integers. An $m \times n$* **matrix** *over $\mathbb{R}$ ($\mathbb{C}$) is a rectangular array of the form*

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}$$

where a_{ij}, $1 \leq i \leq m$, $1 \leq j \leq n$, are real (complex) numbers. We may also write a matrix as $A = [a_{ij}]$.

Definition 1.1.2 (Matrix operations)**.**

1. **Matrix addition***: Let $A = [a_{ij}]$ and $B = [b_{ij}]$ be two $m \times n$ matrices. Then*

   $$[A + B]_{ij} = a_{ij} + b_{ij}$$

   In other words

   $$\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix} + \begin{pmatrix} b_{11} & b_{12} & \cdots & b_{1n} \\ b_{21} & b_{22} & \cdots & b_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ b_{m1} & b_{m2} & \cdots & b_{mn} \end{pmatrix}$$

   $$= \begin{pmatrix} a_{11} + b_{11} & a_{12} + b_{12} & \cdots & a_{1n} + b_{1n} \\ a_{21} + b_{21} & a_{22} + b_{22} & \cdots & a_{2n} + b_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} + b_{m1} & a_{m2} + b_{m2} & \cdots & a_{mn} + b_{mn} \end{pmatrix}$$

2. **Scalar multiplication***: Let $A = [a_{ij}]$ be a $m \times n$ matrix and c be a real (complex) number. Then*

   $$[cA]_{ij} = ca_{ij}$$

In other words

$$c\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix} = \begin{pmatrix} ca_{11} & ca_{12} & \cdots & ca_{1n} \\ ca_{21} & ca_{22} & \cdots & ca_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ ca_{m1} & ca_{m2} & \cdots & ca_{mn} \end{pmatrix}.$$

3. **Matrix multiplication**: *Let* $A = [a_{ij}]$ *be an* $m \times n$ *matrix and* $B = [b_{jk}]$ *be an* $n \times p$ *matrix. The matrix product of* A *and* B *is an* $m \times p$ *matrix and*

$$[AB]_{ik} = \sum_{j=1}^{n} a_{ij}b_{jk} = a_{i1}b_{1k} + a_{i2}b_{2k} + \cdots + a_{in}b_{nk}$$

for $1 \leq i \leq m$, $1 \leq k \leq p$. *Note that the* ik*-th entry of* AB *is the sum of the products of the corresponding entries in the* i*-th row of* A *and the* k*-th column of* B.

Example 1.1.3. *If* A *is a* 3×2 *matrix and* B *is a* 2×2 *matrix, then*

$$\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{32} \end{pmatrix} \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix} = \begin{pmatrix} a_{11}b_{11} + a_{12}b_{21} & a_{11}b_{12} + a_{12}b_{22} \\ a_{21}b_{11} + a_{22}b_{21} & a_{21}b_{12} + a_{22}b_{22} \\ a_{31}b_{11} + a_{32}b_{21} & a_{31}b_{12} + a_{32}b_{22} \end{pmatrix}$$

is a 3×2 *matrix.*

Remarks 1.1.4. *Let* A, B, C *be matrices.*

1. AB *is defined only when the number of columns of* A *is equal to the number of rows of* B.

2. *In general,* $AB \neq BA$ *even when they are both defined and of the same type.*

3. *In general,* $AB = 0$ *does not implies that* $A = 0$ *or* $B = 0$.

4. *In general,* $AB = AC$ *and* $A \neq 0$ *does not implies* $B = C$.

Proposition 1.1.5 (Properties of matrix multiplication)**.** *The following equalities hold true whenever the expressions involved are defined.*

1. $(AB)C = A(BC)$
2. $(A+B)C = AC + BC$ *and* $C(A+B) = CA + CB$
3. $c(AB) = (cA)B = A(cB)$

Proof. We prove the associativity of matrix multiplication. The other two properties are obvious. Write $A = [a_{ij}]$, $B = [b_{jk}]$, $C = [c_{kl}]$. Then

$$\begin{aligned}
[(AB)C]_{il} &= \sum_k [AB]_{ik} c_{kl} \\
&= \sum_k \left(\sum_j a_{ij} b_{jk} \right) c_{kl} \\
&= \sum_k \left(\sum_j a_{ij} b_{jk} c_{kl} \right) \\
&= \sum_j \left(\sum_k a_{ij} b_{jk} c_{kl} \right) \\
&= \sum_j a_{ij} \left(\sum_k b_{jk} c_{kl} \right) \\
&= \sum_j a_{ij} [BC]_{jl} \\
&= [A(BC)]_{il}
\end{aligned}$$

Therefore $(AB)C = A(BC)$. □

Definition 1.1.6 (Transpose)**.** *The* **transpose** *of an* $m \times n$ *matrix* $A = [a_{ij}]$ *is the* $n \times m$ *matrix* A^T *obtained by interchanging rows and columns of* A*, i.e.,*

$$[A^T]_{ji} = a_{ij}$$

for $1 \leq i \leq m$, $1 \leq j \leq n$.

Proposition 1.1.7 (Properties of transpose)**.** *Let* A *and* B *be matrices.*

1. $(A^T)^T = A$
2. $(A+B)^T = A^T + B^T$

3. $(cA)^T = cA^T$
4. $(AB)^T = B^T A^T$

Definition 1.1.8 (Symmetric and anti-symmetric matrices)**.** *Let A be an $n \times n$ matrix.*

1. *We say that A is a* **symmetric matrix** *if $A^T = A$.*
2. *We say that A is an* **anti-symmetric matrix** *(or a* **skew-symmetric matrix***) if $A^T = -A$.*

Definition 1.1.9 (Diagonal matrix)**.** *An $n \times n$ matrix of the form*

$$D = \begin{pmatrix} a_{11} & & & \\ & a_{22} & & \mathit{0} \\ & & \ddots & \\ \mathit{0} & & & a_{nn} \end{pmatrix}$$

is called a **diagonal matrix***.*

Definition 1.1.10 (Zero matrix and identity matrix)**.** *Let m, n be a positive integers.*

1. *The $m \times n$* **zero matrix** *is the matrix which every entry equals to 0.*
2. *The* **identity matrix** *of size n is the matrix*

$$I = \begin{pmatrix} 1 & & & \\ & 1 & & \mathit{0} \\ & & \ddots & \\ \mathit{0} & & & 1 \end{pmatrix}$$

The zero matrix is the identity with respect to addition, that means, $A + 0 = 0 + A = A$ for any $m \times n$ matrix A. The identity matrix is the identity with respect to matrix multiplication, that means, $AI = IA = A$ for any $n \times n$ matrix A.

Let A be an $m \times n$ matrix. The **homogeneous equation** $A\mathbf{x} = \mathbf{0}$ always has the solution $\mathbf{x} = \mathbf{0}$ which is called the **trivial solution**. If $m < n$, the equation $A\mathbf{x} = \mathbf{0}$ always has a **nontrivial solution** $\mathbf{x} \neq \mathbf{0}$.

Proposition 1.1.11. *Suppose A is an $m \times n$ matrix where $m < n$. Then the homogeneous equation $A\mathbf{x} = \mathbf{0}$ has a nontrivial solution $\mathbf{x} \neq \mathbf{0}$.*

Proof. Using Gaussian elimination, one may reduce the matrix A to a row echelon form. There are at most m columns which contain leftmost nonzero leading entries. Since $m < n$, there is at least one column, say the k-th column, which does not contain a nonzero leading entry. Then there is a nontrivial solution whose k-th coordinate is nonzero. □

Definition 1.1.12 (Matrix inverse)**.** *An $n \times n$ matrix A is said to be* **invertible***, if there exists a matrix A^{-1} called the* **inverse** *of A such that*

$$AA^{-1} = A^{-1}A = I$$

where I is the identity matrix.

Inverse only makes sense for **square matrix**, that is, an $n \times n$ matrix. We know that any nonzero number has a multiplicative inverse, but inverse of a nonzero square matrix does not always exist. However the inverse of a matrix is unique whenever it exists. In the next section, we will discuss an important condition for the existence of inverse of a matrix.

Proposition 1.1.13 (Properties of inverse)**.** *Let A and B be two invertible $n \times n$ matrices over real (complex) numbers.*

1. *The inverse A^{-1} is invertible and $(A^{-1})^{-1} = A$*
2. *For any nonnegative integer k, A^k is invertible and $(A^k)^{-1} = (A^{-1})^k$. This allows us to define $A^{-k} = (A^{-1})^k$.*
3. *For any nonzero real (complex) number c, cA is invertible and $(cA)^{-1} = c^{-1}A^{-1}$*
4. *The product AB is invertible and*
$$(AB)^{-1} = B^{-1}A^{-1}$$
5. *A^T is invertible and*
$$(A^T)^{-1} = (A^{-1})^T$$

1.2 Determinant

In this section, we discuss determinant of a square matrix. Determinant can be defined in many different ways. Here we adopt the inductive definition.

Definition 1.2.1 (Determinant). *Let n be a positive integer and*

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix}$$

be an $n \times n$ matrix. The **determinant** *of A is denoted by*

$$\det(A) = \begin{vmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{vmatrix}$$

and is defined inductively by

1. *For $n = 1$, we have $\det(A) = a_{11}$.*
2. *For $n > 1$, we have*

$$\det(A) = a_{11}\det(A_{11}) - a_{12}\det(A_{12}) + \cdots + (-1)^{n+1}a_{1n}\det(A_{1n})$$

where A_{ij}, $1 \le i, j \le n$ is the submatrix of A obtained by deleting the i-th row and the j-th column of A.

Example 1.2.2.

1. *1×1 determinant:*

$$\det((a_{11})) = a_{11}$$

2. *2×2 determinant:*

$$\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} = a_{11}a_{22} - a_{12}a_{21}$$

3. 3×3 *determinant:*

$$\begin{aligned}
& \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} \\
= \ & a_{11} \begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} - a_{12} \begin{vmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{vmatrix} + a_{13} \begin{vmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{vmatrix} \\
= \ & a_{11}(a_{22}a_{33} - a_{23}a_{32}) - a_{12}(a_{21}a_{33} - a_{23}a_{31}) + a_{13}(a_{21}a_{32} - a_{22}a_{31}) \\
= \ & a_{11}a_{22}a_{33} - a_{11}a_{23}a_{32} - a_{12}a_{21}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} - a_{13}a_{22}a_{31}
\end{aligned}$$

In the definition, we define determinant inductively by expansion along the first row. In fact we can find the value of $\det(A)$ by expanding along any row or column of A. We have for any fixed $i = 1, 2, \cdots, n$,

$$\begin{aligned}
& \det(A) \\
= \ & \sum_{j=1}^{n} (-1)^{i+j} a_{ij} \det(A_{ij}) \\
= \ & (-1)^{i+1} a_{i1} \det(A_{i1}) + (-1)^{i+2} a_{i2} \det(A_{i2}) + \cdots + (-1)^{i+n} a_{in} \det(A_{in})
\end{aligned}$$

and for any fixed $j = 1, 2, \cdots, n$,

$$\begin{aligned}
& \det(A) \\
= \ & \sum_{i=1}^{n} (-1)^{i+j} a_{ij} \det(A_{ij}) \\
= \ & (-1)^{1+j} a_{1j} \det(A_{1j}) + (-1)^{2+j} a_{2j} \det(A_{2j}) + \cdots + (-1)^{n+j} a_{nj} \det(A_{nj})
\end{aligned}$$

Example 1.2.3. *We can calculate a* 4×4 *determinant as follows. Here in* *the first step, we expand the determinant along the second column.*

$$
\begin{aligned}
&\begin{vmatrix} 2 & 4 & -2 & 6 \\ 1 & 2 & 5 & 4 \\ 1 & 1 & 2 & 4 \\ 0 & 2 & -6 & 3 \end{vmatrix} \\
= \ & -4\begin{vmatrix} 1 & 5 & 4 \\ 1 & 2 & 4 \\ 0 & -6 & 3 \end{vmatrix} + 2\begin{vmatrix} 2 & -2 & 6 \\ 1 & 2 & 4 \\ 0 & -6 & 3 \end{vmatrix} - \begin{vmatrix} 2 & -2 & 6 \\ 1 & 5 & 4 \\ 0 & -6 & 3 \end{vmatrix} + 2\begin{vmatrix} 2 & -2 & 6 \\ 1 & 5 & 4 \\ 1 & 2 & 4 \end{vmatrix} \\
= \ & -4\left(\begin{vmatrix} 2 & 4 \\ -6 & 3 \end{vmatrix} - \begin{vmatrix} 5 & 4 \\ -6 & 3 \end{vmatrix}\right) + 2\left(2\begin{vmatrix} 2 & 4 \\ -6 & 3 \end{vmatrix} - \begin{vmatrix} -2 & 6 \\ -6 & 3 \end{vmatrix}\right) \\
& -\left(2\begin{vmatrix} 5 & 4 \\ -6 & 3 \end{vmatrix} - \begin{vmatrix} -2 & 6 \\ -6 & 3 \end{vmatrix}\right) + 2\left(2\begin{vmatrix} 5 & 4 \\ 2 & 4 \end{vmatrix} - \begin{vmatrix} -2 & 6 \\ 2 & 4 \end{vmatrix} + \begin{vmatrix} -2 & 6 \\ 5 & 4 \end{vmatrix}\right) \\
= \ & -4(30-39) + 2(60-30) - (78-30) + 2(24-(-20)+(-38)) \\
= \ & 60
\end{aligned}
$$

The above calculation of determinant is not very efficient. Later we will discuss more efficient methods of finding determinant.

There is a direct formula for determinant and it can be proved by induction on n.

Proposition 1.2.4 (Direct formula for determinant)**.** *Let n be a positive integer and $A = [a_{ij}]$. Then*

$$\det(A) = \sum_{\sigma \in S_n} sign(\sigma) a_{1\sigma(1)} a_{2\sigma(2)} \cdots a_{n\sigma(n)}$$

where S_n is the set of all permutations[1] *of $1, 2, \cdots, n$ and $sign(\sigma) = 1, -1$ when σ is a composition of even, odd number of transpositions*[2] *respectively.*

The formula in the above proposition can be used as an alternative definition of determinant. It has an advantage of having a simple and symmetric form. Some properties of determinant, e.g. skew-symmetry, can be proved

[1] Note that the number of elements in S_n, and hence the number of terms in the formula, is $n!$.

[2] A transposition is a permutation which interchanges two numbers and leaves the other numbers unchange. It can be proved that sign(σ) does not depend on how σ is written as composition of transpositions.

easily using the formula. However it is not very efficient to use the formula to calculate the value of determinant.

Next we explain another important way of interpreting the determinant. Write

$$A = [\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_n]$$

where $\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_n$ are column vectors of A. We may consider $\det(A)$ as a real valued function of $\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_n$. Then the determinant is a function from $(\mathbb{R}^n)^n = \mathbb{R}^n \times \mathbb{R}^n \times \cdots \times \mathbb{R}^n$ to $\mathbb{R}$ which is characterized by the following properties.

Theorem 1.2.5 (Characterizing properties of determinant)**.** *The determinant* $\det : (\mathbb{R}^n)^n \to \mathbb{R}$ *is a function characterized by the following properties.*

1. *(Multilinearity) For any* $k = 1, 2, \ldots, n$ *and* $\alpha, \beta \in \mathbb{R}$

$$\begin{aligned} & \det[\mathbf{a}_1, \ldots, \mathbf{a}_{k-1}, \alpha\mathbf{u} + \beta\mathbf{v}, \mathbf{a}_{k+1}, \ldots, \mathbf{a}_n] \\ = \; & \alpha \det[\mathbf{a}_1, \ldots, \mathbf{a}_{k-1}, \mathbf{u}, \mathbf{a}_{k+1}, \ldots, \mathbf{a}_n] + \beta \det[\mathbf{a}_1, \ldots, \mathbf{a}_{k-1}, \mathbf{v}, \mathbf{a}_{k+1}, \ldots, \mathbf{a}_n]. \end{aligned}$$

2. *(Anti-symmetry) For any* $1 \le i < j \le n$,

$$\det[\mathbf{a}_1, \ldots, \mathbf{a}_i, \ldots, \mathbf{a}_j, \ldots, \mathbf{a}_n] = -\det[\mathbf{a}_1, \ldots, \mathbf{a}_j, \ldots, \mathbf{a}_i, \ldots, \mathbf{a}_n].$$

3. *(Determinant of identity) We have*

$$\det[\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n] = 1$$

where

$$\mathbf{e}_i = (0, \ldots, 0, 1, 0 \ldots, 0)^T \in \mathbb{R}^n$$

is the n *column vector with the* i*-th entry equals to* 1 *and all other entries equal to* 0*. In other words,* $\det(I) = 1$ *where* I *is the* $n \times n$ *identity matrix.*

Furthermore, if $f : (\mathbb{R}^n)^n \to \mathbb{R}$ *is a function which is multilinear, anti-symmetric and satisfies*

$$f(\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n) = k,$$

then

$$f(\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n) = k \det(\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n)$$

for any $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n \in \mathbb{R}^n$.

Proof. It follows readily by the formula for determinant (Proposition 1.2.4) that det satisfies the three properties. Suppose $f : (\mathbb{R}^n)^n \to \mathbb{R}$ is a function which is multilinear, anti-symmetric and satisfies $f(\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_n) = k$. Consider the function $g = f - k \det$. Then g is multilinear, anti-symmetric and satisfies $g(\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_n) = 0$. Now the anti-symmetry implies that $g(\mathbf{e}_{i_1}, \mathbf{e}_{i_2}, \dots, \mathbf{e}_{i_n}) = 0$ where $i_1, i_2, \dots, i_n$ is any permutation of $1, 2, \dots, n$. Then it follows by multilinearity that $g(\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n) = 0$ for any vectors $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n \in \mathbb{R}^n$. Therefore $f = k \det$. □

In practice, we usually do not use definition to calculate the determinant because it is not efficient. In stead, we use elementary row or column operations and the following proposition allows us to do so.

Proposition 1.2.6 (Determinant under row and column operations)**.** *Let A be an $n \times n$ matrix.*

1. *If B is obtained from A by multiplying a single row (or column) of A by a constant k, then* $\det(B) = k \det(A)$.
2. *If B is obtained from A by interchanging two rows (or columns) of A, then* $\det(B) = -\det(A)$.
3. *If B is obtained from A by adding a constant multiple of one row (or column) of A to another row (or column) of A, then* $\det(B) = \det(A)$.

Example 1.2.7. *We can calculate a* 4×4 *determinant as follows.*

$$
\begin{aligned}
\begin{vmatrix} 2 & 4 & -2 & 6 \\ 1 & 2 & 5 & 4 \\ 1 & 1 & 2 & 4 \\ 0 & 2 & -6 & 3 \end{vmatrix}
&= 2\begin{vmatrix} 1 & 2 & -1 & 3 \\ 1 & 2 & 5 & 4 \\ 1 & 1 & 2 & 4 \\ 0 & 2 & -6 & 3 \end{vmatrix} \\
&= 2\begin{vmatrix} 1 & 2 & -1 & 3 \\ 0 & 0 & 6 & 1 \\ 0 & -1 & 3 & 1 \\ 0 & 2 & -6 & 3 \end{vmatrix} \\
&= -2(1)\begin{vmatrix} -1 & 3 & 1 \\ 0 & 6 & 1 \\ 2 & -6 & 3 \end{vmatrix} \\
&= -2\begin{vmatrix} -1 & 3 & 1 \\ 0 & 6 & 1 \\ 0 & 0 & 5 \end{vmatrix} \\
&= -2(-1)\begin{vmatrix} 6 & 1 \\ 0 & 5 \end{vmatrix} \\
&= 2(30) \\
&= 60
\end{aligned}
$$

Determinant has the following further properties.

Proposition 1.2.8 (Further properties of determinant)**.** *Let A be an $n \times n$ matrix.*

1. *If A has a row (or column) consisting entirely of zeros, then* $\det(A) = 0$.
2. *If two rows (or columns) of A are identical, then* $\det(A) = 0$.
3. *If A is an upper triangular matrix, that is,*

$$
A = \begin{pmatrix} a_{11} & & & * \\ & a_{22} & & \\ & & \ddots & \\ & 0 & & a_{nn} \end{pmatrix}
$$

Then $\det(A) = a_{11}a_{22}\cdots a_{nn}$. *In particular, the determinant of a diagonal matrix is the product of its diagonal entries.*

4. $\det(cA) = c^n \det(A)$ *for any* $c \in \mathbb{R}$. *(Caution!* $\det(cA) \neq c\det(A)$*)*

5. $\det(A^T) = \det(A)$

The following property of determinant is important and is less obvious.

Proposition 1.2.9. *Let A and B be two $n \times n$ matrices. Then*

$$\det(AB) = \det(A)\det(B).$$

Proof. Write

$$B = [\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n]$$

where $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ are column vectors of B and observe that

$$AB = [A\mathbf{v}_1, A\mathbf{v}_2, \ldots, A\mathbf{v}_n].$$

Consider $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ as variables of the function $f : (\mathbb{R}^n)^n \to \mathbb{R}$ defined by

$$f(\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n) = \det([A\mathbf{v}_1, A\mathbf{v}_2, \ldots, A\mathbf{v}_n]).$$

Then f is obviously multilinear and anti-symmetric. Moreover

$$f(\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n) = \det([A\mathbf{e}_1, A\mathbf{e}_2, \ldots, A\mathbf{e}_n]) = \det(A)$$

where $\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n$ is the standard basis for $\mathbb{R}^n$. Therefore

$$\begin{aligned} \det(AB) &= f(\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n) \\ &= \det(A)\det([\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n]) \text{ (Theorem 1.2.5)} \\ &= \det(A)\det(B). \end{aligned}$$

□

Let $A = [a_{ij}]$ be an $n \times n$ matrix. We define the (i, j) **cofactor** by

$$A_{ij} = (-1)^{i+j}\det(M_{ij})$$

where M_{ij} is the submatrix of A obtained by deleting the i-th row and j-th column of A. Observe that for any fixed $i = 1, 2, \ldots, n$,

$$\sum_{j=1}^{n} a_{ij}A_{ij} = a_{i1}A_{i1} + a_{i2}A_{i2} + \cdots + a_{in}A_{in} = \det(A)$$

since the left hand side is nothing but the expansion of determinant along the i-th row. On the other hand, for $k \neq i$, we have

$$\sum_{j=1}^{n} a_{kj}A_{ij} = a_{k1}A_{i1} + a_{k2}A_{i2} + \cdots + a_{kn}A_{in} = 0$$

since the left hand side is the determinant of the matrix obtained by replacing the i-th row by the k-th row which must be 0 because the i-th and the k-th row are identical. Similarly, we have

$$\sum_{i=1}^{n} A_{ij}a_{ij} = A_{1j}a_{1j} + A_{2j}a_{2j} + \cdots + A_{nj}a_{nj} = \det(A)$$

and

$$\sum_{i=1}^{n} A_{il}a_{ij} = A_{1l}a_{1j} + A_{2l}a_{2j} + \cdots + A_{nl}a_{nj} = 0$$

for $l \neq j$. The above equalities can be summarized into the following identity.

Proposition 1.2.10. *Let* $A = [a_{ij}]$ *be an* $n \times n$ *matrix and* $\mathrm{adj}(A)$ *is the* **adjugate matrix** *of* A*, that is* $[\mathrm{adj}(A)]_{ij} = A_{ji}$ *where* A_{ij} *is the* (i, j) *cofactor of* A*. Then*

$$A\mathrm{adj}(A) = \mathrm{adj}(A)A = \det(A)I$$

where I *is the* $n \times n$ *identity matrix.*

Now we have a simple criterion for a matrix to be invertible and a formula for the inverse of an invertible matrix.

Proposition 1.2.11. *Let* $A = [a_{ij}]$ *be an* $n \times n$ *matrix. Then* A *is invertible, that is, the inverse* A^{-1} *of* A *exists, if and only if* $\det(A) \neq 0$*. Moreover if* A *is invertible, then*

$$A^{-1} = \frac{1}{\det(A)}\mathrm{adj}(A)$$

where $\mathrm{adj}(A)$ *is the adjugate matrix of* A*.*

Proof. Suppose A is invertible. Then the inverse A^{-1} of A exists and we have $AA^{-1} = A^{-1}A = I$ Thus

$$\det(A)\det(A^{-1}) = \det(I) = 1$$

which implies $\det(A) \neq 0$.

Suppose $\det(A) \neq 0$. Then

$$A\left(\frac{1}{\det(A)}\text{adj}(A)\right) = \left(\frac{1}{\det(A)}\text{adj}(A)\right)A = I.$$

Thus $\frac{1}{\det(A)}\text{adj}(A)$ is the inverse of A and hence A is invertible. □

Definition 1.2.12 (Trace)**.** *Let $A = [a_{ij}]$ be an $n \times n$ matrix. The* **trace** *of A is defined by*

$$\text{tr}(A) = a_{11} + a_{22} + \cdots + a_{nn}.$$

Proposition 1.2.13 (Properties of trace)**.** *Let A, B be $n \times n$ matrices and $k \in \mathbb{R}$. Then*

1. $\text{tr}(A + B) = \text{tr}(A) + \text{tr}(B)$
2. $\text{tr}(kA) = k\text{tr}(A)$
3. $\text{tr}(AB) = \text{tr}(BA)$

Proof. The first two properties are obvious. For the third one, let $A = [a_{ij}]$ and $B = [b_{ij}]$. Then

$$\begin{aligned}
\text{tr}(AB) &= \sum_{i=1}^{n}[AB]_{ii} \\
&= \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}b_{ji} \\
&= \sum_{j=1}^{n}\sum_{i=1}^{n} b_{ji}a_{ij} \\
&= \sum_{j=1}^{n}[BA]_{jj} \\
&= \text{tr}(BA)
\end{aligned}$$

□

1.3 Vectors

In mathematics, the term vector refers to an element in any vector space. In these notes, the only vector space we consider is the Euclidean space $\mathbb{R}^n$.

Definition 1.3.1 (Euclidean space)**.** *Let n be a positive integer. The n dimensional* **Euclidean space** *is the set*

$$\mathbb{R}^n = \{(x_1, x_2, \ldots, x_n) : x_i \in \mathbb{R} \textit{ for any } i = 1, 2, \ldots, n\}.$$

Definition 1.3.2 (Vector addition and scaler multiplication)**.**

1. **Vector addition***: Let* $\mathbf{u} = (u_1, u_2, \ldots, u_n), \mathbf{v} = (v_1, v_2, \ldots, v_n) \in \mathbb{R}^n$. *Define*
$$\mathbf{u} + \mathbf{v} = (u_1 + v_1, u_2 + v_2, \ldots, u_n + v_n).$$

2. **Scalar multiplication***: Let* $\mathbf{v} = (v_1, v_2, \ldots, v_n) \in \mathbb{R}^n$ *and* $\alpha \in \mathbb{R}$. *Define*
$$\alpha\mathbf{v} = (\alpha v_1, \alpha v_2, \ldots, \alpha v_n).$$

Next we define scalar product on $\mathbb{R}^n$ which will be used to define distance between two points and angle between two vectors in $\mathbb{R}^n$.

Definition 1.3.3 (Scalar product)**.** *Let* $\mathbf{u} = (u_1, u_2, \ldots, u_n)$,$\mathbf{v}= (v_1, v_2,\ldots,v_n)\in \mathbb{R}^n$. *The* **scalar product***, or* **dot product***, of* $\mathbf{u}$ *and* $\mathbf{v}$ *is defined by*

$$\langle \mathbf{u}, \mathbf{v} \rangle = u_1v_1 + u_2v_2 + \cdots + u_nv_n.$$

Note that the scalar product of two vectors is a number, not a vector.

Proposition 1.3.4 (Properties of scalar product)**.** *Let* $\mathbf{u}, \mathbf{v}, \mathbf{w} \in \mathbb{R}^n$ *and* $\alpha, \beta \in \mathbb{R}$. *Then*

1. *(Bilinear):*
$$\langle \alpha\mathbf{u} + \beta\mathbf{v}, \mathbf{w} \rangle = \alpha\langle \mathbf{u}, \mathbf{w} \rangle + \beta\langle \mathbf{v}, \mathbf{w} \rangle$$

2. *(Symmetric):*
$$\langle \mathbf{v}, \mathbf{u} \rangle = \langle \mathbf{u}, \mathbf{v} \rangle$$

3. *(Positive definite):*
$$\langle \mathbf{v}, \mathbf{v} \rangle \geq 0$$
with equality holds if and only if $\mathbf{v} = \mathbf{0}$.

The following proposition is simple but has many important applications.

Proposition 1.3.5. *Let* $\mathbf{v} \in \mathbb{R}^n$ *be a vector. Suppose*

$$\langle \mathbf{u}, \mathbf{v} \rangle = 0 \text{ for any } \mathbf{u} \in \mathbb{R}^n.$$

Then $\mathbf{v} = \mathbf{0}$.

Proof. Take $\mathbf{u} = \mathbf{v}$. Then we have $\langle \mathbf{v}, \mathbf{v} \rangle = 0$. Therefore $\mathbf{v} = \mathbf{0}$. □

We may use scalar product to define norm which may be consider as length of vectors.

Definition 1.3.6 (Norm)**.** *Let* $\mathbf{v} = (v_1, v_2, \ldots, v_n) \in \mathbb{R}^n$. *The* **norm** *of* $\mathbf{v}$ *is defined as*

$$\|\mathbf{v}\| = \sqrt{\langle \mathbf{v}, \mathbf{v} \rangle} = \sqrt{v_1^2 + v_2^2 + \cdots + v_n^2}.$$

Definition 1.3.7 (Unit vector)**.** *We say that* $\mathbf{v} \in \mathbb{R}^n$ *is a* **unit vector** *if* $\|\mathbf{v}\| = 1$.

Theorem 1.3.8 (Cauchy-Schwarz inequality)**.** *For any* $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$, *we have*

$$|\langle \mathbf{u}, \mathbf{v} \rangle| \leq \|\mathbf{u}\| \|\mathbf{v}\|$$

with equality holds if and only if $\mathbf{u} = \mathbf{0}$ *or* $\mathbf{v} = \alpha \mathbf{u}$ *for some real number* α.

Proof. Suppose $\mathbf{u} = \mathbf{0}$. Then the inequality holds obviously. Suppose $\mathbf{u} \neq \mathbf{0}$. Consider the scalar product

$$\langle t\mathbf{u} - \mathbf{v}, t\mathbf{u} - \mathbf{v} \rangle = t^2 \langle \mathbf{u}, \mathbf{u} \rangle - 2t \langle \mathbf{u}, \mathbf{v} \rangle + \langle \mathbf{v}, \mathbf{v} \rangle$$

which is a quadratic expression in t and is non-negative for any $t \in \mathbb{R}$. Therefore the discriminant satisfies

$$(2\langle \mathbf{u}, \mathbf{v} \rangle)^2 - 4\langle \mathbf{u}, \mathbf{u} \rangle \langle \mathbf{v}, \mathbf{v} \rangle \leq 0$$

which means

$$\langle \mathbf{u}, \mathbf{v} \rangle^2 \leq \langle \mathbf{u}, \mathbf{u} \rangle \langle \mathbf{v}, \mathbf{v} \rangle$$

Now equality holds if and only if there exists $\alpha \in \mathbb{R}$ such that $\|\alpha \mathbf{u} - \mathbf{v}\| = 0$ which means $\mathbf{v} = \alpha \mathbf{u}$. □

Cauchy-Schwarz inequality has two applications. The first one is triangle inequality.

Proposition 1.3.9 (Properties of norm)**.** *Let* $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$ *and* $\alpha \in \mathbb{R}$*.*

1. $\|\alpha\mathbf{v}\| = |\alpha|\|\mathbf{v}\|$
2. $\|\mathbf{v}\| \geq 0$ *with* $\|\mathbf{v}\| = 0$ *if and only if* $\mathbf{v} = \mathbf{0}$*.*
3. *(Triangle inequality):*

$$\|\mathbf{u} + \mathbf{v}\| \leq \|\mathbf{u}\| + \|\mathbf{v}\|$$

4. *(Parallelogram law):*

$$\langle \mathbf{u}, \mathbf{v} \rangle = \frac{1}{4}\left(\|\mathbf{u} + \mathbf{v}\|^2 - \|\mathbf{u} - \mathbf{v}\|^2\right)$$

Proof. The first two properties are obvious. We prove the triangle inequality and parallelogram law.

3. (Triangle inequality)

$$\begin{aligned} \|\mathbf{u} + \mathbf{v}\|^2 &= \langle \mathbf{u} + \mathbf{v}, \mathbf{u} + \mathbf{v} \rangle \\ &= \langle \mathbf{u}, \mathbf{u} \rangle + 2\langle \mathbf{u}, \mathbf{v} \rangle + \langle \mathbf{v}, \mathbf{v} \rangle \\ &\leq \|\mathbf{u}\|^2 + 2\|\mathbf{u}\|\|\mathbf{v}\| + \|\mathbf{v}\|^2 \text{ (Cauchy-Schwarz inequality)} \\ &= (\|\mathbf{u}\| + \|\mathbf{v}\|)^2 \end{aligned}$$

4. (Parallelogram law)

$$\begin{aligned} & \|\mathbf{u} + \mathbf{v}\|^2 - \|\mathbf{u} - \mathbf{v}\|^2 \\ = & \langle \mathbf{u} + \mathbf{v}, \mathbf{u} + \mathbf{v} \rangle - \langle \mathbf{u} - \mathbf{v}, \mathbf{u} - \mathbf{v} \rangle \\ = & (\|\mathbf{u}\|^2 + 2\langle \mathbf{u}, \mathbf{v} \rangle + \|\mathbf{v}\|^2) - (\|\mathbf{u}\|^2 - 2\langle \mathbf{u}, \mathbf{v} \rangle + \|\mathbf{v}\|^2) \\ = & 4\langle \mathbf{u}, \mathbf{v} \rangle \end{aligned}$$

□

The second application is that Cauchy-Schwarz inequality allows us to define angle between two vectors.

Definition 1.3.10 (Angle between two vectors)**.** *Let* $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$ *be two nonzero vectors. The* **angle** *between* $\mathbf{u}$ *and* $\mathbf{v}$ *is the unique* $\theta \in [0, \pi]$ *such that*

$$\cos\theta = \frac{\langle \mathbf{u}, \mathbf{v} \rangle}{\|\mathbf{u}\|\|\mathbf{v}\|}$$

Note that the above definition makes sense, that is, there exists $\theta \in [0, \pi]$ such that $\cos\theta = \frac{\langle \mathbf{u}, \mathbf{v} \rangle}{\|\mathbf{u}\|\|\mathbf{v}\|}$ because $\left| \frac{\langle \mathbf{u}, \mathbf{v} \rangle}{\|\mathbf{u}\|\|\mathbf{v}\|} \right| \leq 1$ for any nonzero vectors $\mathbf{u}, \mathbf{v}$ by Cauchy-Schwarz inequality.

Definition 1.3.11 (Orthogonal vectors)**.** *Let* $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$ *be two vectors. We say that* $\mathbf{u}$ *and* $\mathbf{v}$ *are* **orthogonal** *and write* $\mathbf{u} \perp \mathbf{v}$ *if*

$$\langle \mathbf{u}, \mathbf{v} \rangle = 0$$

Next we introduce a second kind of product which is defined only in $\mathbb{R}^3$.

Definition 1.3.12 (Cross product)**.** *Let* $\mathbf{u} = (u_1, u_2, u_3)$,$\mathbf{v}= (v_1, v_2, v_3)$$\in \mathbb{R}^3$ *be two vectors in* $\mathbb{R}^3$*. The* **cross product***, or* **vector product***, of* $\mathbf{u}$ *and* $\mathbf{v}$ *is defined by*

$$\begin{aligned} \mathbf{u} \times \mathbf{v} = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ u_1 & u_2 & u_3 \\ v_1 & v_2 & v_3 \end{vmatrix} &= \begin{vmatrix} u_2 & u_3 \\ v_2 & v_3 \end{vmatrix} \mathbf{i} - \begin{vmatrix} u_1 & u_3 \\ v_1 & v_3 \end{vmatrix} \mathbf{j} + \begin{vmatrix} u_1 & u_2 \\ v_1 & v_2 \end{vmatrix} \mathbf{k} \\ &= (u_2v_3 - u_3v_2, u_3v_1 - u_1v_3, u_1v_2 - u_2v_1) \end{aligned}$$

where $\mathbf{i} = (1, 0, 0), \mathbf{j} = (0, 1, 0), \mathbf{k} = (0, 0, 1)$.

Proposition 1.3.13 (Properties of cross product)**.**

1. $\mathbf{i} \times \mathbf{j} = \mathbf{k}$*,* $\mathbf{j} \times \mathbf{k} = \mathbf{i}$*,* $\mathbf{k} \times \mathbf{i} = \mathbf{j}$

2. *(Bilinear) For any* $\alpha, \beta \in \mathbb{R}$ *and* $\mathbf{u}, \mathbf{v}, \mathbf{w} \in \mathbb{R}^3$*,*

$$(\alpha\mathbf{u} + \beta\mathbf{v}) \times \mathbf{w} = \alpha\mathbf{u} \times \mathbf{w} + \beta\mathbf{v} \times \mathbf{w}$$

3. *(Anti-symmetric) For any* $\mathbf{u}, \mathbf{v} \in \mathbb{R}^3$*,*

$$\mathbf{v} \times \mathbf{u} = -\mathbf{u} \times \mathbf{v}$$

4. *For any* $\mathbf{u}, \mathbf{v} \in \mathbb{R}^3$*, we have* $\mathbf{u} \times \mathbf{v} \perp \mathbf{u}$ *and* $\mathbf{u} \times \mathbf{v} \perp \mathbf{v}$*, that is*

$$\langle \mathbf{u} \times \mathbf{v}, \mathbf{u} \rangle = \langle \mathbf{u} \times \mathbf{v}, \mathbf{v} \rangle = 0$$

5. *For any* $\mathbf{u}, \mathbf{v} \in \mathbb{R}^3$,
$$\mathbf{u} \times \mathbf{v} = \|\mathbf{u}\|\|\mathbf{v}\| \sin\theta\, \mathbf{n}$$
where θ *is the angle between* $\mathbf{u}$ *and* $\mathbf{v}$*, and* $\mathbf{n}$ *is the unit vector normal to the plane spanned by* $\mathbf{u}$ *and* $\mathbf{v}$ *with direction determined by the right hand rule. In other words,*

   (a) $\|\mathbf{u} \times \mathbf{v}\|$ *is equal to the area of the parallelogram spanned by* $\mathbf{u}$, $\mathbf{v}$.

   (b) $\mathbf{u} \times \mathbf{v}$ *is normal to the plane spanned by* $\mathbf{u}$ *and* $\mathbf{v}$ *with direction determined by the right hand rule.*

6. *(Jacobi identity) For any* $\mathbf{u}, \mathbf{v}, \mathbf{w} \in \mathbb{R}^3$,
$$\mathbf{u} \times (\mathbf{v} \times \mathbf{w}) + \mathbf{v} \times (\mathbf{w} \times \mathbf{u}) + \mathbf{w} \times (\mathbf{u} \times \mathbf{v}) = \mathbf{0}$$

7. *For any* $\mathbf{u}, \mathbf{v}, \mathbf{w} \in \mathbb{R}^3$,
$$\mathbf{u} \times (\mathbf{v} \times \mathbf{w}) = \langle \mathbf{u}, \mathbf{w} \rangle \mathbf{v} - \langle \mathbf{u}, \mathbf{v} \rangle \mathbf{w}$$

Next we define a product which involves three vectors.

Definition 1.3.14 (Scalar triple product)**.** *Let* $\mathbf{u} = (u_1, u_2, u_3), \mathbf{v} = (v_1, v_2, v_3), \mathbf{w} = (w_1, w_2, w_3) \in \mathbb{R}^3$. *The* **scalar triple product** *of* $\mathbf{u}, \mathbf{v}, \mathbf{w}$ *is defined by*
$$\langle \mathbf{u}, \mathbf{v} \times \mathbf{w} \rangle = \begin{vmatrix} u_1 & u_2 & u_3 \\ v_1 & v_2 & v_3 \\ w_1 & w_2 & w_3 \end{vmatrix}.$$

The value $|\langle \mathbf{u}, \mathbf{v} \times \mathbf{w} \rangle|$ is equal to the volume of the parallelepiped spanned by $\mathbf{u}, \mathbf{v}, \mathbf{w}$. The sign of $\langle \mathbf{u}, \mathbf{v} \times \mathbf{w} \rangle$ depends on the orientation of $\mathbf{u}, \mathbf{v}, \mathbf{w}$. It is positive if $\mathbf{u}, \mathbf{v}, \mathbf{w}$ are in right hand orientation and otherwise negative.

Proposition 1.3.15 (Properties of scalar triple product)**.** *Scalar triple product has the following properties.*

1. *Multi-linear*

2. *Anti-symmetric*

3. $\langle \mathbf{i}, \mathbf{j} \times \mathbf{k} \rangle = 1$

Proposition 1.3.16 (Cyclic property of scalar triple product)**.** *Let* $\mathbf{u}, \mathbf{v}, \mathbf{w} \in \mathbb{R}^3$ *be three vectors. We have*

$$\langle \mathbf{u}, \mathbf{v} \times \mathbf{w} \rangle = \langle \mathbf{v}, \mathbf{w} \times \mathbf{u} \rangle = \langle \mathbf{w}, \mathbf{u} \times \mathbf{v} \rangle$$

The following three identities are useful in studying curvature of surfaces.

Proposition 1.3.17.

1. *For any* $\mathbf{u}_1, \mathbf{v}_1, \mathbf{u}_2, \mathbf{v}_2 \in \mathbb{R}^3$,

$$\begin{vmatrix} \langle \mathbf{u}_1, \mathbf{u}_2 \rangle & \langle \mathbf{u}_1, \mathbf{v}_2 \rangle \\ \langle \mathbf{v}_1, \mathbf{u}_2 \rangle & \langle \mathbf{v}_1, \mathbf{v}_2 \rangle \end{vmatrix} = \langle \mathbf{u}_1 \times \mathbf{v}_1, \mathbf{u}_2 \times \mathbf{v}_2 \rangle$$

2. *For any* $\mathbf{u}, \mathbf{v} \in \mathbb{R}^3$,

$$\begin{vmatrix} \langle \mathbf{u}, \mathbf{u} \rangle & \langle \mathbf{u}, \mathbf{v} \rangle \\ \langle \mathbf{v}, \mathbf{u} \rangle & \langle \mathbf{v}, \mathbf{v} \rangle \end{vmatrix} = \|\mathbf{u} \times \mathbf{v}\|^2$$

3. *For any* $\mathbf{x}_{11}, \mathbf{x}_{12}, \mathbf{x}_{21}, \mathbf{x}_{22}, \mathbf{u}, \mathbf{v} \in \mathbb{R}^3$,

$$\begin{aligned} & \begin{vmatrix} \langle \mathbf{x}_{11}, \mathbf{u} \times \mathbf{v} \rangle & \langle \mathbf{x}_{12}, \mathbf{u} \times \mathbf{v} \rangle \\ \langle \mathbf{x}_{21}, \mathbf{u} \times \mathbf{v} \rangle & \langle \mathbf{x}_{22}, \mathbf{u} \times \mathbf{v} \rangle \end{vmatrix} \\ = & \begin{vmatrix} \langle \mathbf{x}_{11}, \mathbf{x}_{22} \rangle - \langle \mathbf{x}_{12}, \mathbf{x}_{21} \rangle & \langle \mathbf{x}_{11}, \mathbf{u} \rangle & \langle \mathbf{x}_{11}, \mathbf{v} \rangle \\ \langle \mathbf{x}_{22}, \mathbf{u} \rangle & \langle \mathbf{u}, \mathbf{u} \rangle & \langle \mathbf{u}, \mathbf{v} \rangle \\ \langle \mathbf{x}_{22}, \mathbf{v} \rangle & \langle \mathbf{v}, \mathbf{u} \rangle & \langle \mathbf{v}, \mathbf{v} \rangle \end{vmatrix} \\ & - \begin{vmatrix} 0 & \langle \mathbf{x}_{12}, \mathbf{u} \rangle & \langle \mathbf{x}_{12}, \mathbf{v} \rangle \\ \langle \mathbf{x}_{21}, \mathbf{u} \rangle & \langle \mathbf{u}, \mathbf{u} \rangle & \langle \mathbf{u}, \mathbf{v} \rangle \\ \langle \mathbf{x}_{21}, \mathbf{v} \rangle & \langle \mathbf{v}, \mathbf{u} \rangle & \langle \mathbf{v}, \mathbf{v} \rangle \end{vmatrix} \end{aligned}$$

We turn our discussion to bases for vector subspaces of $\mathbb{R}^m$.

Definition 1.3.18 (Vector subspace)**.** *We say that a subset* $V \subset \mathbb{R}^m$ *is a* **vector subspace** *of* $\mathbb{R}^m$ *if* V *contains the zero vector* $\mathbf{0}$ *and for any* $\mathbf{u}, \mathbf{v} \in V$, $\alpha, \beta \in \mathbb{R}$, *we have*

$$\alpha \mathbf{u} + \beta \mathbf{v} \in V.$$

In other words, $V \subset \mathbb{R}^m$ is a vector subspace if V contains the zero vector $\mathbf{0}$ and V is closed under addition and scalar multiplication. The set $\{\mathbf{0}\}$ contains only the zero vector is a subset of $\mathbb{R}^m$ which is called the **trivial**

subspace. The whole set $\mathbb{R}^m$ is also a subset of $\mathbb{R}^m$. A vector subspace of $\mathbb{R}^3$ is either the trivial subspace $\{\mathbf{0}\}$, a line passing through $\mathbf{0}$, a plane containing $\mathbf{0}$ or the whole $\mathbb{R}^3$.

Now we introduce the notions of linear independency and spanning set.

Definition 1.3.19 (Linearly independent vectors and spanning set)**.** *Let $V \subset \mathbb{R}^m$ be a vector subspace and $E = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k\} \subset V$ be a set of vectors in V.*

1. *We say that E is* **linearly independent** *if*

$$c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_k\mathbf{v}_k = \mathbf{0}$$

implies $c_1 = c_2 = \cdots = c_k = 0$.

2. *We say that E* **spans** *V if for any $\mathbf{v} \in V$, there exists scalars $\alpha_1, \alpha_2, \ldots, \alpha_k \in \mathbb{R}$ such that*

$$\mathbf{v} = \alpha_1\mathbf{v}_1 + \alpha_2\mathbf{v}_2 + \cdots + \alpha_k\mathbf{v}_k.$$

A set E of vectors in V is linearly independent if the zero vector $\mathbf{0}$ can not be written as a linearly combination of vectors in E in a nontrivial way, meaning that not all coefficients are zero. We say that $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k$ are **linearly dependent** if they are not linearly independent. In other words, $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k$ are linearly dependent if there exists scalars $c_1, c_2, \ldots, c_k$ not all zero such that

$$c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_k\mathbf{v}_k = \mathbf{0}.$$

Proposition 1.3.20. *Let $V \subset \mathbb{R}^m$ be a vector subspace and $E = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k\} \subset V$ be a set of vectors in V. Then E is linearly dependent if and only if there exists $\mathbf{v}_i \in E$ which can be written as a linear combination of other vectors in E, that is,*

$$\mathbf{v}_i = \alpha_1\mathbf{v}_1 + \cdots + \alpha_{i-1}\mathbf{v}_{i-1} + \alpha_{i+1}\mathbf{v}_{i+1} + \cdots + \alpha_k\mathbf{v}_k$$

for some $\alpha_1, \ldots, \alpha_{i-1}, \alpha_{i+1}, \ldots, \alpha_k \in \mathbb{R}$.

Proof. Suppose E is linearly dependent. Then there exists $c_1, \ldots, c_k$, not all zero, such that

$$c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_k\mathbf{v}_k = \mathbf{0}.$$

Let $1 \leq i \leq k$ be such that $c_i \neq 0$. Then

$$\mathbf{v}_i = -\frac{c_1}{c_i}\mathbf{v}_1 - \cdots - \frac{c_{i-1}}{c_i}\mathbf{v}_{i-1} - \frac{c_{i+1}}{c_i}\mathbf{v}_{i+1} - \cdots - \frac{c_k}{c_i}\mathbf{v}_k$$

is a linear combination of other vectors in E.

Suppose there exists $\mathbf{v}_i \in E$ such that

$$\mathbf{v}_i = \alpha_1\mathbf{v}_1 + \cdots + \alpha_{i-1}\mathbf{v}_{i-1} + \alpha_{i+1}\mathbf{v}_{i+1} + \cdots + \alpha_k\mathbf{v}_k$$

for some $\alpha_1, \ldots, \alpha_{i-1}, \alpha_{i+1}, \ldots, \alpha_k \in \mathbb{R}$. Then

$$\alpha_1\mathbf{v}_1 + \cdots + \alpha_{i-1}\mathbf{v}_{i-1} - \mathbf{v}_i + \alpha_{i+1}\mathbf{v}_{i+1} + \cdots + \alpha_k\mathbf{v}_k = \mathbf{0}$$

and the coefficient of $\mathbf{v}_i$ is -1 which is nonzero. Therefore E is linearly dependent. □

The above proposition implies in particular that a set E of vectors in V is linearly dependent if there exists distinct vectors $\mathbf{u}, \mathbf{v} \in E$ such that $\mathbf{v} = \alpha\mathbf{u}$ for some $\alpha \in \mathbb{R}$. Furthermore if $\mathbf{0} \in E$, then E is linearly dependent.

Proposition 1.3.21. *Let $V \subset \mathbb{R}^m$ be a vector subspace and $E \subset V$ be a set of vectors in V. Suppose the vectors in E are*

1. *mutually orthogonal, that is, $\langle \mathbf{u}, \mathbf{v} \rangle = 0$ for any distinct $\mathbf{u}, \mathbf{v} \in E$, and*
2. *nonzero, that is, $\mathbf{v} \neq \mathbf{0}$ for any $\mathbf{v} \in E$.*

Then E is linearly independent.

Proof. Let $E = \{\mathbf{v}_1, \ldots, \mathbf{v}_k\} \subset V$. Suppose

$$c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_k\mathbf{v}_k = \mathbf{0}.$$

For any $1 \leq i \leq k$, we have

$$\begin{aligned} c_1\langle \mathbf{v}_i, \mathbf{v}_i \rangle + c_2\langle \mathbf{v}_i, \mathbf{v}_2 \rangle + \cdots + c_k\langle \mathbf{v}_i, \mathbf{v}_k \rangle &= \langle \mathbf{v}_i, \mathbf{0} \rangle \\ c_i\langle \mathbf{v}_i, \mathbf{v}_i \rangle &= 0 \end{aligned}$$

which implies $c_i = 0$ since $\mathbf{v}_i \neq \mathbf{0}$. Thus $c_i = 0$ for any $1 \leq i \leq k$. Therefore E is linearly independent. □

Proposition 1.3.22. *Let $V \subset \mathbb{R}^m$ be a vector subspace and*

$$E = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k\} \subset V$$

be a set of vectors in V.

1. *Suppose E is linearly independent. Then*
   - *(i) any subset of E is linearly independent.*
   - *(ii) if E does not span V, then there exists $\mathbf{v}_{k+1} \in V$ such that*

     $$F = E \cup \{\mathbf{v}_{k+1}\} = \{\mathbf{v}_1, \ldots, \mathbf{v}_k, \mathbf{v}_{k+1}\}$$

     is linearly independent.
2. *Suppose E spans V. Then*
   - *(i) any set of vectors which contains E spans V.*
   - *(ii) if E is not linearly independent, then there exists $\mathbf{v}_i \in E$ such that*

     $$D = E \backslash \{\mathbf{v}_i\} = \{\mathbf{v}_1, \ldots, \mathbf{v}_{i-1}, \mathbf{v}_{i+1}, \ldots, \mathbf{v}_k\}$$

     spans V.

Proof. 1. Suppose E is linearly independent.

(i) Let $D \subset E$. We may assume $D = \{\mathbf{v}_1, \ldots, \mathbf{v}_r\}$ where $r \leq k$. Suppose

$$c_1\mathbf{v}_1 + \cdots + c_r\mathbf{v}_r = \mathbf{0}.$$

Then

$$c_1\mathbf{v}_1 + \cdots + c_r\mathbf{v}_r + 0\mathbf{v}_{r+1} + \cdots + 0\mathbf{v}_k = \mathbf{0}.$$

Since E is linearly independent, we have $c_1 = c_2 = \cdots = c_r = 0$. Therefore D is linearly independent.

(ii) If E does not span V, then there exists $\mathbf{v}_{k+1} \in V$ which is not a linear combination of vectors in E. Suppose

$$c_1\mathbf{v}_1 + \cdots + c_k\mathbf{v}_k + c_{k+1}\mathbf{v}_{k+1} = \mathbf{0}.$$

Then $c_{k+1} = 0$ for otherwise

$$\mathbf{v}_{k+1} = -\frac{c_1}{c_{k+1}}\mathbf{v}_1 - \cdots - \frac{c_k}{c_{k+1}}\mathbf{v}_k$$

is a linear combination of vectors in E which contradicts the choice of $\mathbf{v}_{k+1}$. Thus

$$c_1\mathbf{v}_1 + \cdots + c_k\mathbf{v}_k = \mathbf{0}$$

which implies $c_1 = \cdots = c_k = 0$. Therefore $F = E \cup \{\mathbf{v}_{k+1}\}$ is linearly independent.

2. Suppose E spans V.

   (i) Let $F \subset V$ be a set of vectors with $E \subset F$. For any $\mathbf{v} \in V$, $\mathbf{v}$ is a linear combination of vectors in F because $E \subset F$ and E spans V. Thus F spans V.

   (ii) If E is not linearly independent, then by Proposition 1.3.20, there exists $\mathbf{v}_i \in E$ such

$$\mathbf{v}_i = \alpha_1\mathbf{v}_1 + \cdots + \alpha_{i-1}\mathbf{v}_{i-1} + \alpha_{i+1}\mathbf{v}_{i+1} + \cdots + \alpha_k\mathbf{v}_k.$$

   Now for any $\mathbf{v} \in V$, since E spans V, there exists $\beta_1, \ldots, \beta_k \in \mathbb{R}$ such that

$$\begin{aligned}
\mathbf{v} &= \beta_1\mathbf{v}_1 + \cdots + \beta_i\mathbf{v}_i + \cdots + \beta_k\mathbf{v}_k \\
&= \beta_1\mathbf{v}_1 + \cdots + \beta_{i-1}\mathbf{v}_{i-1} \\
&\quad +\beta_i(\alpha_1\mathbf{v}_1 + \cdots + \alpha_{i-1}\mathbf{v}_{i-1} + \alpha_{i+1}\mathbf{v}_{i+1} + \cdots + \alpha_k\mathbf{v}_k) \\
&\quad +\beta_{i+1}\mathbf{v}_{i+1} + \cdots + \beta_k\mathbf{v}_k \\
&= (\beta_1 + \beta_i\alpha_1)\mathbf{v}_1 + \cdots + (\beta_{i-1} + \beta_i\alpha_{i-1})\mathbf{v}_{i-1} \\
&\quad +(\beta_{i+1} + \beta_i\alpha_{i+1})\mathbf{v}_{i+1} + \cdots + (\beta_k + \beta_i\alpha_k)\mathbf{v}_k
\end{aligned}$$

   Therefore any vectors in V can be expressed as a linear combination of vectors in $D = E \backslash \{\mathbf{v}_i\}$ which means D spans V.

□

Definition 1.3.23 (Basis). *Let $V \subset \mathbb{R}^m$ be a vector subspace and $E = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\} \subset V$ be a set of vectors in V. We say that E constitutes a* **basis** *for V if*

1. *E is linearly independent, and*

2. *E spans V.*

Example 1.3.24 (Standard basis)**.** *The set* $B = \{\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n\}$ *where*

$$\begin{aligned} \mathbf{e}_1 &= (1, 0, 0, \ldots, 0, 0) \\ \mathbf{e}_2 &= (0, 1, 0, \ldots, 0, 0) \\ &\vdots \\ \mathbf{e}_n &= (0, 0, 0, \ldots, 0, 1) \end{aligned}$$

constitutes a basis for $\mathbb{R}^m$ *and is called the* **standard basis***.*

Theorem 1.3.25. *Let* $V \subset \mathbb{R}^m$ *be a vector subspace and* $E = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\} \subset V$ *be a set of vectors in* V*. Then the following conditions are equivalent.*

1. E *constitutes a basis for* V*.*
2. *For any* $\mathbf{v} \in V$*, there exists unique* $\alpha_1, \alpha_2, \cdots, \alpha_n \in \mathbb{R}$ *such that*

$$\mathbf{v} = \alpha_1 \mathbf{v}_1 + \alpha_2 \mathbf{v}_2 + \cdots + \alpha_n \mathbf{v}_n.$$

Proof. Suppose E constitutes a basis for V. For any $\mathbf{v} \in V$, since E spans V, there exists $\alpha_1, \ldots, \alpha_n$ such that

$$\mathbf{v} = \alpha_1 \mathbf{v}_1 + \alpha_2 \mathbf{v}_2 + \cdots + \alpha_n \mathbf{v}_n.$$

To prove that such coefficients $\alpha_1, \ldots, \alpha_n$ are unique, suppose $\beta_1, \ldots, \beta_n \in \mathbb{R}$ are scalars such that

$$\mathbf{v} = \beta_1 \mathbf{v}_1 + \beta_2 \mathbf{v}_2 + \cdots + \beta_n \mathbf{v}_n.$$

By considering the difference of the two equalities, we have

$$(\alpha_1 - \beta_1)\mathbf{v}_1 + (\alpha_2 - \beta_2)\mathbf{v}_2 + \cdots + (\alpha_n - \beta_n)\mathbf{v}_n = \mathbf{0}.$$

Since E is linearly independent, we must have

$$\alpha_1 - \beta_1 = \alpha_2 - \beta_2 = \cdots = \alpha_n - \beta_n = 0$$

which means the expression of $\mathbf{v}$ as a linear combination of $\mathbf{v}_1, \ldots, \mathbf{v}_n$ is unique.

Suppose for any $\mathbf{v} \in V$, there exists unique $\alpha_1, \cdots, \alpha_n \in \mathbb{R}$ such that

$$\mathbf{v} = \alpha_1 \mathbf{v}_1 + \alpha_2 \mathbf{v}_2 + \cdots + \alpha_n \mathbf{v}_n.$$

It is obvious that E spans V. To prove that E is linearly independent, suppose

$$c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_n\mathbf{v}_n = \mathbf{0}.$$

Since

$$\mathbf{0} = 0\mathbf{v}_1 + 0\mathbf{v}_2 + \cdots + 0\mathbf{v}_n,$$

we have $c_1 = c_2 = \cdots = c_n = 0$ by uniqueness of expression of $\mathbf{0}$ as a linearly combination of E. Thus E is linearly independent. Therefore B constitutes a basis for V. □

The following proposition says that the number of vectors in any set of linearly independent vectors in a vector subspaces V is always less than or equal to the number of vectors in a spanning set for V.

Proposition 1.3.26. *Let $V \subset \mathbb{R}^m$ be a vector subspace. Suppose $E = \{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_r\} \subset V$ spans V and $F = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_s\} \subset V$ is linearly independent. Then $r \geq s$.*

Proof. Suppose $E = \{\mathbf{u}_1, \ldots, \mathbf{u}_r\}$ spans V and $F = \{\mathbf{v}_1, \ldots, \mathbf{v}_s\}$ be any set of s vectors in V. Suppose $r < s$. It suffices to prove that F must be linearly dependent. Since E spans V, any vector $\mathbf{v}_j \in S \subset V$ is a linear combination of vectors in E and we may write

$$\mathbf{v}_j = a_{1j}\mathbf{u}_1 + a_{2j}\mathbf{u}_2 + \cdots + a_{rj}\mathbf{u}_r$$

for some $a_{1j}, \cdots, a_{rj} \in \mathbb{R}$. Collect the above equalities and write them in matrix form as

$$\begin{pmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_s \end{pmatrix} = \begin{pmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_r \end{pmatrix} \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1s} \\ a_{21} & a_{22} & \cdots & a_{2s} \\ \vdots & \vdots & \ddots & \vdots \\ a_{r1} & a_{r2} & \cdots & a_{rs} \end{pmatrix}.$$

By Proposition 1.1.11, there exists $c_1, \ldots, c_s \in \mathbb{R}$, not all zero, such that

$$\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1s} \\ a_{21} & a_{22} & \cdots & a_{2s} \\ \vdots & \vdots & \ddots & \vdots \\ a_{r1} & a_{r2} & \cdots & a_{rs} \end{pmatrix} \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_s \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}.$$

Then

$$
\begin{aligned}
& c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_s\mathbf{v}_s \\
= \; & \begin{pmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_s \end{pmatrix} \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_s \end{pmatrix} \\
= \; & \begin{pmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_r \end{pmatrix} \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1s} \\ a_{21} & a_{22} & \cdots & a_{2s} \\ \vdots & \vdots & \ddots & \vdots \\ a_{r1} & a_{r2} & \cdots & a_{rs} \end{pmatrix} \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_s \end{pmatrix} \\
= \; & \begin{pmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_r \end{pmatrix} \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \\
= \; & \mathbf{0}
\end{aligned}
$$

and at least one of $c_1, \ldots, c_s$ is nonzero. Therefore F is linearly dependent and the proof of the proposition is complete. □

The above proposition has an important consequence that any two bases for V contain the same number of vectors.

Theorem 1.3.27. *Let $V \subset \mathbb{R}^m$ be a vector subspace. Suppose $E = \{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_r\} \subset V$ and $F = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_s\} \subset V$ are two bases for V. Then $r = s$.*

Proof. Since E spans V and F is linearly independent, we have $r \geq s$ by Proposition 1.3.26. Since F spans V and E is linearly independent, we have $s \geq r$ again by Proposition 1.3.26. Therefore we have $r = s$. □

This allows us to define the dimension of V.

Definition 1.3.28 (Dimension)**.** *Let V be a vector subspace of $\mathbb{R}^m$. The* **dimension** *of V is the number of vectors in a basis for V and is denoted by* $\dim(V)$.

Example 1.3.29.

1. *The trivial subspace* $V = \{\mathbf{0}\}$ *has dimension* $\dim(V) = 0$. *The empty set* $E = \emptyset$, *which contains zero vector, is a basis for* $V = \{\mathbf{0}\}$.

2. *The subspace* $V = \mathbb{R}^m$ *has dimension* $\dim(V) = m$. *The standard basis, which contains* n *vectors, is a basis for* $V = \mathbb{R}^m$.

In particular, a basis for $\mathbb{R}^m$ must contain exactly m vectors.

Theorem 1.3.30. *Let* $V \subset \mathbb{R}^m$ *be a vector subspace and* $F \subset V$ *be a set of vectors in* V. *Then*

1. F *is linearly independent if and only if* F *is contained in a basis for* V.

2. F *spans* V *if and only if* F *contains a basis for* V.

Proof. Suppose $\dim(V) = n$.

1. If $F \subset E$ is contained in a basis E for V, then F is linearly independent since E is linearly independent (Proposition 1.3.22).

   Conversely suppose F is linearly independent. If F spans V, then F is a basis for V and we are done. If F does not span V, then there exists $\mathbf{v} \notin F$ such that $F_1 = F \cup \{\mathbf{v}\}$ is linearly independent (Proposition 1.3.22). Repeat this process and get subsets $F \subset F_1 \subset F_2 \subset \cdots$. Since a set of linearly independent vectors in V contains at most $\dim(V) = n$ vectors (Proposition 1.3.26), the process stops in finitely many steps and we obtain a basis for V.

2. If F contains a basis E for V, then F spans V since E spans V (Proposition 1.3.22).

   Suppose F spans V. Let $E \subset F$ be a subset of F which is linearly independent. If E spans V, then E is a basis contained in F and we are done. If E does not span V, then there exists $\mathbf{v} \in F$ which is not a linear combination of vectors in E. Then $E_1 = E \cup \{\mathbf{v}\}$ is linearly independent. Repeat this process and get subsets $E \subset E_1 \subset E_2 \subset \cdots$. Since a set of linearly independent vectors in V contains at most $\dim(V) = n$ vectors (Proposition 1.3.26), the process stops in finitely many steps and we obtain a basis for V.

□

If the dimension of a vector subspace is known, it is easier to determine whether a set of vectors constitutes a basis.

Theorem 1.3.31. *Let $V \subset \mathbb{R}^m$ be a vector subspace with dimension $\dim(V) = n$ and $F = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\} \subset V$ be a set of n vectors in V. Then the following conditions are equivalent.*

1. *F constitutes a basis for V.*
2. *F is linearly independent.*
3. *F spans V.*

Proof. It suffices to prove that F spans V if and only if F is linearly independent.

Suppose F spans V. Then there exists a basis E, which contains n vectors since $\dim(V) = n$, such that $E \subset F$ (Theorem 1.3.30). Hence we must have $F = E$ since F contains n vectors by assumption. Therefore F constitutes a basis for V.

Suppose F is linearly independent. Then there exists a basis E, which contains n vectors since $\dim(V) = n$, such that $F \subset E$ (Theorem 1.3.30). Hence we must have $F = E$ since F contains n vectors by assumption. Therefore F constitutes a basis for V. □

Proposition 1.3.32. *Let $\mathbf{u}, \mathbf{v}, \mathbf{w} \in \mathbb{R}^3$ be three vectors in $\mathbb{R}^3$. The following conditions are equivalent.*

1. *$\mathbf{u}, \mathbf{v}, \mathbf{w}$ are linearly independent.*
2. *$\langle \mathbf{u}, \mathbf{v} \times \mathbf{w} \rangle \neq 0$*

We say that an $n \times n$ matrix A is **nonsingular** if it satisfies the equivalent conditions in the following theorem. We say that A is **singular** if it is not nonsingular, that is, $\det(A) = 0$.

Theorem 1.3.33. *The following conditions for $n \times n$ matrix A are equivalent.*

1. $\det(A) \neq 0$
2. *A is invertible, that is, the inverse A^{-1} of A exists.*
3. *For any n column vector $\mathbf{b}$, the equation $A\mathbf{x} = \mathbf{b}$ has a unique solution for $\mathbf{x}$.*

4. *The homogeneous equation* $A\mathbf{x} = \mathbf{0}$ *has no nontrivial solution, that is, solution for which* $\mathbf{x} \neq \mathbf{0}$.

5. *The column vectors of* A *constitute a basis for* $\mathbb{R}^m$.

Proof. (1)⇒(2). Proposition 1.2.11.
(2)⇒(3). Suppose A is invertible. Then $A\mathbf{x} = \mathbf{b}$ if and only if $\mathbf{x} = A^{-1}\mathbf{b}$. Therefore $A\mathbf{x} = \mathbf{b}$ has a unique solution $\mathbf{x} = A^{-1}\mathbf{b}$ for any $\mathbf{b} \in \mathbb{R}^m$.
(3)⇒(4). Obvious by taking $\mathbf{b} = \mathbf{0}$.
(4)⇒(5). Let $\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_n$ be the column vectors of A. If the homogeneous equation $A\mathbf{x} = \mathbf{0}$ has only trivial solution $\mathbf{x} = \mathbf{0}$, then

$$c_1\mathbf{a}_1 + c_2\mathbf{a}_2 + \cdots + c_n\mathbf{a}_n = \mathbf{0}$$

only when $c_1 = c_2 = \cdots = c_n = 0$. Thus $\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_n$ are linearly independent which implies $\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_n$ constitute a basis for $\mathbb{R}^m$ by Theorem 1.3.31.
(5)⇒(1). Theorem 1.3.31. □

In the last part of this section, we study vector valued function. Suppose $\mathbf{v}(t)$ for $t \in (a, b)$ is a **vector valued function**, that means, $\mathbf{v}$ is a function from open interval (a, b) to $\mathbb{R}^3$. We may write $\mathbf{v}(t) = (x(t), y(t), z(t))$ where $x(t), y(t), z(t)$ are ordinary real valued functions. Thus giving a vector valued function is the same as giving three real valued function. Similar to ordinary function, we say that $\mathbf{v}(t)$ is differentiable if the limit

$$\frac{d\mathbf{v}}{dt} = \lim_{\Delta t \to 0} \frac{\mathbf{v}(t + \Delta t) - \mathbf{v}(t)}{\Delta t}$$

exists and the limit is called the derivative of $\mathbf{v}(t)$ and is denoted by $\frac{d\mathbf{v}}{dt}$ or $\mathbf{v}'(t)$. It is not difficult to see that $\mathbf{v}(t)$ is differentiable if and only if all three functions $x(t), y(t), z(t))$ are differentiable. We have the following rules for derivative of vector valued functions which can be proved by the properties of derivatives of ordinary functions.

Proposition 1.3.34 (Rules for derivative of vector valued functions)**.** *Let* $\mathbf{u}(t), \mathbf{v}(t), \mathbf{w}(t)$ *be differentiable vector valued functions and* $\alpha(t)$ *be real valued function.*

1. $\dfrac{d}{dt}(\mathbf{u} + \mathbf{v}) = \dfrac{d\mathbf{u}}{dt} + \dfrac{d\mathbf{v}}{dt}$

2. $\dfrac{d}{dt}(\alpha\mathbf{v}) = \alpha\dfrac{d\mathbf{v}}{dt} + \dfrac{d\alpha}{dt}\mathbf{v}$

3. $\dfrac{d}{dt}\langle\mathbf{u},\mathbf{v}\rangle = \langle\dfrac{d\mathbf{u}}{dt},\mathbf{v}\rangle + \langle\mathbf{u},\dfrac{d\mathbf{v}}{dt}\rangle$

4. $\dfrac{d}{dt}(\mathbf{u}\times\mathbf{v}) = \dfrac{d\mathbf{u}}{dt}\times\mathbf{v} + \mathbf{u}\times\dfrac{d\mathbf{v}}{dt}$

5. $\dfrac{d}{dt}\langle\mathbf{u},\mathbf{v}\times\mathbf{w}\rangle = \langle\dfrac{d\mathbf{u}}{dt},\mathbf{v}\times\mathbf{w}\rangle + \langle\mathbf{u},\dfrac{d\mathbf{v}}{dt}\times\mathbf{w}\rangle + \langle\mathbf{u},\mathbf{v}\times\dfrac{d\mathbf{w}}{dt}\rangle$

The following lemma will be used from time to time in these notes and therefore we include the proof here.

Lemma 1.3.35. *Let* $\mathbf{u}(t)$ *and* $\mathbf{v}(t)$ *be two vector valued functions.*

1. *If* $\langle\mathbf{u}(t),\mathbf{v}(t)\rangle$ *is constant, then for any* t*, we have*

$$\langle\mathbf{u}'(t),\mathbf{v}(t)\rangle = -\langle\mathbf{u}(t),\mathbf{v}'(t)\rangle.$$

2. *If* $\|\mathbf{v}(t)\|$ *is constant, then for any* t*, we have*

$$\langle\mathbf{v}'(t),\mathbf{v}(t)\rangle = 0.$$

Proof. Differentiate $\langle\mathbf{u}(t),\mathbf{v}(t)\rangle = C$, where C is constant, with respect to t, we have

$$\langle\mathbf{u}'(t),\mathbf{v}(t)\rangle + \langle\mathbf{u}(t),\mathbf{v}'(t)\rangle = 0$$

and the first statement follows readily. The second statement is obtained by taking $\mathbf{u}(t) = \mathbf{v}(t)$. □

1.4 Orthogonal matrices and rigid transformations

An important interpretation of matrices is that they associate naturally with linear transformation.

Definition 1.4.1 (Linear transformation)**.** *A function* $L : \mathbb{R}^n \to \mathbb{R}^m$ *is called a* **linear transformation** *if for any* $\mathbf{u},\mathbf{v}\in\mathbb{R}^n$ *and* $\alpha,\beta\in\mathbb{R}$*, then*

$$L(\alpha\mathbf{u}+\beta\mathbf{v}) = \alpha L(\mathbf{u}) + \beta L(\mathbf{v})$$

Example 1.4.2 (Linear transformations associated with matrices)**.** *Let A be an $m \times n$ matrix. Define a function $L_A : \mathbb{R}^n \to \mathbb{R}^m$ by*

$$L_A(\mathbf{v}) = A\mathbf{v}$$

for $\mathbf{v} = (v_1, v_2, \ldots, v_n)^T$. Here we use the column vector notation where $\mathbf{v}$ is an n column vector and $A\mathbf{v}$ is an m column vector. Then L_A is a linear transformation which is called the linear transformation associated with A.

Conversely for any linear transformation $L : \mathbb{R}^n \to \mathbb{R}^m$, if we take

$$A = [L(\mathbf{e}_1), \ldots, L(\mathbf{e}_n)]$$

where $\mathbf{e}_1, \ldots, \mathbf{e}_n$ are n column vectors in the standard basis for $\mathbb{R}^n$, then $L_A = L$ where L_A is the linear transformation associated with A. Thus we have

Proposition 1.4.3 (Matrix representation of linear transformation)**.** *Let $L : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation. Then there is an $m \times n$ matrix A such that $L_A = L$ where L_A is the linearly transformation associated with A. The matrix A is called the* **matrix representation** *of L.*

Therefore there is a one-to-one correspondence between $m \times n$ matrices and linear transformations from $\mathbb{R}^n$ to $\mathbb{R}^m$. The following proposition says that the matrix multiplication associates with composition of linear transformations. This is one of the major reasons why matrix multiplication is defined in such a way.

Proposition 1.4.4. *Let A and an $k \times m$ matrix and B be an $m \times n$ matrix. Let L_A and L_B be the linear transformation associated with A and B respectively. Then the matrix representing $L_A \circ L_B$ is AB. In other words,*

$$L_{AB} = L_A \circ L_B.$$

Next we consider an important class of 3×3 matrices which correspond to rotation in $\mathbb{R}^3$.

Definition 1.4.5 (Orthogonal and special orthogonal matrix)**.** *Let Q be an $n \times n$ matrix.*

1. *We say that Q is an* **orthogonal matrix** *if $Q^{-1} = Q^T$ where Q^T is the transpose of Q.*

2. *We say that Q is a* **special orthogonal matrix** *if Q is an orthogonal matrix and* $\det(Q) = 1$.

Definition 1.4.6 (Orthonormal basis)**.** *Let $V \subset \mathbb{R}^m$ be a vector subspace. We say that a set $E = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ of vectors constitutes an* **orthonormal basis** *for V if they satisfy the following conditions.*

1. *E constitutes a basis for V.*

2. *E is mutually orthogonal, that is, $\langle \mathbf{v}_i, \mathbf{v}_j \rangle = 0$ whenever $i \neq j$.*

3. *E consists of unit vectors, that is, $\|\mathbf{v}_i\| = 1$ for $i = 1, 2, \ldots, n$.*

Proposition 1.4.7. *The following conditions for an $n \times n$ matrix Q are equivalent.*

1. *Q is an orthogonal matrix.*

2. *The column vectors of Q constitute an orthonormal basis for $\mathbb{R}^n$.*

3. *For any $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$,*
$$\langle Q\mathbf{u}, Q\mathbf{v} \rangle = \langle \mathbf{u}, \mathbf{v} \rangle$$

4. *For any $\mathbf{v} \in \mathbb{R}^n$,*
$$\|Q\mathbf{v}\| = \|\mathbf{v}\|$$

Note that if Q is an orthogonal matrix, then $\det(Q) = \pm 1$. If $\det(Q) = 1$, that is, Q is a special orthogonal matrix, then Q corresponds to a rotation in $\mathbb{R}^n$. If $\det(Q) = -1$, then Q corresponds to a reflection composites with a rotation in $\mathbb{R}^n$.

Proposition 1.4.8. *Suppose Q is a special 3×3 orthogonal matrix.*

1. *For any $\mathbf{u}, \mathbf{v} \in \mathbb{R}^3$, we have*
$$Q(\mathbf{u} \times \mathbf{v}) = Q\mathbf{u} \times Q\mathbf{v}.$$

2. *For any $\mathbf{u}, \mathbf{v}, \mathbf{w} \in \mathbb{R}^3$, we have*
$$\langle Q\mathbf{u}, Q\mathbf{v} \times Q\mathbf{w} \rangle = \langle \mathbf{u}, \mathbf{v} \times \mathbf{w} \rangle.$$

Proof. First we prove (2). For any $\mathbf{u}, \mathbf{v} \in \mathbb{R}^3$, we have

$$\begin{aligned} \langle Q\mathbf{u}, Q\mathbf{v} \times Q\mathbf{w} \rangle &= \det([Q\mathbf{u}, Q\mathbf{v}, Q\mathbf{w}]) \\ &= \det(Q[\mathbf{u}, \mathbf{v}, \mathbf{w}]) \\ &= \det(Q)\det([\mathbf{u}, \mathbf{v}, \mathbf{w}]) \\ &= \det([\mathbf{u}, \mathbf{v}, \mathbf{w}]) \\ &= \langle \mathbf{u}, \mathbf{v} \times \mathbf{w} \rangle. \end{aligned}$$

Now we use (2) to prove (1). Let $\mathbf{u}, \mathbf{v} \in \mathbb{R}^3$. For any $\mathbf{w} \in \mathbb{R}^3$, we have

$$\begin{aligned} \langle \mathbf{w}, Q(\mathbf{u} \times \mathbf{v}) \rangle &= \langle QQ^{-1}\mathbf{w}, Q(\mathbf{u} \times \mathbf{v}) \rangle \\ &= \langle Q^{-1}\mathbf{w}, \mathbf{u} \times \mathbf{v} \rangle \text{ (Proposition 1.4.7)} \\ &= \langle QQ^{-1}\mathbf{w}, Q\mathbf{u} \times Q\mathbf{v} \rangle \text{ (by (2))} \\ &= \langle \mathbf{w}, Q\mathbf{u} \times Q\mathbf{v} \rangle \end{aligned}$$

Therefore $Q(\mathbf{u} \times \mathbf{v}) = Q\mathbf{u} \times Q\mathbf{v}$ by Proposition 1.3.5. □

Definition 1.4.9 (Rigid transformation)**.** *A* **rigid transformation** *of* $\mathbb{R}^n$ *is a function* $T : \mathbb{R}^n \to \mathbb{R}^n$ *of the form*

$$T(\mathbf{v}) = Q\mathbf{v} + \mathbf{a}$$

for some $n \times n$ *orthogonal matrix* Q *and constant vector* $\mathbf{a} \in \mathbb{R}^n$*. If furthermore* $\det(Q) = 1$*, that is,* Q *is a special orthogonal matrix, we say that* T *is orientation preserving. If* $\det(Q) = -1$*, we say that* T *is orientation reversing.*

Proposition 1.4.10. *The following conditions for a function* $T : \mathbb{R}^3 \to \mathbb{R}^3$ *are equivalent.*

1. *T is a rigid transformation.*

2. *T preserves distance between two points, that is, for any* $\mathbf{u}, \mathbf{v} \in \mathbb{R}^3$*,*

   $$\|T(\mathbf{u}) - T(\mathbf{v})\| = \|\mathbf{u} - \mathbf{v}\|$$

3. *T is a composition of a rotation, and/or a translation, and/or a reflection.*

1.5 Eigenvalues, eigenvectors and diagonalization

Eigenvalues and eigenvectors are important in many aspects in linear algebra.

Definition 1.5.1 (Eigenvalues and eigenvectors)**.** *Let A be an $n \times n$ matrix. If λ is a complex number*[3] *and $\boldsymbol{\xi}$ is a non-zero*[4] *complex vector such that*

$$A\boldsymbol{\xi} = \lambda\boldsymbol{\xi},$$

then we say that λ is an **eigenvalue** *of A and $\boldsymbol{\xi}$ is an* **eigenvector** *of A associated with λ.*

To find eigenvalues of a matrix, we need to solve the characteristic equation.

Definition 1.5.2 (Characteristic polynomial and characteristic equation)**.** *Let A be an $n \times n$ matrix. The* **characteristic polynomial** *of A is the degree n polynomial $\det(xI - A)$ in x, where I is the identity matrix. The* **characteristic equation** *of A is the degree n polynomial equation*

$$\det(xI - A) = 0.$$

Note that the equality $A\boldsymbol{\xi} = \lambda\boldsymbol{\xi}$ is equivalent to $(\lambda I - A)\boldsymbol{\xi} = \mathbf{0}$. Now λ is an eigenvalue of A if and only if there exists nonzero vector $\boldsymbol{\xi}$ such that $(\lambda I - A)\boldsymbol{\xi} = \mathbf{0}$ which is equivalent to $\det(\lambda I - A) = 0$. To find an eigenvector associated with the eigenvalue λ, one needs to find $\boldsymbol{\xi} \neq \mathbf{0}$ such that $(\lambda I - A)\boldsymbol{\xi} = \mathbf{0}$.

Proposition 1.5.3. *Let A be an $n \times n$ matrix.*

1. *A complex number λ is an eigenvalue of A if and only if λ is a root to the characteristic equation $\det(xI - A) = 0$.*

2. *Let λ be an eigenvalue of A. Then $\boldsymbol{\xi}$ is an eigenvector of A associated with λ if and only if $\boldsymbol{\xi} \neq \mathbf{0}$ and $(\lambda I - A)\boldsymbol{\xi} = \mathbf{0}$.*

Note that a polynomial equation of degree n has at least one complex root and at most n distinct root. We have

[3]The set of complex numbers is $\mathbb{C} = \{a + bi : a, b \in \mathbb{R}\}$ where $i^2 = -1$. Note that a real number is also a complex number. Even when the matrix is real, we would consider complex eigenvalues and eigenvectors.

[4]Eigenvalue of a matrix may be 0 but eigenvector is by definition a non-zero vector.

Proposition 1.5.4. *An $n \times n$ matrix has at least one eigenvalue and has at most n distinct eigenvalues.*

Next we discuss diagonalization of matrices.

Definition 1.5.5. *Let A and B be $n \times n$ matrices. We say that A and B are* **similar** *if there exists invertible matrix P such that*

$$P^{-1}AP = B.$$

Proposition 1.5.6. *Similarity of matrices satisfies the following properties.*

1. *(Reflexive) For any A, we have A is similar to A.*
2. *(Symmetric) If A is similar to B, then B is similar to A.*
3. *(Transitive) If A is similar to B and B is similar to C, then A is similar to C.*

In mathematics, we say that a relation is an **equivalence relation** if it is reflexive, symmetric and transitive. Thus similarity of matrices defines an equivalence relations on the set of $n \times n$ matrices.

Proposition 1.5.7. *Suppose A and B are similar $n \times n$ matrices. Then*

1. *A and B have the same characteristic polynomial.*
2. *λ is an eigenvalue of A if and only if it is an eigenvalue of B.*
3. $\det(A) = \det(B)$
4. $\operatorname{tr}(A) = \operatorname{tr}(B)$

Proof. Suppose A and B are similar. Then there exists invertible matrix P such that $B = P^{-1}AP$.

1. Since

   $$\det(xI - B) = \det(xI - P^{-1}AP) = \det(P^{-1}(xI - A)P) = \det(xI - A),$$

   the characteristic polynomials of A and B are the same.

2. The statement follows easily by the fact that the eigenvalues of a matrix is exactly the roots of the characteristic polynomial of the matrix (Proposition 1.5.3).

3. $\det(B) = \det(P^{-1}AP) = \det(P)^{-1}\det(A)\det(P) = \det(A)$

4.
$$\begin{aligned} \operatorname{tr}(B) &= \operatorname{tr}(P^{-1}AP) \\ &= \operatorname{tr}(APP^{-1}) \text{ (Proposition 1.2.13)} \\ &= \operatorname{tr}(A) \end{aligned}$$

□

Definition 1.5.8 (Diagonalization)**.** *An $n \times n$ matrix A is* **diagonalizable** *if there exists invertible matrix P such that*

$$P^{-1}AP = D$$

is a diagonal matrix and we say that P **diagonalizes** *A. In other words, a matrix A is diagonalizable if and only if A is similar to a diagonal matrix.*

A matrix P diagonalizes A if and only if the column vectors of P are linearly independent eigenvectors of A.

Proposition 1.5.9. *let A be an $n \times n$ matrix and $P = [\boldsymbol{\xi}_1, \boldsymbol{\xi}_2, \ldots, \boldsymbol{\xi}_n]$ where $\boldsymbol{\xi}_1, \boldsymbol{\xi}_2, \ldots, \boldsymbol{\xi}_n$ are column vectors of P. Then the following statements are equivalent.*

1. *P is invertible and $P^{-1}AP = D$ where*
$$D = \begin{pmatrix} \lambda_1 & & & \mathit{0} \\ & \lambda_2 & & \\ & & \ddots & \\ \mathit{0} & & & \lambda_n \end{pmatrix}.$$

2. *The vectors $\boldsymbol{\xi}_1, \boldsymbol{\xi}_2, \ldots, \boldsymbol{\xi}_n$ are linearly independent eigenvectors of A associated with $\lambda_1, \lambda_2, \ldots, \lambda_n$ respectively.*

Proof. Observe that

$$\begin{aligned} & AP = PD \\ \Leftrightarrow\ & [A\boldsymbol{\xi}_1, A\boldsymbol{\xi}_2, \ldots, A\boldsymbol{\xi}_n] = [\lambda_1\boldsymbol{\xi}_1, \lambda_2\boldsymbol{\xi}_2, \ldots, \lambda_n\boldsymbol{\xi}_n] \\ \Leftrightarrow\ & A\boldsymbol{\xi}_i = \lambda_i\boldsymbol{\xi}_i \text{ for } i = 1, 2, \ldots, n \end{aligned}$$

Therefore $P^{-1}AP = D$ if and only if $\boldsymbol{\xi}_1, \boldsymbol{\xi}_2, \ldots, \boldsymbol{\xi}_n$ are linearly independent eigenvectors of A associated with $\lambda_1, \lambda_2, \ldots, \lambda_n$ respectively. □

To diagonalize a matrix A, we need to find all eigenvalues of A and as many linearly independent eigenvectors as possible for each eigenvalue. For each eigenvalue λ, the number of linearly independent eigenvectors associated with λ cannot be larger than the multiplicity of λ as a root of the characteristic equation.

Definition 1.5.10 (Algebraic and geometric multiplicity of eigenvalue)**.** *Let A be an $n \times n$ matrix and λ be an eigenvalue of A.*

1. *The* **algebraic multiplicity** $m_a(\lambda)$ *of λ is the multiplicity of λ as a root of the polynomial equation* $\det(xI - A) = 0$*, that means the largest positive integer k such that* $\det(xI - A)$ *is divisible by* $(x - \lambda)^k$*.*

2. *The* **geometric multiplicity** $m_g(\lambda)$ *of λ is maximum number of linearly independent eigenvectors associated with λ.*

The algebraic and geometric multiplicity of an eigenvalue satisfies the following inequality.

Proposition 1.5.11. *Let A be an $n \times n$ matrix and λ be an eigenvalue of A. Let $m_a(\lambda)$ be the algebraic multiplicity and $m_g(\lambda)$ be the geometric multiplicity of λ. Then we have*

$$1 \leq m_g(\lambda) \leq m_a(\lambda) \leq n.$$

Proof. There is at least one eigenvector $\boldsymbol{\xi}$ associated with λ and eigenvector is by definition nonzero. Thus we have $m_g(\lambda) \geq 1$. On the other hand, the characteristic equation is of degree n and thus we have $m_a(\lambda) \leq n$. Suppose $m_g(\lambda) = k$. Then there exists k linearly independent eigenvectors $\boldsymbol{\xi}_1, \boldsymbol{\xi}_2, \ldots, \boldsymbol{\xi}_k \in \mathbb{C}^n$ of A associated with λ. We are going to prove that the algebraic multiplicity of λ satisfies $m_a(\lambda) \geq k$. Now there exists (Theorem 1.3.30) $n - k$ vectors $\mathbf{v}_{k+1}, \ldots, \mathbf{v}_n \in \mathbb{C}^n$ such that $\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_k, \mathbf{v}_{k+1}, \ldots, \mathbf{v}_n$ constitute a basis for $\mathbb{C}^n$. Using these vectors as column vectors, the $n \times n$ matrix

$$P = [\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_k, \mathbf{v}_{k+1}, \ldots, \mathbf{v}_n]$$

is nonsingular (Theorem 1.3.33). Consider the matrix $B = P^{-1}AP$ which must be of the form

$$B = \begin{pmatrix} \lambda I & C \\ 0 & D \end{pmatrix}$$

where I is the $k \times k$ identity matrix, 0 is the $(n-k) \times k$ zero matrix, C is a $k \times (n-k)$ matrix and D is an $(n-k) \times (n-k)$ matrix. Note that since A and B are similar, the characteristic equation of A and B are the same (Proposition 1.5.7). Observe that

$$\det(xI - B) = \begin{vmatrix} (x-\lambda)I & -C \\ 0 & xI - D \end{vmatrix} = (x-\lambda)^k \det(xI - D).$$

We see that the algebraic multiplicity of λ as root of the characteristic equation of B is as least k and therefore the algebraic multiplicity of λ as root of the characteristic equation of A is as least k. □

Note that by **fundamental theorem of algebra**[5], the sum of algebraic multiplicities of all eigenvalues of an $n \times n$ matrix is n.

Theorem 1.5.12. *Let A be an $n \times n$ matrix and $\lambda_1, \lambda_2, \ldots, \lambda_k$ be distinct eigenvalues of A. Suppose $\boldsymbol{\xi}_1, \boldsymbol{\xi}_2, \ldots, \boldsymbol{\xi}_k$ are eigenvectors associated with $\lambda_1, \lambda_2, \ldots, \lambda_k$ respectively. Then $\boldsymbol{\xi}_1, \boldsymbol{\xi}_2, \ldots, \boldsymbol{\xi}_k$ are linearly independent. More generally, suppose*

$$E = \{\boldsymbol{\xi}_{11}, \ldots, \boldsymbol{\xi}_{1m_1}, \boldsymbol{\xi}_{21}, \ldots, \boldsymbol{\xi}_{2m_2}, \ldots, \boldsymbol{\xi}_{k1}, \ldots, \boldsymbol{\xi}_{km_k}\}$$

are vectors such that for each $i = 1, 2, \ldots, k$,

$$E_i = \{\boldsymbol{\xi}_{i1}, \ldots, \boldsymbol{\xi}_{im_i}\}$$

is a set of linearly independent eigenvectors associated with λ_i. Then the vectors in E are linearly independent.

Proof. We prove the first part of the statement by induction on k. When $k = 1$, the vector $\boldsymbol{\xi}$ is linearly independent since eigenvector is nonzero by definition. Assume that any $k-1$ eigenvectors associated with distinct eigenvectors are linearly independent. Let $\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_k$ be eigenvectors associated with distinct eigenvalues $\lambda_1, \ldots, \lambda_k$. Suppose

$$c_1\boldsymbol{\xi}_1 + c_2\boldsymbol{\xi}_2 + \cdots + c_k\boldsymbol{\xi}_k = \mathbf{0}.$$

Multiplying $A - \lambda_k I$ from the left to both sides, we have

$$\begin{aligned} c_1(A - \lambda_k I)\boldsymbol{\xi}_1 + \cdots + c_{k-1}(A - \lambda_k I)\boldsymbol{\xi}_{k-1} + c_k(A - \lambda_k I)\boldsymbol{\xi}_k &= \mathbf{0} \\ c_1(\lambda_1 - \lambda_k)\boldsymbol{\xi}_1 + \cdots + c_{k-1}(\lambda_{k-1} - \lambda_k)\boldsymbol{\xi}_{k-1} &= \mathbf{0}. \end{aligned}$$

[5]One way of stating the fundamental theorem of algebra is that the sum of the multiplicities of all roots of a polynomial equation of degree n is n.

By induction hypothesis, the vectors $\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_{k-1}$ are linearly independent which implies $c_i(\lambda_i - \lambda_k) = 0$ for any $i = 1, 2, \ldots, k-1$. Now $\lambda_i - \lambda_k \neq 0$ for $i \neq k$ since $\lambda_1, \ldots, \lambda_k$ are distinct. Hence we have $c_i = 0$ for $i = 1, 2, \ldots, k-1$. It follows that $c_k \boldsymbol{\xi}_k = \mathbf{0}$ which implies $c_k = 0$ since $\boldsymbol{\xi}_k \neq \mathbf{0}$ being an eigenvector. Therefore $\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_k$ are linearly independent.

For the more general statement, suppose

$$\boldsymbol{\eta}_1 + \boldsymbol{\eta}_2 + \cdots + \boldsymbol{\eta}_k = \mathbf{0}$$

where

$$\boldsymbol{\eta}_i = c_{i1}\boldsymbol{\xi}_{i1} + \cdots + c_{im_i}\boldsymbol{\xi}_{im_i}$$

for $i = 1, 2, \ldots, k$. Observe that $A\boldsymbol{\eta}_i = \lambda_i \boldsymbol{\eta}_i$, and by the first part of the proof, we must have

$$\boldsymbol{\eta}_1 = \boldsymbol{\eta}_2 = \cdots = \boldsymbol{\eta}_k = \mathbf{0}.$$

Note that $\boldsymbol{\xi}_{i1}, \ldots, \boldsymbol{\xi}_{im_i}$ are linearly independent which implies $c_{i1} = c_{i2} = \cdots = c_{im_i} = 0$. Therefore the vectors in E are linearly independent. □

Theorem 1.5.13. *Let A be an $n \times n$ matrix. Then the following statements are equivalent.*

1. *A is diagonalizable.*
2. *There exists n linearly independent eigenvectors of A.*
3. *For each eigenvalue λ of A, we have $m_g(\lambda) = m_a(\lambda)$ where $m_g(\lambda)$ and $m_g(\lambda)$ are the geometric multiplicity and algebraic multiplicity of λ respectively.*

Proof. The first two statements are equivalent by Proposition 1.5.9. We are going to prove that (2) and (3) are equivalent. Let $\lambda_1, \ldots, \lambda_k$ be all eigenvalues of A. Suppose there exists n linearly independent eigenvectors of A for which m_i of them are associated with λ_i for $1 \leq i \leq k$. Then $m_i \leq m_g(\lambda_i)$ by definition of m_g and thus

$$\begin{aligned} n &= m_1 + m_2 + \cdots + m_k \\ &\leq m_g(\lambda_1) + m_g(\lambda_2) + \cdots + m_g(\lambda_k) \\ &\leq m_a(\lambda_1) + m_a(\lambda_2) + \cdots + m_a(\lambda_k) \text{ (Proposition 1.5.11)} \\ &= n \text{ (Fundamental theorem of algebra).} \end{aligned}$$

Therefore $m_g(\lambda_i) = m_a(\lambda_i)$ for any $i = 1, 2, \ldots, k$.

Suppose for each $i = 1, 2, \ldots, k$, we have $m_g(\lambda_i) = m_a(\lambda_i)$ and let $\boldsymbol{\xi}_{i1}, \ldots, \boldsymbol{\xi}_{im_i}$, where $m_i = m_g(\lambda_i) = m_a(\lambda_i)$ be linearly independent eigenvectors associated with λ_i. Then by Theorem 1.5.12, the vectors

$$\boldsymbol{\xi}_{11}, \ldots, \boldsymbol{\xi}_{1m_1}, \boldsymbol{\xi}_{21}, \ldots, \boldsymbol{\xi}_{2m_2}, \ldots, \boldsymbol{\xi}_{k1}, \ldots, \boldsymbol{\xi}_{km_k}$$

are linearly independent. By fundamental theorem of algebra, we have

$$m_1 + \cdots + m_k = m_a(\lambda_1) + \cdots + m_a(\lambda_k) = n.$$

Therefore we have n linearly independent eigenvectors which implies A is diagonalizable. □

In particular, we have

Proposition 1.5.14. *Let A be an $n \times n$ matrix. Suppose A has n distinct eigenvalues. Then A is diagonalizable.*

Note that the converse of the above theorem is false. That is, a diagonalizable $n \times n$ matrix may have less than n distinct eigenvalues.

Theorem 1.5.15 (Cayley-Hamilton theorem)**.** *Let A be an $n \times n$ matrix and $p(x) = \det(xI - A)$ be its characteristic polynomial. Then $p(A) = 0$.*

Proof. Let $B = xI - A$ and

$$p(x) = \det(B) = x^n + c_{n-1}x^{n-1} + \cdots + c_1 x + c_0$$

be the characteristic polynomial of A. Consider $B = xI - A$ as an $n \times n$ matrix whose entries are polynomial in x. Write the adjugate $\operatorname{adj}(B)$ of B as a polynomial of degree $n - 1$ in x with matrix coefficients

$$\operatorname{adj}(B) = B_{n-1}x^{n-1} + \cdots + B_1 x + B_0$$

where the coefficients B_i are $n \times n$ constant matrices. On one hand, we have

$$\begin{aligned} \det(B)I &= (x^n + c_{n-1}x^{n-1} + \cdots + c_1 x + c_0)I \\ &= Ix^n + c_{n-1}Ix^{n-1} + \cdots + c_1 Ix + c_0 I. \end{aligned}$$

One the other hand, we have

$$\begin{aligned} B\operatorname{adj}(B) &= (xI - A)(B_{n-1}x^{n-1} + \cdots + B_1 x + B_0) \\ &= B_{n-1}x^n + (B_{n-2} - AB_{n-1})x^{n-1} + \cdots + (B_0 - AB_1)x - AB_0. \end{aligned}$$

Comparing the coefficients of

$$\det(B)I = B\text{adj}(B),$$

we get

$$\begin{aligned} I &= B_{n-1} \\ c_{n-1}I &= B_{n-2} - AB_{n-1} \\ &\vdots \\ c_1 I &= B_0 - AB_1 \\ c_0 I &= -AB_0. \end{aligned}$$

Now we multiply the first equation by A^n, the second equation by A^{n-1}, and so on, and the last one by I. Then adding up the resulting equations, we obtain

$$\begin{aligned} p(A) &= A^n + c_{n-1}A^{n-1} + \cdots + c_1 A + c_0 I \\ &= A^n B_{n-1} + (A^{n-1}B_{n-2} - A^n B_{n-1}) + \cdots + (AB_0 - A^2 B_1) - AB_0 \\ &= 0. \end{aligned}$$

□

1.6 Self-adjoint operator

Self-adjoint operators are linear operators which satisfy $\langle L(\mathbf{u}), \mathbf{v}\rangle = \langle \mathbf{u}, L(\mathbf{v})\rangle$. They form an important class of linear operators. To understand them, we need to extend our studies in the previous sections to vector space over $\mathbb{C}$, complex inner product space and linear operators on these spaces.

Definition 1.6.1 (Inner product)**.** *The* **inner product** *of two vectors* $\mathbf{w}, \mathbf{z} \in \mathbb{C}^n$ *is defined by*

$$\langle \mathbf{w}, \mathbf{z}\rangle = w_1\overline{z_1} + w_2\overline{z_2} + \cdots + w_n\overline{z_n}$$

for $\mathbf{w} = (w_1, w_2, \ldots, w_n), \mathbf{z} = (z_1, z_2, \ldots, z_n) \in \mathbb{C}^n$.

Inner product has the following properties.

Proposition 1.6.2 (Properties of inner product)**.** *Let* $\mathbf{u}, \mathbf{v}, \mathbf{w} \in \mathbb{C}^n$ *and* $\alpha, \beta \in \mathbb{C}$. *Then*

1. *(Linear in first argument):*

$$\langle \alpha\mathbf{u} + \beta\mathbf{v}, \mathbf{w}\rangle = \alpha\langle \mathbf{u}, \mathbf{w}\rangle + \beta\langle \mathbf{v}, \mathbf{w}\rangle$$

2. *(Conjugate symmetric):*

$$\langle \mathbf{v}, \mathbf{u}\rangle = \overline{\langle \mathbf{u}, \mathbf{v}\rangle}$$

3. *(Positive definite):*

$$\langle \mathbf{v}, \mathbf{v}\rangle \geq 0$$

with equality holds if and only if $\mathbf{v} = \mathbf{0}$.

Note that by linearity in first argument and conjugate symmetry, inner product is conjugate linear in the second argument, that is,

$$\langle \mathbf{w}, \alpha\mathbf{u} + \beta\mathbf{v}\rangle = \overline{\alpha}\langle \mathbf{w}, \mathbf{u}\rangle + \overline{\beta}\langle \mathbf{w}, \mathbf{v}\rangle.$$

A subset $V \subset \mathbb{C}^m$ is a complex vector subspace if $\mathbf{0} \in V$ and for any $\mathbf{u}, \mathbf{v} \in V$, $\alpha, \beta \in \mathbb{C}$, we have

$$\alpha\mathbf{u} + \beta\mathbf{v} \in V.$$

A linear operator on a complex vector subspace $V \subset \mathbb{C}^m$ is a function $L : V \to V$ such that for any $\mathbf{u}, \mathbf{v} \in V$, $\alpha, \beta \in \mathbb{C}$, we have

$$L(\alpha\mathbf{u} + \beta\mathbf{v}) = \alpha L(\mathbf{u}) + \beta L(\mathbf{v}).$$

A linear operator on a n dimensional vector subspace $V \subset \mathbb{C}^m$ can be represented by a $n \times n$ matrix with respect to a basis for V. When we talk about matrix representation, we need to specify the order of vectors in the basis. We called a basis $E = (\mathbf{v}_1, \ldots, \mathbf{v}_n)$ whose vectors are ordered an **ordered basis**.

Definition 1.6.3 (Matrix representation of linear operator)**.** *Let* $V \subset \mathbb{C}^m$ *be a vector subspace with dimension* $\dim(V) = n$ *and* $L : V \to V$ *be a linear operator on* V*. Let* $E = (\mathbf{u}_1, \ldots, \mathbf{u}_n)$ *be an ordered basis for* V*. Then for each* $j = 1, 2, \ldots, n$*, we may write*

$$L(\mathbf{u}_j) = a_{1j}\mathbf{u}_1 + a_{2j}\mathbf{u}_2 + \cdots + a_{nj}\mathbf{u}_n$$

for some complex numbers $a_{1j}, a_{2j}, \ldots, a_{nj} \in \mathbb{C}$. *We say that the* $n \times n$ *matrix*

$$A_E = [a_{ij}] = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix}$$

is a **matrix representation** *of* L *with respect to the ordered basis* E.

One may write the equalities

$$L(\mathbf{u}_j) = a_{1j}\mathbf{u}_1 + a_{2j}\mathbf{u}_2 + \cdots + a_{nj}\mathbf{u}_n$$

for $j = 1, 2, \ldots, n$, as

$$L(\mathbf{u}_1, \ldots, \mathbf{u}_n) = (\mathbf{u}_1, \ldots, \mathbf{u}_n)A_E.$$

The matrix representation A_E has the following interpretation. For any $\mathbf{v} \in V$, if we write

$$\mathbf{v} = \alpha_1\mathbf{u}_1 + \alpha_2\mathbf{u}_2 + \cdots + \alpha_n\mathbf{u}_n,$$

then

$$L(\mathbf{v}) = \beta_1\mathbf{u}_1 + \beta_2\mathbf{u}_2 + \cdots + \beta_n\mathbf{u}_n$$

where

$$\begin{pmatrix} \beta_1 \\ \vdots \\ \beta_n \end{pmatrix} = A_E \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{pmatrix}.$$

We may choose different bases for V and obtain different matrix representations of L. However, two matrix representations of a linear operator are always similar.

Proposition 1.6.4. *Let* $V \subset \mathbb{C}^m$ *be a vector subspace with* $\dim(V) = n$ *and* $L : V \to V$ *be a linear operator on* V. *Let* $E = (\mathbf{u}_1, \ldots, \mathbf{u}_n)$ *and* $F = (\mathbf{v}_1, \ldots, \mathbf{v}_n)$ *be two ordered basis for* V. *Let* A_E *and* A_F *be the matrix representation of* L *with respect to bases* E *and* F *respectively. Then* A_E *and* A_F *are similar matrices.*

Proof. We need to find an invertible matrix P such that $A_F = P^{-1}A_E P$. We may write

$$L(\mathbf{u}_1, \ldots, \mathbf{u}_n) = (\mathbf{u}_1, \ldots, \mathbf{u}_n)A_E$$

and

$$L(\mathbf{v}_1, \ldots, \mathbf{v}_n) = (\mathbf{v}_1, \ldots, \mathbf{v}_n)A_F.$$

For each $j = 1, 2, \ldots, n$, write

$$\mathbf{v}_j = p_{1j}\mathbf{u}_1 + p_{2j}\mathbf{u}_2 + \cdots + p_{nj}\mathbf{u}_n,$$

where $p_{1j}, \ldots, p_{nj} \in \mathbb{C}$ which can be written as

$$(\mathbf{v}_1, \ldots, \mathbf{v}_n) = (\mathbf{u}_1, \ldots, \mathbf{u}_n)P$$

where $P = [p_{ij}]$. Note also that

$$(\mathbf{v}_1, \ldots, \mathbf{v}_n)P^{-1} = (\mathbf{u}_1, \ldots, \mathbf{u}_n).$$

Thus we have

$$\begin{aligned} L(\mathbf{v}_1, \ldots, \mathbf{v}_n) &= L((\mathbf{u}_1, \ldots, \mathbf{u}_n)P) \\ &= L(\mathbf{u}_1, \ldots, \mathbf{u}_n)P \\ &= (\mathbf{u}_1, \ldots, \mathbf{u}_n)A_E P \\ &= (\mathbf{v}_1, \ldots, \mathbf{v}_n)P^{-1}A_E P. \end{aligned}$$

This means $A_F = P^{-1}A_E P$. Therefore A_E and A_F are similar. □

Proposition 1.6.5. *Let $V \subset \mathbb{C}^m$ be a linear operator and $L : V \to V$ be a linear operator on V. Let E and F be two ordered bases for V. Suppose A_E and A_F are the matrix representations of L with respect to E and F respectively. Then the following statements holds.*

1. *A_E and A_F have the same characteristic polynomial.*
2. *A_E and A_F have the same set of eigenvalues.*
3. $\det(A_E) = \det(A_F)$
4. $\operatorname{tr}(A_E) = \operatorname{tr}(A_F)$

The above proposition allows us to define the determinant and trace of a linear operator.

Definition 1.6.6 (Determinant and trace of linear operator)**.** *Let $V \subset \mathbb{C}^m$ be a vector subspace with* $\dim(V) = n$ *and $L : V \to V$ be a linear operator on V. The* **determinant** *and* **trace** *of L is the determinant and trace of a matrix representation of L respectively.*

Definition 1.6.7 (Eigenvalues and eigenvectors of linear operators)**.** *Let $V \subset \mathbb{C}^m$ be a vector subspace and $L : V \to V$ be a linear operator. Suppose $\lambda \in \mathbb{C}$ and $\boldsymbol{\xi} \in V$ is a nonzero vector such that*

$$L(\boldsymbol{\xi}) = \lambda\boldsymbol{\xi}.$$

Then we say that λ is an **eigenvalue** *of L and $\boldsymbol{\xi}$ is an* **eigenvector** *of L associated with λ.*

It is not difficult to see that λ is an eigenvalue of L if and only if λ is an eigenvalue of the matrix representation A_E. In fact

$$\boldsymbol{\xi} = \alpha_1\mathbf{u}_1 + \alpha_2\mathbf{u}_2 + \cdots + \alpha_n\mathbf{u}_n$$

is an eigenvector of L associated with λ if and only if $(\alpha_1, \ldots, \alpha_n) \in \mathbb{C}^n$ is an eigenvector of A_E associated with λ.

Definition 1.6.8 (Self-adjoint operator)**.** *Let $V \subset \mathbb{C}^m$ be a complex vector subspace and $L : V \to V$ be a linear operator on V. We say that L is* **self-adjoint** *if for any* $\mathbf{u}, \mathbf{v} \in V$*, we have*

$$\langle L(\mathbf{u}), \mathbf{v}\rangle = \langle \mathbf{u}, L(\mathbf{v})\rangle.$$

Let's study the matrix representation of a self-adjoint operator. For complex matrix, it is more natural to consider conjugate transpose in stead transpose.

Definition 1.6.9 (Conjugate transpose)**.** *Let $A = [a_{ij}]$ be an $n \times n$ complex matrix. The* **conjugate transpose** *of A is defined by*

$$A^* = \overline{A}^T.$$

In other words, the ij-th entry of the conjugate transpose A^* of A is

$$[A^*]_{ij} = \overline{a_{ji}}.$$

A linear operator is self-adjoint if and only if its matrix representation with respect to an orthonormal basis is Hermitian.

Definition 1.6.10 (Hermitian and unitary matrix)**.** *Denote by* $A^* = \overline{A}^T$ *the* **conjugate transpose** *of an* $n \times n$ *complex matrix* A*.*

1. *An* $n \times n$ *matrix* H *is said to be* **Hermitian** *if* $H^* = H$*.*
2. *An* $n \times n$ *matrix* U *is said to be* **unitary** *if* U *is invertible and* $U^* = U^{-1}$*.*

Proposition 1.6.11. *Let* $V \subset \mathbb{C}^m$ *be a complex vector subspace and* E *be an ordered orthonormal basis. Let* $L : V \to V$ *be a linear operator on* V*. Then* L *is self-adjoint if and only if the matrix representation* A_E *of* L *with respect to* E *is Hermitian.*

Proof. Let $E = (\mathbf{u}_1, \ldots, \mathbf{u}_n)$ be an ordered orthonormal basis. Suppose the matrix representation of L with respect to E is $A_E = [a_{ij}]$. Now for $i = 1, 2, \ldots, n$,

$$L(\mathbf{u}_i) = a_{1i}\mathbf{u}_1 + a_{2i}\mathbf{u}_2 + \cdots + a_{ni}\mathbf{u}_n.$$

Since E is an orthonormal basis, we have

$$\langle L(\mathbf{u}_i), \mathbf{u}_j \rangle = \langle a_{1i}\mathbf{u}_1 + a_{2i}\mathbf{u}_2 + \cdots + a_{ni}\mathbf{u}_n, \mathbf{u}_j \rangle = a_{ji}$$

and similarly

$$\langle \mathbf{u}_i, L(\mathbf{u}_j) \rangle = \langle \mathbf{u}_i, a_{1j}\mathbf{u}_1 + a_{2j}\mathbf{u}_2 + \cdots + a_{nj}\mathbf{u}_n \rangle = \overline{a_{ij}}.$$

Now if L is self-adjoint, we have $a_{ji} = \overline{a_{ij}}$ for any $1 \leq i, j \leq n$ which means A_E is Hermitian.

Conversely if A_E is Hermitian, then $a_{ji} = \overline{a_{ij}}$ which implies $\langle L(\mathbf{u}_i), \mathbf{u}_j \rangle = \langle \mathbf{u}_i, L(\mathbf{u}_j) \rangle$ for any $1 \leq i, j \leq n$. Then it follows readily that L is self-adjoint. □

Let λ and $\boldsymbol{\xi}$ be eigenvalue and eigenvectors of a self-adjoint operator L. Then the image under L of a vector orthogonal to $\boldsymbol{\xi}$ is orthogonal to $\boldsymbol{\xi}$.

Theorem 1.6.12. *Let* $V \subset \mathbb{C}^m$ *be a vector subspace and* $L : V \to V$ *be a self-adjoint operator. Let* $\boldsymbol{\xi}$ *be an eigenvector of* L*. Then for any vector* $\boldsymbol{\eta} \in V$ *with* $\langle \boldsymbol{\xi}, \boldsymbol{\eta} \rangle = 0$*, we have* $\langle \boldsymbol{\xi}, L(\boldsymbol{\eta}) \rangle = 0$*.*

Proof. 1. Suppose $\langle \boldsymbol{\xi}, \boldsymbol{\eta} \rangle = 0$. Then

$$\begin{aligned} \langle \boldsymbol{\xi}, L(\boldsymbol{\eta}) \rangle &= \langle L(\boldsymbol{\xi}), \boldsymbol{\eta} \rangle \\ &= \langle \lambda \boldsymbol{\xi}, \boldsymbol{\eta} \rangle \\ &= \lambda \langle \boldsymbol{\xi}, \boldsymbol{\eta} \rangle \\ &= 0 \end{aligned}$$

□

Definition 1.6.13 (Orthogonal complement)**.** *Let $V \subset \mathbb{C}^m$ be a vector subspace and $W \subset V$ be a vector subspace of V. The* **orthogonal complement** *of W in V is defined by*

$$W^\perp = \{\mathbf{w}^\perp \in V : \langle \mathbf{w}, \mathbf{w}^\perp \rangle = 0 \textit{ for any } \mathbf{w} \in W\}$$

Proposition 1.6.14. *Let $V \subset \mathbb{C}^m$ be a vector subspace with* $\det(V)$*. Let $W \subset V$ be a vector subspace of V and $W^\perp$ be the orthogonal complement of W in V. Then*

1. $W \cap W^\perp = \{\mathbf{0}\}$
2. *For any $\mathbf{v} \in W$, there exists unique decomposition*

   $$\mathbf{v} = \mathbf{w} + \mathbf{w}^\perp$$

   such that $\mathbf{w} \in W$ and $\mathbf{w}^\perp \in W^\perp$.
3. $\dim(W) + \dim(W^\perp) = \dim(V)$

Proof. 1. Suppose $\mathbf{v} \in W \cap W^\perp$. Then $\mathbf{v} \in W$ and $\mathbf{v} \in W^\perp$ which implies $\langle \mathbf{v}, \mathbf{v} \rangle = 0$. Thus we must have $\mathbf{v} = \mathbf{0}$. Therefore $W \cap W^\perp = \{\mathbf{0}\}$.

2. Suppose $\mathbf{v} \in V$. By **extreme value theorem**[6], there exists $\mathbf{w} \in W$ such that $\|\mathbf{v} - \mathbf{w}\| \leq \|\mathbf{v} - \mathbf{z}\|$ for any $\mathbf{z} \in W$. Let $\mathbf{w}^\perp = \mathbf{v} - \mathbf{w}$. Suppose there exist $\mathbf{z} \in W$ such that $\langle \mathbf{z}, \mathbf{w}^\perp \rangle \neq 0$. By replacing $\mathbf{z}$ by

[6] The extreme value theorem says that if f is a continuous function defined on a closed and bounded set D and f is bounded from below, then f attains its minimum on D. That means there exists $z \in D$ such that $f(z) \leq f(x)$ for any $x \in D$.

its multiple, we may assume that $\langle \mathbf{z}, \mathbf{w}^\perp \rangle = \alpha > 0$ and $\|\mathbf{z}\| = 1$. Note that $\mathbf{w} + \alpha\mathbf{z} \in W$ and

$$\begin{aligned}
\|\mathbf{v} - (\mathbf{w} + \alpha\mathbf{z})\|^2 &= \|\mathbf{w}^\perp - \alpha\mathbf{z}\|^2 \\
&= \|\mathbf{w}^\perp\|^2 - \langle \alpha\mathbf{z}, \mathbf{w}^\perp \rangle - \langle \mathbf{w}^\perp, \alpha\mathbf{z} \rangle + \|\alpha\mathbf{z}\|^2 \\
&= \|\mathbf{w}^\perp\|^2 - \alpha^2 - \overline{\alpha}^2 + |\alpha|^2 \\
&= \|\mathbf{w}^\perp\|^2 - \alpha^2
\end{aligned}$$

which contradicts the construction of $\mathbf{w}$ that $\|\mathbf{v}-\mathbf{z}\| \geq \|\mathbf{v}-\mathbf{w}\| = \|\mathbf{w}^\perp\|$ for any $\mathbf{z} \in W$.. Hence we have $\langle \mathbf{z}, \mathbf{w}^\perp \rangle = 0$ for any $\mathbf{z} \in W$. This means $\mathbf{w}^\perp \in W^\perp$ and $\mathbf{v} = \mathbf{w} + \mathbf{w}^\perp$ is the required decomposition.

To prove uniqueness, suppose

$$\mathbf{v} = \mathbf{w} + \mathbf{w}^\perp = \mathbf{z} + \mathbf{z}^\perp$$

where $\mathbf{w}, \mathbf{z} \in W$ and $\mathbf{w}^\perp, \mathbf{z}^\perp \in W^\perp$. Then

$$\mathbf{w} - \mathbf{z} = \mathbf{z}^\perp - \mathbf{w}^\perp$$

is a vector which lies in both W and $W^\perp$. The vector $\mathbf{w} - \mathbf{z} = \mathbf{z}^\perp - \mathbf{w}^\perp$ lies in both W and $W^\perp$ which implies $\mathbf{w} - \mathbf{z} = \mathbf{z}^\perp - \mathbf{w}^\perp = \mathbf{0}$ by (1). Therefore the decomposition $\mathbf{v} = \mathbf{w} + \mathbf{w}^\perp$ is unique.

3. Let $\mathbf{w}_1, \ldots, \mathbf{w}_p \in W$ be a basis for W and $\mathbf{w}_1^\perp, \ldots, \mathbf{w}_q^\perp \in W^\perp$ be a basis for $W^\perp$. For any $\mathbf{v} \in V$, by (2) there exists unique decomposition $\mathbf{v} = \mathbf{w} + \mathbf{w}^\perp$ where $\mathbf{w} \in W$ and $\mathbf{w}^\perp \in W^\perp$. Then there exists constants $\alpha_1, \ldots, \alpha_p, \beta_1, \ldots, \beta_q \in \mathbb{C}$ such that

$$\begin{aligned}
\mathbf{v} &= \mathbf{w} + \mathbf{w}^\perp \\
&= \alpha_1\mathbf{w}_1, \cdots, \alpha_p\mathbf{w}_p + \beta_1\mathbf{w}_1^\perp, \ldots, \beta_q\mathbf{w}_q^\perp.
\end{aligned}$$

On the other hand, the vectors $\mathbf{w}_1, \ldots, \mathbf{w}_p, \mathbf{w}_1^\perp, \ldots, \mathbf{w}_q^\perp$ are linearly independent since they are nonzero mutually orthogonal vectors (Proposition 1.3.21), and thus constitute a basis for V. It follows that $p+q = n$ which means $\dim(W) + \dim(W^\perp) = \dim(V)$.

□

Theorem 1.6.15 (Spectral theorem for self-adjoint operator)**.** *Let $V \subset \mathbb{C}^m$ be a vector subspace and $L : V \to V$ be a linear operator on V. Then the following statements hold.*

1. *All eigenvalues of* L *are real.*

2. *There exists eigenvectors* $\boldsymbol{\xi}_1, \boldsymbol{\xi}_2, \ldots, \boldsymbol{\xi}_n$ *of* L *which constitute an orthonormal basis for* V*.*

Proof. 1. Suppose $\boldsymbol{\xi}$ is an eigenvector of L which means $\boldsymbol{\xi} \neq \mathbf{0}$ and $L(\boldsymbol{\xi}) = \lambda\boldsymbol{\xi}$ where $\lambda \in \mathbb{C}$.

$$\begin{aligned} \lambda\langle\boldsymbol{\xi},\boldsymbol{\xi}\rangle &= \langle\lambda\boldsymbol{\xi},\boldsymbol{\xi}\rangle \\ &= \langle L(\boldsymbol{\xi}),\boldsymbol{\xi}\rangle \\ &= \langle\boldsymbol{\xi}, L(\boldsymbol{\xi})\rangle \\ &= \langle\boldsymbol{\xi},\lambda\boldsymbol{\xi}\rangle \\ &= \overline{\lambda}\langle\boldsymbol{\xi},\boldsymbol{\xi}\rangle \end{aligned}$$

Since $\boldsymbol{\xi} \neq 0$, we have $\overline{\lambda} = \lambda$ which means λ is real.

2. We prove the statement by induction on $\dim(V)$. Suppose $\dim(V) = 1$. Let $\boldsymbol{\xi} \in V$ be a unit vector. Then $L(\boldsymbol{\xi}) = \lambda\boldsymbol{\xi}$ for some scalar λ since $\dim(V) = 1$. Thus $\boldsymbol{\xi}$ constitutes a basis for V.

   Assume that the statement holds for any vector subspace of dimension k. Let $V \subset \mathbb{C}^m$ be a subspace with $\dim(V) = k+1$. By fundamental theorem of algebra, the characteristic polynomial of L has at least one root $\lambda \in \mathbb{C}$. Then λ is an eigenvalue of L which means there exists unit vector $\boldsymbol{\xi}_0 \in V$ such that $L(\boldsymbol{\xi}_0) = \lambda\boldsymbol{\xi}_0$. Let

   $$W = \{\mathbf{w} \in V : \mathbf{w} = \alpha\boldsymbol{\xi}_0 \text{ for some } \alpha \in \mathbb{C}\}$$

   and

   $$W^\perp = \{\mathbf{w}^\perp \in V : \langle\boldsymbol{\xi}_0, \mathbf{w}^\perp\rangle = 0\}$$

   be the orthogonal complement of W in V which is of dimension k by Proposition 1.6.14. By Theorem 1.6.12, we have $L(\mathbf{w}^\perp) \in W^\perp$ for any $\mathbf{w}^\perp \in W^\perp$. Thus the restriction $L|_{W^\perp} : W^\perp \to W^\perp$ of L on $W^\perp$ can be considered a linear operator on $W^\perp$. It is not difficult to see that $L|_{W^\perp}$ is self-adjoint. Note that $\dim(W^\perp) = k$. By induction hypothesis, there exists eigenvectors $\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_k \in W^\perp$ of $L|_{W^\perp}$ which constitute a basis for $W^\perp$. Now $\boldsymbol{\xi}_0, \boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_k, \boldsymbol{\xi}_{k+1} \in V$ are eigenvectors of L which constitute an orthonomal basis for V. This completes the induction step and the proof of the theorem.

□

Theorem 1.6.16 (Spectral theorem for Hermitian matrices)**.** *Let H be an $n \times n$ Hermitian matrix. Then the following statements hold.*

1. *All eigenvalues of H are real.*

2. *There exists an orthonormal basis for $\mathbb{C}^n$ which consists of eigenvectors of H.*

3. *There exists special unitary matrix U which diagonalizes H, that is, U^*HU is a diagonal matrix.*

Proof. Let $L_H : \mathbb{C}^n \to \mathbb{C}^n$ be the linear operator defined by $L_H(\mathbf{v}) = H\mathbf{v}$, where we consider $\mathbf{v}$ as column vector. Then L_H is represented by the matrix H with respect to the standard basis. Thus L_H is a self-adjoint operator since H is Hermitian (Proposition 1.6.11). By spectral theorem (Theorem 1.6.15), all eigenvalues of L_H are real and there exists orthonormal basis $\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_n$ for $\mathbb{C}^n$ consisting of eigenvectors of L_H which are also eigenvectors of H. This proves the first two statements. For the third statement, note that the matrix $U = [\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_n]$ diagonalizes H (Proposition 1.5.9). Since $\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_n$ constitute an orthonormal basis, U is a unitary matrix. One may multiply a suitable complex number to the first column of U making its determinant equals to one and keeping the matrix unitary. Then the resulting matrix is a special unitary matrix which diagonalizes H. □

Note that a real matrix is Hermitian if and only if it is symmetric and is unitary if and only if it is orthogonal. Thus we have the following spectral theorem for real symmetric matrices.

Theorem 1.6.17 (Spectral theorem for real symmetric matrices)**.** *Let S be an $n \times n$ real symmetric matrix. Then the following statements hold.*

1. *All eigenvalues of S are real.*

2. *There exists an orthonormal basis for $\mathbb{R}^n$ which consists of eigenvectors of S.*

3. *There exists special orthogonal matrix Q which diagonalizes S, that is, Q^TSQ is a diagonal matrix.*

1.7 Some transcendental functions

In this section, we discuss the most basic transcendental functions namely, exponential function, logarithmic function, trigonometric functions and hyperbolic functions. We will give the definitions, list some basic identities, and do some calculus on them.

Definition 1.7.1. *The exponential function, logarithmic function, trigonometric functions and hyperbolic functions are defined as follows.*

1. **Exponential function**:

$$e^x = \sum_{k=0}^{\infty} \frac{x^k}{k!} = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots \text{ for } x \in \mathbb{R}$$

2. **Trigonometric functions**: *There are 6 trigonometric functions which are defined as follows.*

Cosine: $$\cos x = \sum_{k=0}^{\infty} \frac{(-1)^k x^{2k}}{(2k)!} = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \frac{x^6}{6!} + \cdots \text{ for } x \in \mathbb{R}$$

Sine: $$\sin x = \sum_{k=0}^{\infty} \frac{(-1)^k x^{2k+1}}{(2k+1)!} = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + \cdots \text{ for } x \in \mathbb{R}$$

Tangent: $$\tan x = \frac{\sin x}{\cos x} \text{ for } x \neq \frac{(2k+1)\pi}{2},\ k \in \mathbb{Z}$$

Cotangent: $$\cot x = \frac{\cos x}{\sin x} \text{ for } x \neq k\pi,\ k \in \mathbb{Z}$$

Secant: $$\sec x = \frac{1}{\cos x} \text{ for } x \neq \frac{(2k+1)\pi}{2},\ k \in \mathbb{Z}$$

Cosecant: $$\csc x = \frac{1}{\sin x} \text{ for } x \neq k\pi,\ k \in \mathbb{Z}$$

3. **Hyperbolic functions**: *There are 6 hyperbolic functions which are defined as follows.*
Hyperbolic cosine:

$$\cosh x = \frac{e^x + e^{-x}}{2} = \sum_{k=0}^{\infty} \frac{x^{2k}}{(2k)!} = 1 + \frac{x^2}{2!} + \frac{x^4}{4!} + \frac{x^6}{6!} + \cdots \text{ for } x \in \mathbb{R}$$

Hyperbolic sine:

$$\sinh x = \frac{e^x - e^{-x}}{2} = \sum_{k=0}^{\infty} \frac{x^{2k+1}}{(2k+1)!} = x + \frac{x^3}{3!} + \frac{x^5}{5!} + \frac{x^7}{7!} + \cdots \text{ for } x \in \mathbb{R}$$

Hyperbolic tangent:

$$\tanh x = \frac{\sinh x}{\cosh x} \text{ for } x \in \mathbb{R}$$

Hyperbolic cotangent:

$$\coth x = \frac{\cosh x}{\sinh x} \text{ for } x \neq 0$$

Hyperbolic secant:

$$\text{sech} x = \frac{1}{\cosh x} \text{ for } x \in \mathbb{R}$$

Hyperbolic cosecant:

$$\text{csch} x = \frac{1}{\sinh x} \text{ for } x \neq 0$$

The exponential function can be interpreted as a certain limit which can be used as an alternative definition.

Theorem 1.7.2. *The exponential function satisfies*

$$e^x = \lim_{n\to\infty} \left(1 + \frac{x}{n}\right)^n$$

for any $x \in \mathbb{R}$.

Another important transcendental function is logarithm which is the inverse of the exponential function. Note that the exponential function has the property that for any $x > 0$, there exists a unique $y \in \mathbb{R}$ such that $e^y = x$.

Definition 1.7.3 (Logarithmic function)**.** *The* **logarithmic function** *is the function* $\ln : \mathbb{R}^+ \to \mathbb{R}$ *defined for* $x > 0$ *by*

$$y = \ln x \text{ if } e^y = x.$$

In other words, $\ln x$ *is the inverse function of the exponential function.*

The transcendental functions satisfy the following identities.

Proposition 1.7.4 (Identities for transcendental functions)**.**

1. *Exponential function:*
   (a) $e^{x+y} = e^x e^y$
   (b) $e^{x-y} = \dfrac{e^x}{e^y}$
   (c) $e^{kx} = (e^x)^k$ *for* $k \in \mathbb{Z}$
2. *Logarithmic function:*
   (a) $\ln(xy) = \ln x + \ln y$
   (b) $\ln \dfrac{x}{y} = \ln x - \ln y$
   (c) $\ln(x^k) = k \ln x$ *for* $k \in \mathbb{Z}$
3. *Trigonometric identities:*
   (a) $\cos^2 x + \sin^2 x = 1; \quad \sec^2 x - \tan^2 x = 1; \quad \csc^2 x - \cot^2 x = 1$
   (b) $\cos(-x) = \cos x; \quad \sin(-x) = -\sin x; \quad \tan(-x) = -\tan x$
   (c) $\cos(x+y) = \cos x \cos y - \sin x \sin y;$
   $\sin(x+y) = \sin x \cos y + \cos x \sin y;$
   $\tan(x+y) = \dfrac{\tan x + \tan y}{1 - \tan x \tan y}$
   (d) $\cos 2x = \cos^2 x - \sin^2 x = 2\cos^2 x - 1 = 1 - 2\sin^2 x;$
   $\sin 2x = 2 \sin x \cos x;$
   $\tan 2x = \dfrac{2 \tan x}{1 - \tan^2 x}$
4. *Hyperbolic identities:*
   (a) $\cosh^2 x - \sinh^2 x = 1; \quad \operatorname{sech}^2 x + \tanh^2 x = 1; \quad \coth^2 x - \operatorname{csch}^2 x = 1$
   (b) $\cosh(-x) = \cosh x; \quad \sinh(-x) = -\sinh x; \quad \tanh(-x) = -\tanh x$
   (c) $\cosh(x+y) = \cosh x \cosh y + \sinh x \sinh y;$
   $\sinh(x+y) = \sinh x \cosh y + \cosh x \sinh y;$
   $\tanh(x+y) = \dfrac{\tanh x + \tanh y}{1 + \tanh x \tanh y}$

(d) $\cosh 2x = \cosh^2 x + \sinh^2 x = 2\cosh^2 x - 1 = 1 + 2\sinh^2 x$;
$\sinh 2x = 2\sinh x \cosh x$;
$\tanh 2x = \dfrac{2\tanh x}{1 + \tanh^2 x}$

Proposition 1.7.5 (Derivatives of transcendental functions)**.**

1. *Exponential function:*
$$\frac{d}{dx} e^x = e^x$$
2. *Logarithmic function:*
$$\frac{d}{dx} \ln x = \frac{1}{x}$$
3. *Trigonometric functions:*
$$\frac{d}{dx}\cos x = -\sin x; \qquad \frac{d}{dx}\sin x = \cos x;$$
$$\frac{d}{dx}\tan x = \sec^2 x; \qquad \frac{d}{dx}\cot x = -\csc^2 x;$$
$$\frac{d}{dx}\sec x = \sec x \tan x; \qquad \frac{d}{dx}\csc x = -\csc x \cot x$$
4. *Inverse trigonometric functions*[7]*:*
$$\frac{d}{dx}\cos^{-1} x = -\frac{1}{\sqrt{1-x^2}};$$
$$\frac{d}{dx}\sin^{-1} x = \frac{1}{\sqrt{1-x^2}};$$
$$\frac{d}{dx}\tan^{-1} x = \frac{1}{1+x^2}$$

[7]Here we define the inverse trigonometric functions as follows.

(a) $\cos^{-1} : [-1,1] \to [0,\pi]$: For $x \in [-1,1]$, $y = \cos^{-1} x$ is the unique value $0 \le y \le \pi$ such that $\cos y = x$.

(b) $\sin^{-1} : [-1,1] \to [-\frac{\pi}{2}, \frac{\pi}{2}]$: For $x \in [-1,1]$, $y = \sin^{-1} x$ is the unique value $-\frac{\pi}{2} \le y \le \frac{\pi}{2}$ such that $\sin y = x$.

(c) $\tan^{-1} : \mathbb{R} \to (-\frac{\pi}{2}, \frac{\pi}{2})$: For $x \in \mathbb{R}$, $y = \tan^{-1} x$ is the unique value $-\frac{\pi}{2} < y < \frac{\pi}{2}$ such that $\tan y = x$.

5. *Hyperbolic functions:*

$$\frac{d}{dx}\cosh x = \sinh x; \qquad \frac{d}{dx}\sinh x = \cosh x;$$
$$\frac{d}{dx}\tanh x = \operatorname{sech}^2 x; \qquad \frac{d}{dx}\coth x = -\operatorname{csch}^2 x;$$
$$\frac{d}{dx}\operatorname{sech} x = -\operatorname{sech} x \tanh x; \quad \frac{d}{dx}\operatorname{csch} x = -\operatorname{csch} x \coth x$$

6. *Inverse hyperbolic functions*[8]*:*

$$\frac{d}{dx}\cosh^{-1} x = \frac{1}{\sqrt{x^2-1}};$$
$$\frac{d}{dx}\sinh^{-1} x = \frac{1}{\sqrt{x^2+1}};$$
$$\frac{d}{dx}\tanh^{-1} x = \frac{1}{1-x^2}$$

Proposition 1.7.6 (Integrals of transcendental functions)**.**

1. *Exponential function:*

$$\int e^x dx = e^x + C$$

2. *Logarithmic function:*

$$\int \frac{1}{x} dx = \ln|x| + C$$

[8] The inverse hyperbolic functions can be expressed in terms of logarithm as follows.

(a) $\cosh^{-1} : [1, +\infty) \to [0, +\infty)$: $\cosh^{-1} x = \ln(x + \sqrt{x^2-1})$.

(b) $\sinh^{-1} : \mathbb{R} \to \mathbb{R}$: $\sinh^{-1} x = \ln(x + \sqrt{x^2+1})$.

(c) $\tanh^{-1} : (-1, 1) \to \mathbb{R}$: $\tanh^{-1} x = \frac{1}{2}\ln(\frac{1+x}{1-x})$.

3. *Trigonometric functions:*

$$\int \cos x dx = \sin x + C; \qquad \int \sin x dx = -\cos x + C;$$

$$\int \tan x dx = \ln \sec x; \qquad \int \cot x = \ln \sin x + C;$$

$$\int \sec x dx = \ln|\sec x + \tan x| + C; \qquad \int \csc x dx = \ln|\csc x - \cot x| + C$$

4. *Hyperbolic functions:*

$$\int \cosh x dx = \sinh x + C; \qquad \int \sinh x dx = \cosh x + C;$$

$$\int \tanh x dx = \ln \cosh x; \qquad \int \coth x = \ln \sinh x + C;$$

$$\int \text{sech} x dx = \tan^{-1} \sinh x + C; \qquad \int \text{csch} x dx = \ln|\text{csch} x - \coth x| + C$$

Exercise 1

1. Let $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$. Prove the polarization identity

$$\langle \mathbf{u}, \mathbf{v} \rangle = \frac{1}{4}\left(\|\mathbf{u}+\mathbf{v}\|^2 - \|\mathbf{u}-\mathbf{v}\|^2\right)$$

2. Let $\mathbf{u}, \mathbf{v} \in \mathbb{R}^3$. Prove that if $\langle \mathbf{u}, \mathbf{w} \rangle = \langle \mathbf{v}, \mathbf{w} \rangle$ for any $\mathbf{w} \in \mathbb{R}^3$, then $\mathbf{u} = \mathbf{v}$.

3. Prove that for any $\mathbf{u}, \mathbf{v} \in \mathbb{R}^3$, we have $\mathbf{u} \times \mathbf{v}$ is orthogonal to both $\mathbf{u}$ and $\mathbf{v}$.

4. Prove that for any $\mathbf{u}, \mathbf{v} \in \mathbb{R}^3$, we have

$$\|\mathbf{u} \times \mathbf{v}\|^2 + \langle \mathbf{u}, \mathbf{v} \rangle^2 = \|\mathbf{u}\|^2 \|\mathbf{v}\|^2$$

5. Let $\mathbf{u}(t), \mathbf{v}(t)$ be two differentiable vector valued functions. Prove that

$$\frac{d}{dt}\langle \mathbf{u}, \mathbf{v} \rangle = \langle \frac{d\mathbf{u}}{dt}, \mathbf{v} \rangle + \langle \mathbf{u}, \frac{d\mathbf{v}}{dt} \rangle$$

6. Let $\mathbf{v}(t)$ be a differentiable vector valued function. Suppose $\|\mathbf{v}\|$ is a constant independent of t. Prove that $\dfrac{d\mathbf{v}}{dt}$ is orthogonal to $\mathbf{v}$ for any t.

7. Let $\mathbf{u}, \mathbf{v}, \mathbf{w} \in \mathbb{R}^3$.

   (a) Prove that

   $$\langle \mathbf{u} \times (\mathbf{v} \times \mathbf{w}), \mathbf{x} \rangle = \langle \mathbf{u}, \mathbf{w} \rangle \langle \mathbf{v}, \mathbf{x} \rangle - \langle \mathbf{u}, \mathbf{v} \rangle \langle \mathbf{w}, \mathbf{x} \rangle$$

   for any $\mathbf{x} \in \mathbb{R}^3$. (Hint: use $\langle \mathbf{u}_1 \times \mathbf{v}_1, \mathbf{u}_2 \times \mathbf{v}_2 \rangle = \begin{vmatrix} \langle \mathbf{u}_1, \mathbf{u}_2 \rangle & \langle \mathbf{u}_1, \mathbf{v}_2 \rangle \\ \langle \mathbf{v}_1, \mathbf{u}_2 \rangle & \langle \mathbf{v}_1, \mathbf{v}_2 \rangle \end{vmatrix}$)

   (b) Prove that

   $$\mathbf{u} \times (\mathbf{v} \times \mathbf{w}) = \langle \mathbf{u}, \mathbf{w} \rangle \mathbf{v} - \langle \mathbf{u}, \mathbf{v} \rangle \mathbf{w}$$

   (c) Prove the Jacobi identity

   $$\mathbf{u} \times (\mathbf{v} \times \mathbf{w}) + \mathbf{v} \times (\mathbf{w} \times \mathbf{u}) + \mathbf{w} \times (\mathbf{u} \times \mathbf{v}) = \mathbf{0}$$

8. Let $Q = [\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3]$ be a 3×3 matrix where $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$ are the column vectors of Q. Show that Q is an orthogonal matrix, that is $Q^{-1} = Q^T$ if and only if $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$ constitute an orthonormal basis for $\mathbb{R}$

9. Let A be an $n \times n$ matrix.

   (a) Prove that

   $$\langle \mathbf{u}, A\mathbf{v} \rangle = \langle A^T \mathbf{u}, \mathbf{v} \rangle$$

   for any $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$.

   (b) Prove that if

   $$\langle A\mathbf{u}, A\mathbf{v} \rangle = \langle \mathbf{u}, \mathbf{v} \rangle$$

   for any $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$, then A is an orthogonal matrix.

10. Prove the following hyperbolic identities.

    (a) $\cosh^2 x - \sinh^2 x = 1$

    (b) $\cosh(x + y) = \cosh x \cosh y + \sinh x \sinh y$

    (c) $\sinh(x + y) = \cosh x \sinh y + \sinh x \cosh y$

11. Prove that

(a) $\dfrac{d}{dx}\cosh x = \sinh x$

(b) $\dfrac{d}{dx}\sinh x = \cosh x$

(c) $\dfrac{d}{dx}\tanh x = \dfrac{1}{\cosh^2 x}$

2 Curves

2.1 Regular parametrized curves

The subject we are going to study in this chapter is curves in two or three dimensional Euclidean space. The first problem is how do we define curves. One may say that curves are one dimensional objects in the Euclidean space. However it is not easy to define what dimension is for a general subset of the Euclidean space. Moreover we would also like the curves to be sufficiently smooth. In differential geometry, this is done by considering regular parametrized curves. Intuitively, it is the trajectory of a moving particle.

Definition 2.1.1 (Regular parametrized curves)**.** *A* **regular parametrized curve** *is a differentiable function* $\mathbf{r} : (a, b) \to \mathbb{R}^n$*,* $n = 2$ *or* 3*, such that* $\mathbf{r}'(t) \neq \mathbf{0}$ *for any* $t \in (a, b)$*.*

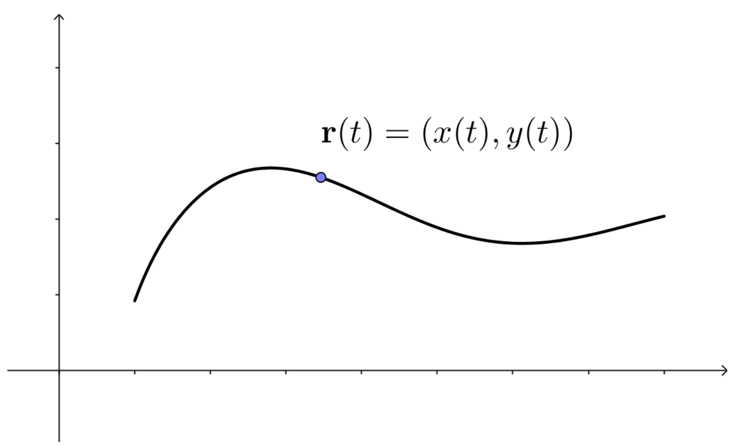

Figure 1: Regular parametrized curve

In daily language, curve usually refers to a collection of points. Here, by parametrized curve, we mean a function from an open interval to $\mathbb{R}^2$ or $\mathbb{R}^3$. However, if two such functions have the same image, we may also consider them to be the same as curves and say that the two functions are two different parametrization of the curve.

One usually requires, though not necessary, the function defining a parametrized curve to be injective[9]. However when we consider a closed curve, e.g. a cir-

[9]A function f is injective if $f(x) = f(y)$ implies $x = y$. In other words, any two distinct elements in the domain of f cannot have the same image.

cle, which has no end points, we will need more than one parametrization function.

A curve on $\mathbb{R}^2$ is called a **plane curve** and a curve in $\mathbb{R}^3$ is called a **space curve**. The requirement $\mathbf{r}'(t) \neq \mathbf{0}$ guarantees that the curve does not have a sharp turning point and tangent to the curve can be defined everywhere.

Example 2.1.2.

1. *Straight line: Let (x_0, y_0) and (x_1, y_1) be two points on $\mathbb{R}^2$. The function*

$$\mathbf{r}(t) = ((1-t)x_0 + tx_1, (1-t)y_0 + ty_1), \text{ for } 0 < t < 1$$

defines a regular plane curve which is a **straight line segment** *joining (x_0, y_0) and (x_1, y_1). Yes, a straight line on the plane is a curve.*

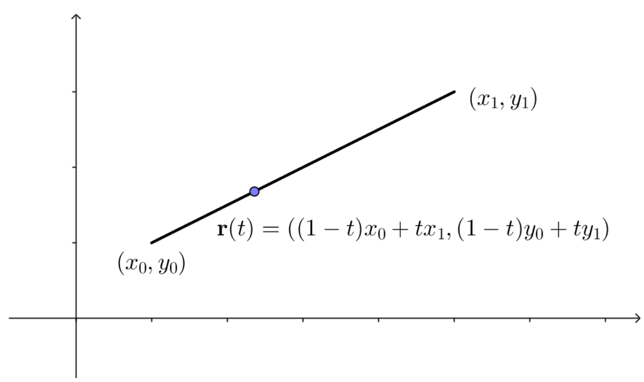

Figure 2: Straight line segment

2. *Circle: Let $r > 0$ be a positive real number. The function*

$$\mathbf{r}(\theta) = (r\cos\theta, r\sin\theta), \text{ for } 0 < \theta < 2\pi$$

defines a **circle** *with radius r centered at the origin.*

3. *Cycloid: The function*

$$\mathbf{r}(\theta) = (\theta - \sin\theta, 1 - \cos\theta), \text{ for } 0 < \theta < 2\pi$$

defines a curve which is called a **cycloid**.

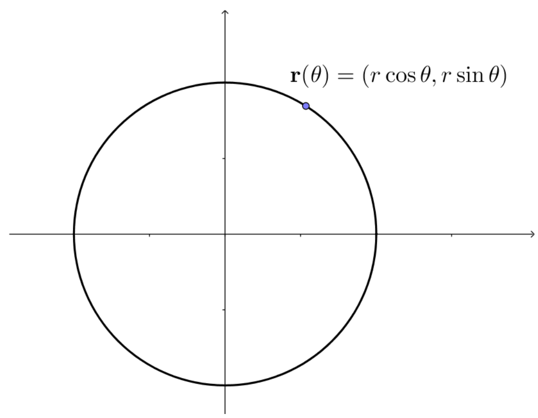

Figure 3: Circle

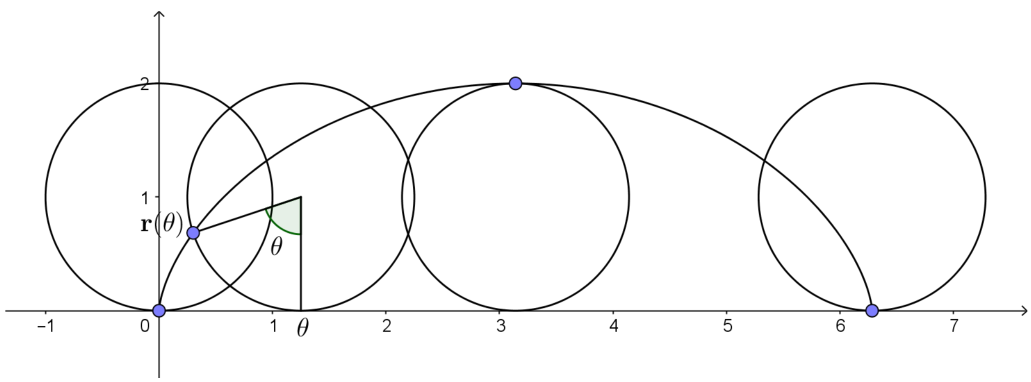

Figure 4: Cycloid

4. *Helix: The function*

$$\mathbf{r}(\theta) = (a\cos\theta, a\sin\theta, b\theta), \text{ for } \theta \in \mathbb{R}$$

defines a curve which is called a **helix**.

The following example illustrates a curve which has an irregular point.

Example 2.1.3. *Let* $\mathbf{r}(t) = (t^2, t^3)$. *Then* $\mathbf{r}'(t) = (2t, 3t^2)$ *and* $\mathbf{r}'(0) = (0, 0)$. *Therefore* $\mathbf{r}(t)$ *is not regular at* $t = 0$.

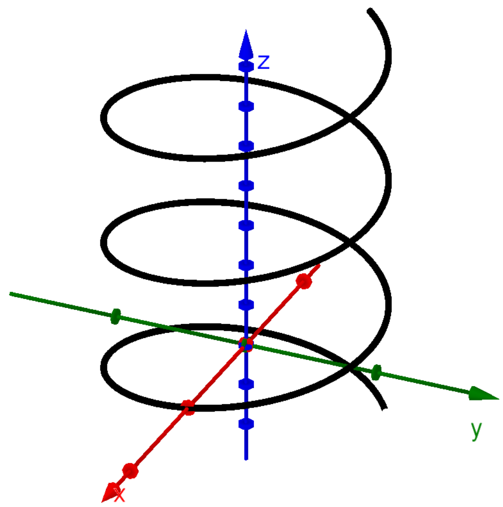

Figure 5: Helix

2.2 Arc length

The first geometric quantity associated with a curve we study is its arc length. It is a generalization of lengths of line segments to curves. The idea is to cut a curve into n small pieces and approximate each piece with a small line segment. Then sum up the lengths of the line segments and let n goes to infinity. This is obtained by an integral.

Definition 2.2.1 (Arc length)**.** *Let* $\mathbf{r} : (a, b) \to \mathbb{R}^n$ *be a regular parametrized curve. Then the* **arc length** *of* $\mathbf{r}$ *is defined by*

$$l = \int_a^b \|\mathbf{r}'(t)\| dt.$$

The first thing we check is that arc length is really a generalization of length of line segment.

Example 2.2.2 (Arc length of line segments)**.** *Let*

$$\mathbf{r}(t) = ((1 - t)x_0 + tx_1, (1 - t)y_0 + ty_1),\ 0 < t < 1,$$

be the line segment joining (x_0, y_0) *and* (x_1, y_1)*. Now*

$$\begin{aligned} \mathbf{r}'(t) &= (x_1 - x_0, y_1 - y_0) \\ \|\mathbf{r}'(t)\| &= \sqrt{(x_1 - x_0)^2 + (y_1 - y_0)^2} \end{aligned}$$

Thus the arc length of $\mathbf{r}$ *is*

$$\begin{aligned} \int_0^1 \|\mathbf{r}'(t)\| dt &= \int_0^1 \sqrt{(x_1 - x_0)^2 + (y_1 - y_0)^2} dt \\ &= \sqrt{(x_1 - x_0)^2 + (y_1 - y_0)^2} \end{aligned}$$

which is exactly the length of line segment joining (x_0, y_0) *and* (x_1, y_1)*.*

The arc length of a circle of radius r is known to be $2\pi r$ even for primary students. Now may give a rigorous proof for this simple fact.

Example 2.2.3 (Arc length of circles)**.** *Let* $\mathbf{r}(\theta) = (r\cos\theta, r\sin\theta)$*,* $0 < \theta < 2\pi$*, be the circle with radius* $r > 0$ *centered at the origin. Now*

$$\begin{aligned} \mathbf{r}'(t) &= (-r\sin\theta, r\cos\theta) \\ \|\mathbf{r}'(t)\| &= \sqrt{r^2\sin^2\theta + r^2\cos^2\theta} \\ &= r \end{aligned}$$

Thus the arc length of $\mathbf{r}$ *is*

$$\begin{aligned} \int_0^{2\pi} \|\mathbf{r}'(t)\| dt &= \int_0^{2\pi} r dt \\ &= 2\pi r \end{aligned}$$

Proposition 2.2.4 (Arc length of graphs of functions)**.**

1. *(Rectangular coordinates): The arc length of the curve given by the graph of function* $y = f(x)$*,* $a < x < b$*, in rectangular coordinates is*

$$l = \int_a^b \sqrt{1 + f'^2} dx.$$

2. *(Polar coordinates): The arc length of the curve given by the graph of function* $r = r(\theta)$*,* $\alpha < \theta < \beta$*, in polar coordinates is*

$$l = \int_\alpha^\beta \sqrt{r^2 + r'^2} d\theta.$$

Proof. 1. Parametrized the graph of $y = f(x)$ by $\mathbf{r}(t) = (t, f(t))$, $a < t < b$. Then

$$\begin{aligned} \mathbf{r}'(t) &= (1, f') \\ \|\mathbf{r}'(t)\| &= \sqrt{1 + f'^2} \end{aligned}$$

Therefore the arc length is

$$l = \int_a^b \|\mathbf{r}'(t)\| dt = \int_a^b \sqrt{1 + f'^2} dx.$$

2. Parametrized the graph of $r = r(\theta)$ by $\mathbf{r}(\theta) = (r(\theta)\cos\theta, r(\theta)\sin\theta)$, $\alpha < \theta < \beta$. The rest is left for the reader as exercise.

□

Definition 2.2.5 (Arc length parametrization)**.** *We say that* $\mathbf{r}(s)$ *is an* **arc length parametrized** *curve, or* $\mathbf{r}(s)$ *is* **parametrized by arc length***, if* $\|\mathbf{r}'(s)\| = 1$ *for any* s*.*

Using arc length parametrization has a lot of advantage. For example, it makes calculating the arc length of a curve very easy.

Proposition 2.2.6. *Let* $\mathbf{r}(s)$*, be an arc length parametrized curve. Then for* $a < b$*, the arc length of* $\mathbf{r}(s)$ *from* $s = a$ *to* $s = b$ *is* $b - a$*.*

Proof. Since $\mathbf{r}(s)$ is an arc length parametrization, we have $\|\mathbf{r}'(s)\| = 1$. Therefore the arc length from $s = a$ to $s = b$ is

$$\int_a^b \|\mathbf{r}'(s)\| ds = \int_a^b ds = [s]_a^b = b - a.$$

□

There are many other geometric quantities which are easier to be calculated using arc length parametrization. We may also use arc length parametrization to prove certain statements concerning curves because it always exists and is unique.

Theorem 2.2.7 (Existence and uniqueness of arc length parametrization)**.** *Let* $\mathbf{r}(t)$ *be a regular parametrized curve. Then there exists increasing differentiable function* $s = s(t)$ *such that when* $\mathbf{r}(s)$ *is considered as a function of* s*, it is an arc length parametrized curve. Moreover if* $s_1(t)$ *and* $s_2(t)$ *are two such functions, then* $s_2 - s_1$ *is a constant.*

Proof. Let

$$s(t) = \int_\alpha^t \|\mathbf{r}'(u)\| du.$$

By fundamental theorem of calculus, we have $s'(t) = \|\mathbf{r}'(t)\|$. Then when $\mathbf{r}(s)$ is considered as a function of s and by chain rule, we obtain

$$\frac{d\mathbf{r}}{dt} = \frac{ds}{dt}\frac{d\mathbf{r}}{ds} = \|\mathbf{r}'(t)\| \frac{d\mathbf{r}}{ds}.$$

Thus $\dfrac{d\mathbf{r}}{ds}$ is an unit vector which means $\mathbf{r}(s)$ is an arc length parametrization.

Suppose $s_1(t), s_2(t)$ are two increasing differentiable functions such that $\mathbf{r}(s_1)$ and $\mathbf{r}(s_2)$ are arc length parametrizations. Then

$$\frac{ds_2}{dt}\frac{d\mathbf{r}}{ds_2} = \frac{d\mathbf{r}}{dt} = \frac{ds_1}{dt}\frac{d\mathbf{r}}{ds_1}$$

which implies

$$\left|\frac{ds_2}{dt}\right| = \left|\frac{ds_1}{dt}\right|.$$

Since both $s_1(t), s_2(t)$ are increasing functions, we have $\dfrac{ds_2}{dt} = \dfrac{ds_1}{dt}$ and it follows that $s_2 - s_1$ is a constant. □

To find the arc length parametrization of $\mathbf{r}(t)$, we do the following three steps.

1. Find the arc length $s(t)$ as a function of t by

$$s(t) = \int_a^t \|\mathbf{r}(u)\| du.$$

2. Express $t = t(s)$ in terms of s. In other words, make t the subject in $s = s(t)$.

3. Substitute $t(s)$ into t in $\mathbf{r}(t)$ to get the arc length parametrization $\mathbf{r}(s)$.

Example 2.2.8 (Arc length parametrization of helix)**.** *Let $a, b > 0$ be constants. Find an arc length parametrization of the helix $\mathbf{r}(\theta) = (a\cos\theta, a\sin\theta, b\theta)$.*

Solution. We have

$$\begin{aligned} \mathbf{r}'(\theta) &= (-a\sin\theta, a\cos\theta, b) \\ \|\mathbf{r}'(\theta)\| &= \sqrt{a^2+b^2} \end{aligned}$$

If we let $c = \sqrt{a^2+b^2}$ and

$$s(\theta) = \int_0^\theta \|\mathbf{r}'(t)\| dt = \int_0^\theta c dt = c\theta,$$

then

$$\theta = \frac{s}{c}$$

and

$$\mathbf{r}(s) = \left(a\cos\frac{s}{c}, a\sin\frac{s}{c}, \frac{b}{c}s\right)$$

is an arc length parametrization of the helix. □

Example 2.2.9 (Arc length parametrization of catenary)**.** *Find an arc length parametrization of the catenary* $\mathbf{r}(t) = (t, \cosh t)$*.*

Solution. We have

$$\begin{aligned} \mathbf{r}'(t) &= (1, \sinh t) \\ \|\mathbf{r}'(\theta)\| &= \sqrt{1+\sinh^2 t} = \cosh t \end{aligned}$$

Let

$$s(t) = \int_0^t \|\mathbf{r}'(u)\| du = \int_0^t \cosh u du = \sinh t,$$

then

$$t = \sinh^{-1} s = \ln(s + \sqrt{s^2+1})$$

and

$$\begin{aligned} \mathbf{r}(s) &= (\ln(s+\sqrt{s^2+1}), \cosh\ln(s+\sqrt{s^2+1})) \\ &= (\ln(s+\sqrt{s^2+1}), \cosh t) \\ &= (\ln(s+\sqrt{s^2+1}), \sqrt{s^2+1}) \end{aligned}$$

is an arc length parametrization of the catenary. □

Example 2.2.10 (Tractrix). *The* **tractrix** *is a curve parametrized by*

$$\mathbf{r}(t) = (\text{sech}t, t - \tanh t),\ \ t > 0.$$

Find the arc length parametrization of the tractrix.
Note: The tractrix may also be parametrized by

$$\mathbf{r}(\theta) = \left(\sin\theta, \ln\left(\cot\frac{\theta}{2}\right) - \cos\theta\right),\ \ 0 < \theta < \frac{\pi}{2}.$$

Suppose L is the tangent to the tractrix at $\mathbf{r}(\theta)$ and P is the point of intersection of L and the y-axis. Then the angle between L and the y-axis is θ and the distance between $\mathbf{r}(\theta)$ and P is always 1.

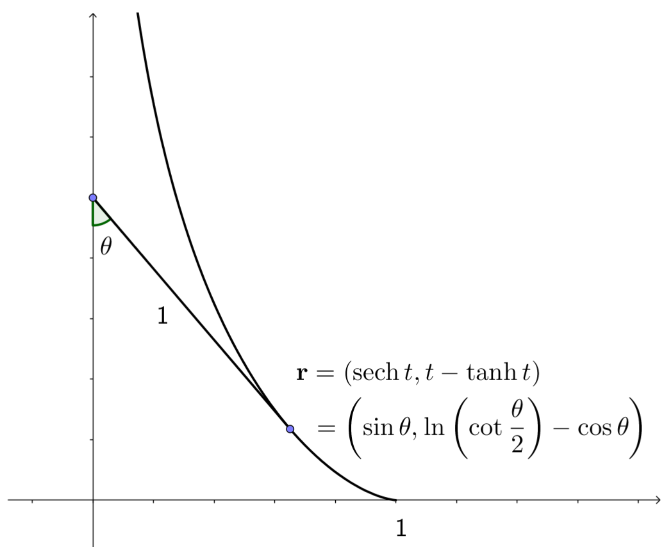

Figure 6: Tractrix

Solution. We have

$$\begin{aligned} \mathbf{r}'(t) &= (-\text{sech}t\tanh t, 1 - \text{sech}^2 t) = (-\text{sech}t\tanh t, \tanh^2 t) \\ \|\mathbf{r}'(t)\| &= \tanh t\sqrt{\text{sech}^2 t + \tanh^2 t} = \tanh t. \end{aligned}$$

Thus the arc length function is

$$\begin{aligned} s(t) &= \int_0^t \tanh u du \\ &= \int_0^t \frac{\sinh u}{\cosh u} du \\ &= \int_0^t \frac{d\cosh u}{\cosh u} \\ &= [\ln\cosh u]_0^t \\ &= \ln\cosh t \end{aligned}$$

Thus

$$e^s = \cosh t.$$

Now

$$\begin{aligned} x(s) &= \text{sech} t \\ &= \frac{1}{\cosh t} \\ &= e^{-s} \\ y(s) &= t - \tanh t \\ &= \cosh^{-1} e^s - \sqrt{1 - \text{sech}^2 t} \\ &= \ln(e^s + \sqrt{e^{2s} - 1}) - \sqrt{1 - e^{-2s}}. \end{aligned}$$

Therefore the arc length parametrization is

$$\mathbf{r}(s) = (e^{-s}, \ln(e^s + \sqrt{e^{2s} - 1}) - \sqrt{1 - e^{-2s}}), \;\; s > 0.$$

□

Although arc length always exists for any curve, it is in general very difficult to write it down explicitly.

We conclude this section by proving a simple geometric fact that straight line is the shortest curve joining two given points.

Theorem 2.2.11. *Let $\mathbf{r}(t)$ be a regular parametrized curve with $\mathbf{r}(a) = \mathbf{r}_0$ and $\mathbf{r}(b) = \mathbf{r}_1$. Then the arc length l of the curve from $t = a$ to $t = b$ satisfies*

$$l \geq \|\mathbf{r}_1 - \mathbf{r}_0\|$$

with equality holds if and only if $\mathbf{r}(t)$ is a line segment joining $\mathbf{r}_0$ and $\mathbf{r}_1$.

Proof. Let

$$\mathbf{a} = \frac{\mathbf{r}_1 - \mathbf{r}_0}{\|\mathbf{r}_1 - \mathbf{r}_0\|}.$$

Since $\mathbf{a}$ is a unit vector, we have $\langle \mathbf{r}'(t), \mathbf{a} \rangle \leq \|\mathbf{r}'(t)\|$ for any t and the arc length of the curve satisfies

$$\begin{aligned} l &= \int_a^b \|\mathbf{r}'(t)\| dt \\ &\geq \int_a^b \langle \mathbf{r}'(t), \mathbf{a} \rangle dt \\ &= \langle \mathbf{r}(b) - \mathbf{r}(a), \mathbf{a} \rangle \\ &= \langle \mathbf{r}_1 - \mathbf{r}_0, \mathbf{a} \rangle \\ &= \|\mathbf{r}_1 - \mathbf{r}_0\|. \end{aligned}$$

The equality holds if and only if

$$\mathbf{r}'(t) = \langle \mathbf{r}'(t), \mathbf{a} \rangle \mathbf{a}$$

for any t which means

$$\mathbf{r}'(t) = \alpha(t)\mathbf{a}$$

for some positive valued function $\alpha(t)$. Therefore

$$\mathbf{r}(t) = \beta(t)\mathbf{a}$$

where $\beta(t)$ is a differentiable function such that $\beta'(t) = \alpha(t)$, which implies that $\mathbf{r}(t)$ is a straight line segment. □

2.3 Curve curvature

The curvature of a curve describes how rapidly it is bending. To make a rigorous definition, we need the unit tangent and unit normal vectors to the curve.

Definition 2.3.1 (Unit tangent and normal vector)**.** *Let* $\mathbf{r}(t)$ *be a regular parametrized curve.*

1. *The* **unit tangent vector** *to the curve at* $\mathbf{r}(t)$ *is defined by*

$$\mathbf{T}(t) = \frac{\mathbf{r}'(t)}{\|\mathbf{r}'(t)\|}.$$

In particular if $\mathbf{r}(s)$ *is an arc length parametrization, then*

$$\mathbf{T}(s) = \mathbf{r}'(s).$$

2. *Suppose* $\mathbf{T}'(t) \neq 0$. *We define the* **unit normal vector** *to the curve at* $\mathbf{r}(t)$ *by*

$$\mathbf{N}(t) = \frac{\mathbf{T}'(t)}{\|\mathbf{T}'(t)\|}.$$

In particular if $\mathbf{r}(s)$ *is an arc length parametrization, then*

$$\mathbf{N}(s) = \frac{\mathbf{T}'(s)}{\|\mathbf{T}'(s)\|} = \frac{\mathbf{r}''(s)}{\|\mathbf{r}''(s)\|}.$$

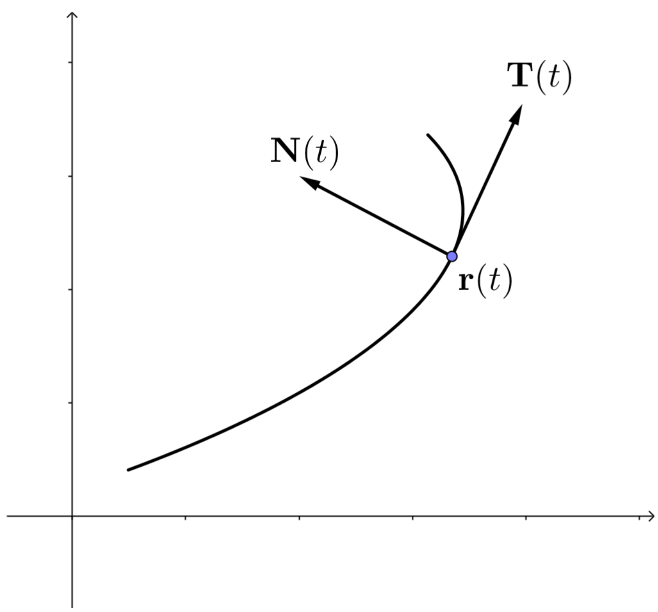

Figure 7: Unit tangent and unit normal vector

We give some useful formulas for calculation.

Proposition 2.3.2. *Let* $\mathbf{r}(t)$ *be a regular parametrized curve and* $\mathbf{N}(t)$ *be the unit normal vector. We have*

1. $\dfrac{d}{dt}\|\mathbf{r}'\| = \dfrac{\langle \mathbf{r}', \mathbf{r}'' \rangle}{\|\mathbf{r}'\|}$

2. $\mathbf{T}' = \dfrac{\mathbf{r}''}{\|\mathbf{r}'\|} - \dfrac{\langle \mathbf{r}', \mathbf{r}''\rangle}{\|\mathbf{r}'\|^3}\mathbf{r}'$

Proof. 1. We have

$$\begin{aligned}\frac{d}{dt}\|\mathbf{r}'\| &= \frac{d}{dt}\sqrt{\langle \mathbf{r}', \mathbf{r}'\rangle} \\ &= \frac{\langle \mathbf{r}'', \mathbf{r}'\rangle + \langle \mathbf{r}', \mathbf{r}''\rangle}{2\sqrt{\langle \mathbf{r}', \mathbf{r}'\rangle}} \\ &= \frac{\langle \mathbf{r}', \mathbf{r}''\rangle}{\|\mathbf{r}'\|}\end{aligned}$$

2. We have

$$\begin{aligned}\mathbf{T}' &= \frac{\|\mathbf{r}'\|\mathbf{r}'' - (\frac{d}{dt}\|\mathbf{r}'\|)\mathbf{r}'}{\|\mathbf{r}'\|^2} \\ &= \frac{\mathbf{r}''}{\|\mathbf{r}'\|} - \frac{\langle \mathbf{r}', \mathbf{r}''\rangle}{\|\mathbf{r}'\|^3}\mathbf{r}'\end{aligned}$$

□

Now we define the curvature of a curve. There are many different ways to write down its definition. Here we define it as the magnitude of the derivative of the unit tangent vector with respect to arc length.

Definition 2.3.3 (Curve curvature)**.** *Let* $\mathbf{r}(t)$ *be a regular parametrized curve and* $\mathbf{T}(t)$ *be the unit tangent to the curve at* $\mathbf{r}(t)$. *Then the* **curvature** *of the curve at* $\mathbf{r}(t)$ *is*

$$\kappa(t) = \frac{\|\mathbf{T}'(t)\|}{\|\mathbf{r}'(t)\|}.$$

In particular if $\mathbf{r}(s)$ *is an arc length parametrized curve, the curvature is*

$$\kappa(s) = \|\mathbf{T}'(s)\|$$

Curvature is a geometric property which is used to measure how much 'banding' a curve has. The first thing we expect is that a curve has zero curvature if and only if it is a straight line segment.

Proposition 2.3.4. *Let* $\mathbf{r}(t)$ *be a regular parametrized curve. Then the curvature satisfies* $\kappa(t) = 0$ *for any* $a < t < b$ *if and only if* $\mathbf{r}(t)$ *is a straight line segment joining* $\mathbf{r}_0$ *and* $\mathbf{r}_1$.

Proof. Suppose $\mathbf{r}$ is a straight line segment. Then

$$\mathbf{r}(t) = \mathbf{a} + \alpha(t)\mathbf{b}$$

for some increasing function $\alpha(t)$ and constant vector $\mathbf{a}$, $\mathbf{b}$ with $\|\mathbf{b}\| = 1$. Thus

$$\mathbf{T} = \frac{\mathbf{r}'}{\|\mathbf{r}'\|} = \frac{\alpha'\mathbf{b}}{\|\alpha'\mathbf{b}\|} = \mathbf{b}$$

is a constant unit vector which implies $\mathbf{T}'(t) = \mathbf{0}$ for any $a < t < b$. Therefore

$$\kappa(t) = \frac{\|\mathbf{T}'(t)\|}{\|\mathbf{r}'(t)\|} = 0$$

for any $a < t < b$.

Conversely, suppose $\kappa(t) = 0$ for any $a < t < b$. Then

$$\mathbf{T}'(t) = \mathbf{0}$$

for any $a < t < b$ which implies

$$\frac{\mathbf{r}'(t)}{\|\mathbf{r}'(t)\|} = \mathbf{b}$$

is a constant unit vector. Thus

$$\mathbf{r}'(t) = \alpha(t)\mathbf{b}$$

for some positive valued function $\alpha(t)$ and therefore

$$\mathbf{r}(t) = \mathbf{a} + \beta(t)\mathbf{b}$$

where $\beta(t)$ is a differentiable function such that $\beta'(t) = \alpha(t)$ and $\mathbf{a}$ is a constant vector. It follows that $\mathbf{r}$ is a straight line segment. □

Now we give the formula for finding the curvature of plane and space curves.

Proposition 2.3.5 (Formulas for curvature)**.** *Let* $\mathbf{r}(t)$ *be a regular parametrized curve.*

1. *Suppose* $\mathbf{r}(t) = (x(t), y(t))$ *is a plane curve. Then*

$$\kappa(t) = \frac{|x'y'' - x''y'|}{(x'^2 + y'^2)^{\frac{3}{2}}}.$$

2. *Suppose* $\mathbf{r}(t) = (x(t), y(t), z(t))$ *is a space curve. Then*

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

Proof. Let $\mathbf{r}(t)$ be a regular parametrized curve. Then

$$\mathbf{T} = \frac{\mathbf{r}'}{\|\mathbf{r}'\|}$$

By Proposition 2.3.2,

$$\frac{d}{dt}\|\mathbf{r}'\| = \frac{\langle \mathbf{r}', \mathbf{r}'' \rangle}{\|\mathbf{r}'\|}.$$

Thus we have

$$\mathbf{T}' = \frac{\|\mathbf{r}'\|\mathbf{r}'' - (\frac{d}{dt}\|\mathbf{r}'\|)\mathbf{r}'}{\|\mathbf{r}'\|^2} = \frac{\|\mathbf{r}'\|^2\mathbf{r}'' - \langle \mathbf{r}', \mathbf{r}'' \rangle \mathbf{r}'}{\|\mathbf{r}'\|^3}.$$

Therefore the curvature is

$$\kappa = \frac{\|\mathbf{T}'\|}{\|\mathbf{r}'\|} = \left\| \frac{\|\mathbf{r}'\|^2\mathbf{r}'' - \langle \mathbf{r}', \mathbf{r}'' \rangle \mathbf{r}'}{\|\mathbf{r}'\|^4} \right\|.$$

1. Suppose $\mathbf{r}(t) = (x(t), y(t))$ is a plane curve. Then

$$\begin{aligned}
\mathbf{r}' &= (x', y') \\
\mathbf{r}'' &= (x'', y'') \\
\langle \mathbf{r}', \mathbf{r}'' \rangle &= x'x'' + y'y'' \\
\|\mathbf{r}'\|^2\mathbf{r}'' - \langle \mathbf{r}', \mathbf{r}'' \rangle \mathbf{r}' &= (x'^2 + y'^2)(x'', y'') - (x'x'' + y'y'')(x', y') \\
&= (y'^2x'' - x'y'y'', x'^2y'' - x'y'x'') \\
&= (x'y'' - x''y')(-y', x')
\end{aligned}$$

Therefore the curvature of $\mathbf{r}$ is

$$\kappa = \frac{|x'y'' - x''y'|\sqrt{y'^2 + x'^2}}{(\sqrt{x'^2 + y'^2})^4} = \frac{|x'y'' - x''y'|}{(x'^2 + y'^2)^{\frac{3}{2}}}.$$

2. Suppose $\mathbf{r}(t) = (x(t), y(t), z(t))$ is a space curve. Then

$$\begin{aligned} & \left\| \|\mathbf{r}'\|^2\mathbf{r}'' - \langle \mathbf{r}', \mathbf{r}''\rangle \mathbf{r}' \right\|^2 \\ = & \langle \|\mathbf{r}'\|^2\mathbf{r}'' - \langle \mathbf{r}', \mathbf{r}''\rangle \mathbf{r}', \|\mathbf{r}'\|^2\mathbf{r}'' - \langle \mathbf{r}', \mathbf{r}''\rangle \mathbf{r}'\rangle \\ = & \|\mathbf{r}'\|^4\|\mathbf{r}''\|^2 - 2\langle \mathbf{r}', \mathbf{r}''\rangle^2\|\mathbf{r}'\|^2 + \langle \mathbf{r}', \mathbf{r}''\rangle^2\|\mathbf{r}'\|^2 \\ = & \|\mathbf{r}'\|^4\|\mathbf{r}''\|^2 - \langle \mathbf{r}', \mathbf{r}''\rangle^2\|\mathbf{r}'\|^2 \\ = & \|\mathbf{r}'\|^2(\|\mathbf{r}'\|^2\|\mathbf{r}''\|^2 - \langle \mathbf{r}', \mathbf{r}''\rangle^2) \\ = & \|\mathbf{r}'\|^2\|\mathbf{r}' \times \mathbf{r}''\|^2. \end{aligned}$$

Therefore the curvature is

$$\begin{aligned} \kappa &= \frac{\|\mathbf{r}'\|\|\mathbf{r}' \times \mathbf{r}''\|}{\|\mathbf{r}'\|^4} \\ &= \frac{\|\mathbf{r}' \times \mathbf{r}''\|}{\|\mathbf{r}'\|^3}. \end{aligned}$$

□

If $\mathbf{r}(s)$ is an arc length parametrized curve, we have a simple formula to calculate the curvature.

Theorem 2.3.6. *Suppose* $\mathbf{r}(s)$ *is an arc length parametrized curve. Then*

1. $\kappa(s) = \|\mathbf{r}''(s)\|$
2. $\mathbf{T}'(s) = \kappa(s)\mathbf{N}(s)$

Proof. Since $\mathbf{r}(s)$ is an arc length parametrization, we have $\|\mathbf{r}'(s)\| = 1$ and

$$\begin{cases} \mathbf{T}(s) = \mathbf{r}'(s) \\ \mathbf{T}'(s) = \mathbf{r}''(s). \end{cases}$$

1. Now the curvature is

$$\begin{aligned} \kappa(s) &= \frac{\|\mathbf{T}'(s)\|}{\|\mathbf{r}'(s)\|} \\ &= \|\mathbf{r}''(s)\|. \end{aligned}$$

2. We also have

$$\mathbf{T}'(s) = \mathbf{r}''(s) = \|\mathbf{r}''(s)\|\frac{\mathbf{T}'(s)}{\|\mathbf{T}'(s)\|} = \kappa(s)\mathbf{N}.$$

□

Intuitively, a circle is a curve which is banding in a uniform way. Thus we expect the curvature at any point of a circle is the same. Moreover the larger the radius of the circle the smaller the curvature is expected. We are going to show that the curvature of a circle is uniform and is equal to the reciprocal of its radius.

Example 2.3.7 (Circle)**.** *Let* $\mathbf{r}(\theta) = (r\cos\theta, r\sin\theta)$*,* $0 < \theta < 2\pi$*, be the circle of radius* $r > 0$ *centered at the origin. Then*

$$\begin{cases} \mathbf{r}'(\theta) = (-r\sin\theta, r\cos\theta) \\ \mathbf{r}''(\theta) = (-r\cos\theta, -r\sin\theta) \end{cases}$$

Thus

$$\begin{aligned} \kappa(\theta) &= \frac{|x'y'' - x''y'|}{(x'^2 + y'^2)^{\frac{3}{2}}} \\ &= \frac{|r^2\sin^2\theta + r^2\cos^2\theta|}{(r^2\cos^2\theta + r^2\sin^2\theta)^{\frac{3}{2}}} \\ &= \frac{1}{r}. \end{aligned}$$

Example 2.3.8 (Cycloid)**.** *The* **cycloid** *is the curve parametrized by*

$$\mathbf{r}(\theta) = (\theta - \sin\theta, 1 - \cos\theta), \ \textit{for } \theta \in (0, 2\pi).$$

Show that the curvature of the cycloid is

$$\kappa = \frac{1}{2^{\frac{3}{2}}\sqrt{1 - \cos\theta}}.$$

Proof. Observe that

$$\begin{cases} \mathbf{r}'(\theta) = (1 - \cos\theta, \sin\theta) \\ \mathbf{r}''(\theta) = (\sin\theta, \cos\theta) \end{cases}$$

Therefore the curvature of the cycloid is

$$\begin{aligned}\kappa(\theta) &= \frac{|x'y'' - x''y'|}{(x'^2 + y'^2)^{\frac{3}{2}}} \\ &= \frac{|(1-\cos\theta)\cos\theta - \sin\theta\sin\theta|}{((1-\cos\theta)^2 + (-\sin\theta)^2)^{\frac{3}{2}}} \\ &= \frac{1-\cos\theta}{(2-2\cos\theta)^{\frac{3}{2}}} \\ &= \frac{1}{2^{\frac{3}{2}}\sqrt{1-\cos\theta}}.\end{aligned}$$

□

Let's see some examples of curvature of space curves.

Example 2.3.9 (Helix)**.** *Let $a, b > 0$ be constants. The space curve $\mathbf{r}(\theta) = (a\cos\theta, a\sin\theta, b\theta)$, $\theta \in \mathbb{R}$, is called a* **helix***. Then*

$$\begin{cases}\mathbf{r}'(\theta) = (-a\sin\theta, a\cos\theta, b) \\ \mathbf{r}''(\theta) = (-a\cos\theta, -a\sin\theta, 0)\end{cases}$$

We have

$$\mathbf{r}' \times \mathbf{r}'' = (ab\sin\theta, -ab\cos\theta, a^2)$$

$$\begin{aligned}\kappa(\theta) &= \frac{\|\mathbf{r}' \times \mathbf{r}''\|}{\|\mathbf{r}'\|^3} \\ &= \frac{a\sqrt{a^{+}b^2}}{(a^2+b^2)^{\frac{3}{2}}} \\ &= \frac{a}{a^2+b^2}\end{aligned}$$

Observe that the curvature is constant in this case.

Proposition 2.3.10 (Curvature of graphs of functions)**.**

1. *(Rectangular coordinates): The curvature of the curve given by the graph of function $y = f(x)$ in rectangular coordinates is*

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

2. *(Polar coordinates): The curvature of the curve given by the graph of function $r = r(\theta)$ in polar coordinates is*

$$\kappa(\theta) = \frac{|r^2 + 2r'^2 - rr''|}{(r^2 + r'^2)^{\frac{3}{2}}}.$$

Proof. 1. Parametrized the graph of $y = f(x)$ by $\mathbf{r}(t) = (t, f(t))$. Then

$$\begin{cases} \mathbf{r}'(t) = (1, f') \\ \mathbf{r}''(t) = (0, f'') \end{cases}$$

Thus

$$\begin{aligned} \kappa &= \frac{|x'y'' - x''y'|}{(x'^2 + y'^2)^{\frac{3}{2}}} \\ &= \frac{|f''|}{(1^2 + f'^2)^{\frac{3}{2}}} \end{aligned}$$

2. Parametrized the graph of $r = r(\theta)$ by $\mathbf{r}(\theta) = (r\cos\theta, r\sin\theta)$. Then

$$\begin{cases} \mathbf{r}'(\theta) = (r'\cos\theta - r\sin\theta, r'\sin\theta + r\cos\theta) \\ \mathbf{r}''(\theta) = (r''\cos\theta - 2r'\sin\theta - r\cos\theta, r''\sin\theta + 2r'\cos\theta - r\sin\theta) \end{cases}$$

Thus

$$\begin{aligned} \kappa &= \frac{|x'y'' - x''y'|}{(x'^2 + y'^2)^{\frac{3}{2}}} \\ &= \frac{|2r'^2 - rr'' + r^2|}{(r^2 + r'^2)^{\frac{3}{2}}} \end{aligned}$$

□

Example 2.3.11 (Catenary)**.** *The* **catenary** *is the curve given by the graph of the function $y = \cosh x$. Show that the curvature of the catenary is*

$$\kappa = \frac{1}{\cosh^2 x}.$$

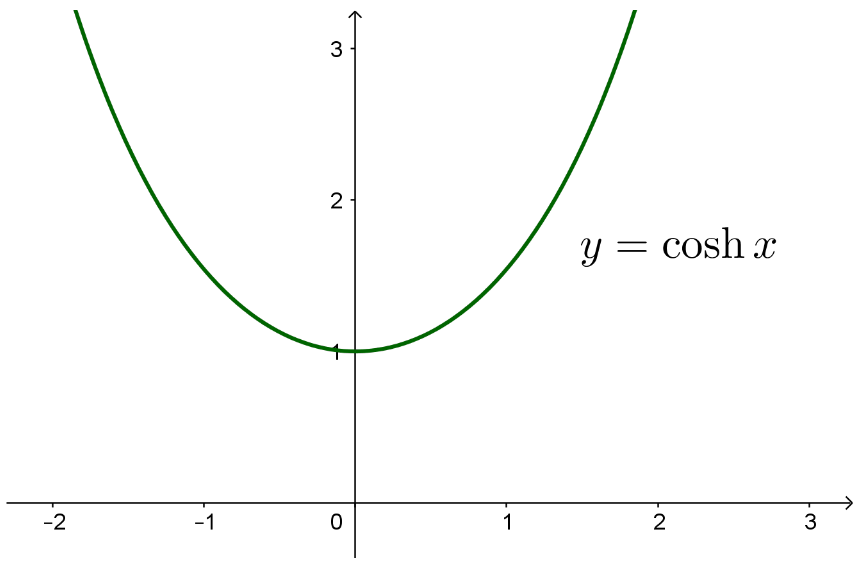

Figure 8: Catenary

Proof. Observe that

$$\begin{cases} f'(x) = \sinh x, \\ f''(x) = \cosh x. \end{cases}$$

By Proposition 2.3.10, the curvature of the catenary is

$$\begin{aligned} \kappa &= \frac{|f''|}{(1+f'^2)^{\frac{3}{2}}} \\ &= \frac{\cosh x}{(1+\sinh^2 x)^{\frac{3}{2}}} \\ &= \frac{\cosh x}{(\cosh^2 x)^{\frac{3}{2}}} \\ &= \frac{1}{\cosh^2 x} \end{aligned}$$

□

Let's summarize the above calculation in the following table.

Parametrized Curve	Arc length	Curvature
Plane curve $\mathbf{r}(t) = (x(t), y(t))$, $a < t < b$	$\int_a^b \\|\mathbf{r}'\\| dt$	$\kappa(t) = \frac{\|x'y'' - x''y'\|}{(x'^2 + y'^2)^{\frac{3}{2}}}$
Space curve $\mathbf{r}(t) = (x(t), y(t), z(t))$, $a < t < b$	$\int_a^b \\|\mathbf{r}'\\| dt$	$\kappa(t) = \frac{\\|\mathbf{r}' \times \mathbf{r}''\\|}{\\|\mathbf{r}'\\|^3}$
Arc length parametrized curve $\mathbf{r}(s)$ with $\\|\mathbf{r}(s)\\| = 1$ $a < s < b$	$b - a$	$\kappa(s) = \\|\mathbf{r}''(s)\\|$
Circle $\mathbf{r}(\theta) = (r\cos\theta, r\sin\theta)$, $0 < \theta < 2\pi$	$2\pi r$	$\kappa = \frac{1}{r}$
Cycloid $\mathbf{r}(\theta) = (\theta - \sin\theta, \cos\theta)$, $\theta \in (0, 2\pi)$	8	$\frac{1}{2^{\frac{3}{2}}\sqrt{1-\cos\theta}}$
Helix $\mathbf{r}(\theta) = (a\cos\theta, a\sin\theta, b\theta)$, $0 < \theta < 2\pi$	$2\pi\sqrt{a^2 + b^2}$	$\frac{a}{a^2 + b^2}$
Graph of function $y = f(z)$ in rectangular coordinates $\mathbf{r}(t) = (t, f(t))$, $a < t < b$	$\int_a^b \sqrt{1 + f'^2} dx$	$\frac{\|f''\|}{(1 + f'^2)^{\frac{3}{2}}}$
Graph of function $r = r(\theta)$ in polar coordination $\mathbf{r}(\theta) = (r\cos\theta, r\sin\theta)$, $\alpha < \theta < \beta$	$\int_\alpha^\beta \sqrt{r^2 + r'^2} d\theta$	$\frac{\|r^2 + 2r'^2 - rr''\|}{(r^2 + r'^2)^{\frac{3}{2}}}$

Another way to interpret the curvature of a curve is that is the change of angle of tangent vector with respect to arc length.

Proposition 2.3.12. *Let* $\mathbf{r}(s)$ *be an arc length parametrized plane curve and* $\theta(s)$ *be the angle between* $\mathbf{T}$ *and positive* x*-axis. Then*

$$\kappa(s) = \left|\frac{d\theta}{ds}\right|.$$

Proof. Suppose $\mathbf{r}(s) = (x(s), y(s))$. Then $\mathbf{T} = \mathbf{r}' = (x', y')$ and $\|\mathbf{r}'\| = \sqrt{x'^2 + y'^2} = 1$. Now

$$\theta = \tan^{-1} \frac{y'}{x'}.$$

Thus

$$\begin{aligned}\frac{d\theta}{ds} &= \frac{1}{1+\frac{y'^2}{x'^2}}\left(\frac{x'y''-y'x''}{x'^2}\right)\\ &= x'y''-y'x''\end{aligned}$$

since $x'^2 + y'^2 = 1$. Therefore

$$\kappa = |x'y'' - y'x''| = \left|\frac{d\theta}{ds}\right|.$$

□

This inspires us to give a sign for the curvature.

Definition 2.3.13 (Signed curvature)**.** *Let* $\mathbf{r}(t) = (x(t), y(t))$ *be a regular parametrized curve. The* **signed curvature***, also denoted by* κ*, of* $\mathbf{r}$ *is*

$$\kappa(t) = \frac{d\theta}{ds} = \frac{x'y'' - y'x''}{(x'^2 + y'^2)^{\frac{3}{2}}}$$

where θ *is the angle between the unit tangent vector* $\mathbf{T}$ *and the positive* x*-axis so that* $\mathbf{T} = (\cos\theta, \sin\theta)$*.*

The objects we have studied up to now are open curves. We are going to explains a theorem which concerns closed curves.

Definition 2.3.14 (Simple closed curve)**.** *A regular* **simple closed curve** *in* $\mathbb{R}^2$ *is a closed and bounded connected subset* $C \subset \mathbb{R}^2$ *such that for any point* $p \in C$*, we may find an open set* $U_p \subset \mathbb{R}^2$ *containing* p *such that* $U_p \cap C$ *is the image of a regular parametrized curve.*

There is a natural orientation which leads to a natural sign of curvature on a regular simple closed curve. The **Jordan curve theorem** asserts that a simple closed curve in $\mathbb{R}^2$ separates the plane into two regions, one bounded and another unbounded. We say that a regular parametrization of a simple closed curve is **positively oriented** if the region bounded by the curve is to the left of the tangent direction.

On a regular simple closed curve, we may find a positively oriented regular parametrization $\mathbf{r}(t)$, $a \le t \le b$, such that $\mathbf{r}$ is injective on (a, b) and $\mathbf{r}(a) = \mathbf{r}(b)$. Define a function $\theta(t)$, $a \le t \le b$, which is continuous and $\theta(t)$ is the angle between the unit tangent vector $\mathbf{T}(t)$ and the positive x-axis so that $\mathbf{T} = (\cos\theta, \sin\theta)$. The choice of $\theta(t)$ is not unique but any two choices are different by a multiple of 2π. Then since $\mathbf{T}(a) = \mathbf{T}(b)$, the value $\theta(b) - \theta(a)$ must be a multiple of 2π. For regular simple closed curve, we must have $\theta(b) - \theta(a) = 2\pi$.

Theorem 2.3.15. *Let $\mathbf{r}(t)$, $a \le t \le b$, be a positively oriented regular parametrization of a regular simple closed curve C such that $\mathbf{r}(t)$ is injective on (a, b) and $\mathbf{r}(a) = \mathbf{r}(b)$. Let $\theta(t)$ be a continuous function such that $\theta(t)$ is the angle between the unit tangent vector $\mathbf{T}(t)$ and the positive x-axis so that $\mathbf{T} = (\cos\theta, \sin\theta)$. Then $\theta(b) - \theta(a) = 2\pi$.*

Sketch of proof. We are going to deform the simple closed curve C. When we deform the curve, the quantity $\theta(b) - \theta(a)$ must keep constant. This is because $\theta(b) - \theta(a)$ would change continuously when the curve is being deformed and the quantity only takes integer values which forces it to be constant. Now a regular simple closed curve can always be deformed regularly into the unit circle. (This is where the assumption that the closed curve C is simple is being used.) By considering the positive oriented parametrization $\mathbf{r}(t) = (\cos t, \sin t)$, $0 \le t \le 2\pi$, of the unit circle, the unit tangent vector is $\mathbf{T}(t) = (-\sin t, \cos t)$ and we see that an angle function can be chosen to be $\theta(t) = t + \frac{\pi}{2}$. Now we have $\theta(2\pi) - \theta(0) = 2\pi$ and the proof of the theorem is complete. □

Signed curvature of a simple closed curve can be considered as the continuous version of exterior angles of a polygon. The following theorem is the continuous version of the theorem for sum of exterior angles of polygon.

Theorem 2.3.16. *Let C be a simple closed curve and κ be the signed curvature defined by positively oriented parametrization. Then*

$$\int_C \kappa ds = 2\pi.$$

Sketch of proof. Let $\mathbf{r}(t)$, $a \le t \le b$ be a positively oriented parametrization of C so that $\mathbf{r}(t)$ is injective on (a, b) and $\mathbf{r}(a) = \mathbf{r}(b)$. Let $\theta(t)$ be the angle between $\mathbf{T}(t)$ and positive x-axis which is a continuous function such that $\mathbf{T} = (\cos\theta, \sin\theta)$. Now $\kappa = \dfrac{d\theta}{ds}$ and using Theorem 2.3.15, we have

$$\int_C \kappa ds = \int_C \frac{d\theta}{ds} ds = \int_C d\theta = \theta(b) - \theta(a) = 2\pi$$

□

The curvature of a curve can be interpreted in at least two more ways. First, a regular parametrized curve $\mathbf{r}(t)$ can be considered as the displacement of a moving particle at time t. Then $\mathbf{v} = \mathbf{r}'$ is the velocity of the particle. We may write the acceleration $\mathbf{a} = \mathbf{r}''$ as a linear combination of orthogonal vectors $\mathbf{T}$ and $\mathbf{N}$. It is known that the projection of $\mathbf{a}$ along $\mathbf{T}$ is $\dfrac{d\|\mathbf{v}\|}{dt}$ and the normal component depends on the velocity of the particle and the curvature of the curve.

Proposition 2.3.17. *Let $\mathbf{r}(t)$ be a regular parametrized curve. Then*

$$\mathbf{a} = \mathbf{r}'' = \frac{dv}{dt}\mathbf{T} + \kappa v^2 \mathbf{N}$$

where $v = \|\mathbf{v}\| = \|\mathbf{r}'\|$.

Proof. First, we have

$$\mathbf{r}'(t) = v(t)\mathbf{T}(t).$$

Let s be an arc length parameter, that means $s(t)$ is a function such that $\dfrac{ds}{dt} = \|\mathbf{r}'(t)\|$. Then $\dfrac{d}{ds}\mathbf{T} = \kappa\mathbf{N}$ by Theorem 2.3.6 and we have

$$\begin{aligned}\mathbf{r}'' &= \frac{dv}{dt}\mathbf{T} + v\frac{d}{dt}\mathbf{T}\\ &= \frac{dv}{dt}\mathbf{T} + v\frac{ds}{dt}\frac{d}{ds}\mathbf{T}\\ &= \frac{dv}{dt}\mathbf{T} + \kappa v^2\mathbf{N}.\end{aligned}$$

□

In the view of the above proposition, we may also define curvature to be the normal component of acceleration divided by the square of speed, that is,

$$\kappa(t) = \frac{\langle \mathbf{r}''(t), \mathbf{N}\rangle}{\|\mathbf{r}'(t)\|^2}.$$

There is one more way to interpret the curvature of a curve. When we consider $\mathbf{r}(t)$ as the displacement of a moving particle, we try to find a circle which is closest to the trajectory of the particle at a certain point on the curve. Then the curvature of the curve at that point can be interpreted as the reciprocal of the radius of that circle.

Let's summarize the fact about curvature of a curve in the following proposition.

Proposition 2.3.18. *Let* $\mathbf{r}(t)$ *be a regular parametrized curve. Let* $s(t)$ *be an arc length parameter, that is,* $\frac{ds}{dt} = \|\mathbf{r}'(t)\|$ *or equivalently* $\left\|\frac{d\mathbf{r}}{ds}\right\| = 1$. *Let* $\mathbf{T}$ *and* $\mathbf{N}$ *be the unit tangent and normal vectors, which can be considered as vector valued functions of* t *or* s*, respectively. The curvature* κ *of the curve is characterized by any of the following conditions.*

1. $$\kappa(t) = \frac{\|\mathbf{T}'(t)\|}{\|\mathbf{r}'(t)\|}$$

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

3. *If* $\mathbf{r} = (x, y)$ *is a plane curve, we have*
$$\kappa = \frac{|x'y'' - x''y'|}{(x'^2 + y'^2)^{\frac{3}{2}}}.$$

4. *If* $\mathbf{r} = (x, y, z)$ *is a space curve, we have*
$$\kappa = \frac{\|\mathbf{r}' \times \mathbf{r}''\|}{\|\mathbf{r}'\|^3}.$$

5. $$\kappa = \left\|\frac{d^2\mathbf{r}}{ds^2}\right\|$$

6. *If* $\mathbf{r} = (x, y)$ *is a plane curve and* θ *is the angle between* $\mathbf{T}$ *and the positive* x*-axis, that is,* $\mathbf{T} = (\cos\theta, \sin\theta)$*, then we have*
$$\kappa = \frac{d\theta}{ds}.$$

7. $$\mathbf{r}'' = \frac{dv}{dt}\mathbf{T} + \kappa v^2\mathbf{N}, \text{ where } v = \|\mathbf{r}'(t)\|.$$

2.4 Frenet frame

In this section, we study space curve. We have define unit normal to a curve by $\mathbf{N} = \frac{\mathbf{T}'}{\|\mathbf{T}'\|}$. For space curve, there is one more direction which is orthogonal to the unit tangent vector which is called binormal.

Definition 2.4.1 (Binormal)**.** *Let* $\mathbf{r}(t)$ *be a space curve with curvature* $\kappa(t) > 0$ *for any* t*. We define the unit* **binormal** *to the curve by*

$$\mathbf{B}(t) = \mathbf{T}(t) \times \mathbf{N}(t).$$

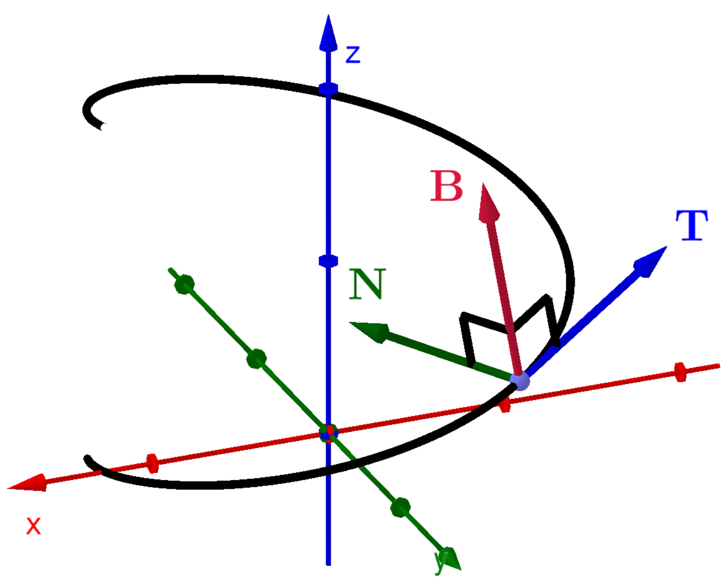

Figure 9: Binormal

Note that **T**, **N**, **B** form a orthonormal basis for $\mathbb{R}^3$. This basis depends on t and we may consider it moving along the curve. We call it the **Frenet frame** and this is the simplest example of moving frame. We would like to study how this frame moves along the curve.

Definition 2.4.2 (Torsion)**.** *Let* $\mathbf{r}(t)$ *be a space curve with curvature* $\kappa(t) > 0$ *for any* t*. The* **torsion** *of the curve at* $\mathbf{r}(t)$ *is defined by*

$$\tau = \left\langle \frac{d\mathbf{N}}{ds}, \mathbf{B} \right\rangle$$

where s *is a arc length parameter, which means* $\frac{ds}{dt} = \|\mathbf{r}'(t)\|$*. Equivalently, we have*

$$\tau(t) = \left\langle \frac{\mathbf{N}'(t)}{\|\mathbf{r}'(t)\|}, \mathbf{B}(t) \right\rangle$$

Here we give a formula for finding τ.

Proposition 2.4.3. *Let* $\mathbf{r}(t)$ *be a space curve with curvature* $\kappa(t) > 0$ *for any* t*. Then*

$$\tau = \frac{\langle \mathbf{r}' \times \mathbf{r}'', \mathbf{r}''' \rangle}{\|\mathbf{r}' \times \mathbf{r}''\|^2}.$$

Proof. Note that

$$\begin{aligned}
\mathbf{r}' &= \|\mathbf{r}'\|\mathbf{T} \\
\mathbf{r}'' &= \frac{d}{dt}(\|\mathbf{r}'\|\mathbf{T}) \\
&= \frac{d\|\mathbf{r}'\|}{dt}\mathbf{T} + \|\mathbf{r}'\|\mathbf{T}' \\
&= \frac{\langle \mathbf{r}', \mathbf{r}'' \rangle}{\|\mathbf{r}'\|}\mathbf{T} + \|\mathbf{r}'\|(\kappa\|\mathbf{r}'\|\mathbf{N}) \quad \text{(Proposition 2.3.2)} \\
&= \frac{\langle \mathbf{r}', \mathbf{r}'' \rangle}{\|\mathbf{r}'\|}\mathbf{T} + \kappa\|\mathbf{r}'\|^2\mathbf{N}
\end{aligned}$$

So

$$\begin{aligned}
\mathbf{r}' \times \mathbf{r}'' &= \|\mathbf{r}'\|\mathbf{T} \times \left(\frac{\langle \mathbf{r}', \mathbf{r}'' \rangle}{\|\mathbf{r}'\|}\mathbf{T} + \kappa\|\mathbf{r}'\|^2\mathbf{N} \right) \\
&= \kappa\|\mathbf{r}'\|^3\mathbf{T} \times \mathbf{N} \\
&= \kappa\|\mathbf{r}'\|^3\mathbf{B}
\end{aligned}$$

Thus

$$\begin{aligned} \|\mathbf{r}' \times \mathbf{r}''\| &= \kappa \|\mathbf{r}'\|^3 \\ \mathbf{r}''' &= \frac{d}{dt}\left(\frac{\langle \mathbf{r}', \mathbf{r}''\rangle}{\|\mathbf{r}'\|}\right)\mathbf{T} + \frac{\langle \mathbf{r}', \mathbf{r}''\rangle}{\|\mathbf{r}'\|}\mathbf{T}' + \frac{d}{dt}\left(\kappa\|\mathbf{r}'\|^2\right)\mathbf{N} + \kappa\|\mathbf{r}'\|^2\mathbf{N}' \\ &= \frac{d}{dt}\left(\frac{\langle \mathbf{r}', \mathbf{r}''\rangle}{\|\mathbf{r}'\|}\right)\mathbf{T} + \frac{\langle \mathbf{r}', \mathbf{r}''\rangle}{\|\mathbf{r}'\|}(\kappa\mathbf{N}) + \frac{d}{dt}\left(\kappa\|\mathbf{r}'\|^2\right)\mathbf{N} + \kappa\|\mathbf{r}'\|^2\mathbf{N}' \end{aligned}$$

Note that $\langle \mathbf{T}, \mathbf{B}\rangle = \langle \mathbf{N}, \mathbf{B}\rangle = 0$. Thus

$$\begin{aligned} \langle \mathbf{r}' \times \mathbf{r}'', \mathbf{r}'''\rangle &= \langle \kappa\|\mathbf{r}'\|^3\mathbf{B}, \kappa\|\mathbf{r}'\|^2\mathbf{N}'\rangle \\ &= \kappa\|^2\mathbf{r}'\|^5\langle \mathbf{N}', \mathbf{B}\rangle \\ &= (\kappa\|\mathbf{r}'\|^3)^2\langle \frac{\mathbf{N}'}{\|\mathbf{r}'\|}, \mathbf{B}\rangle \\ &= \|\mathbf{r}' \times \mathbf{r}''\|^2\tau \end{aligned}$$

Therefore

$$\tau = \frac{\langle \mathbf{r}' \times \mathbf{r}'', \mathbf{r}'''\rangle}{\|\mathbf{r}' \times \mathbf{r}''\|^2}.$$

□

Theorem 2.4.4 (Frenet formula)**.** *Let* $\mathbf{r}(s)$ *be a regular space curve parametrized by arc length with curvature* $\kappa(s) > 0$ *for any* s*. Then*

$$\begin{cases} \mathbf{T}'(s) &= \quad\quad\;\; \kappa\mathbf{N} \\ \mathbf{N}'(s) &= -\kappa\mathbf{T} \quad\quad\quad + \tau\mathbf{B} \\ \mathbf{B}'(s) &= \quad\quad -\tau\mathbf{N} \end{cases}$$

We may write the formula in matrix form

$$\frac{d}{ds}\begin{pmatrix} \mathbf{T} \\ \mathbf{N} \\ \mathbf{B} \end{pmatrix} = \begin{pmatrix} 0 & \kappa & 0 \\ -\kappa & 0 & \tau \\ 0 & -\tau & 0 \end{pmatrix}\begin{pmatrix} \mathbf{T} \\ \mathbf{N} \\ \mathbf{B} \end{pmatrix}$$

Proof. First by definition of curvature (Definition 2.3.3), we have

$$\mathbf{T}'(s) = \kappa\mathbf{N}.$$

We are going to use the following fact. If $\mathbf{u}(t)$ and $\mathbf{v}(t)$ are two vector valued functions with $\langle \mathbf{u}, \mathbf{v}\rangle$ being constant, then $\langle \mathbf{u}', \mathbf{v}\rangle = -\langle \mathbf{u}, \mathbf{v}'\rangle$ (See

Lemma 1.3.35). In particular if $\|\mathbf{v}\|$ is constant, then $\langle \mathbf{v}', \mathbf{v} \rangle = 0$. Now since $\langle \mathbf{N}, \mathbf{T} \rangle = 0$ for any s, we have

$$\begin{aligned} \langle \mathbf{N}'(s), \mathbf{T} \rangle &= -\langle \mathbf{N}, \mathbf{T}'(s) \rangle \\ &= -\langle \mathbf{N}, \kappa \mathbf{N} \rangle \\ &= -\kappa \end{aligned}$$

Moreover since $\|\mathbf{N}\| = 1$ is a constant, we have

$$\langle \mathbf{N}'(s), \mathbf{N} \rangle = 0.$$

Observe that $\mathbf{T}$, $\mathbf{N}$, $\mathbf{B}$ constitute an orthonormal basis for $\mathbb{R}^3$. We get

$$\begin{aligned} \mathbf{N}'(s) &= \langle \mathbf{N}'(s), \mathbf{T} \rangle \mathbf{T} + \langle \mathbf{N}'(s), \mathbf{N} \rangle \mathbf{N} + \langle \mathbf{N}'(s), \mathbf{B} \rangle \mathbf{B} \\ &= -\kappa \mathbf{T} + \tau \mathbf{B} \end{aligned}$$

Applying the same argument to $\mathbf{B}'(s)$, we have

$$\begin{cases} \langle \mathbf{B}'(s), \mathbf{T} \rangle = -\langle \mathbf{B}, \mathbf{T}'(s) \rangle = 0 \\ \langle \mathbf{B}'(s), \mathbf{N} \rangle = -\langle \mathbf{B}, \mathbf{N}'(s) \rangle = -\tau \end{cases}$$

since $\langle \mathbf{B}, \mathbf{T} \rangle = \langle \mathbf{B}, \mathbf{N} \rangle = 0$ and

$$\langle \mathbf{B}'(s), \mathbf{B} \rangle = 0$$

since $\|\mathbf{B}\| = 1$ is constant. Therefore

$$\begin{aligned} \mathbf{B}'(s) &= \langle \mathbf{B}'(s), \mathbf{T} \rangle \mathbf{T} + \langle \mathbf{B}'(s), \mathbf{N} \rangle \mathbf{N} + \langle \mathbf{B}'(s), \mathbf{B} \rangle \mathbf{B} \\ &= -\tau \mathbf{N} \end{aligned}$$

□

In the last section, we define plane curve as a curve in $\mathbb{R}^2$. However, we would also call a space curve a plane curve if it lies on a plane in $\mathbb{R}^3$.

Definition 2.4.5 (Plane curve)**.** *We say that a space curve* $\mathbf{r}$ *is a* **plane curve** *if there exists a unit vector* $\mathbf{n}$ *such that*

$$\langle \mathbf{r}, \mathbf{n} \rangle = a$$

is a constant.

The vector $\mathbf{n}$ in the above definition is the unit normal vector of the plane containing the curve. The curvature κ of a curve measures how far a curve is away from straight and the torsion τ measure how far it is away from a plane curve.

Proposition 2.4.6. *Let* $\mathbf{r}(t)$ *be a regular parametrized space curve with curvature* $\kappa(t) > 0$ *for any* t*. Then* $\mathbf{r}$ *is a plane curve if and only if its torsion* $\tau(t) = 0$ *for any* t*.*

Proof. By Proposition 2.2.7, we may consider the arc length parametrization $\mathbf{r}(s)$ of the curve. Suppose $\mathbf{r}(s)$ is a plane curve. Then there exists constant unit vector $\mathbf{n}$ such that

$$\langle \mathbf{r}, \mathbf{n} \rangle = a$$

is a constant. Observe that

$$\begin{cases} \langle \mathbf{r}'(s), \mathbf{n} \rangle = \dfrac{d}{ds}\langle \mathbf{r}, \mathbf{n} \rangle = 0 \\ \langle \mathbf{r}''(s), \mathbf{n} \rangle = \dfrac{d}{ds}\langle \mathbf{r}', \mathbf{n} \rangle = 0 \\ \langle \mathbf{r}'''(s), \mathbf{n} \rangle = \dfrac{d}{ds}\langle \mathbf{r}'', \mathbf{n} \rangle = 0 \end{cases}$$

which implies that $\langle \mathbf{r}' \times \mathbf{r}'', \mathbf{r}''' \rangle = 0$. Therefore

$$\tau = \frac{\langle \mathbf{r}' \times \mathbf{r}'', \mathbf{r}''' \rangle}{\|\mathbf{r}' \times \mathbf{r}''\|^2} = 0$$

Conversely Suppose $\tau(s) = 0$ for any s. Then by Frenet formula (Theorem 2.4.4), we have

$$\mathbf{B}'(s) = -\tau \mathbf{N} = \mathbf{0}$$

for any s. Thus the binormal $\mathbf{B}$ is a constant vector and

$$\begin{aligned} \frac{d}{ds}\langle \mathbf{r}, \mathbf{B} \rangle &= \langle \mathbf{r}', \mathbf{B} \rangle + \langle \mathbf{r}, \mathbf{B}' \rangle \\ &= \langle \mathbf{T}, \mathbf{B} \rangle - \tau \langle \mathbf{r}, \mathbf{N} \rangle \\ &= 0 \end{aligned}$$

for any s. Therefore $\langle \mathbf{r}, \mathbf{B} \rangle$ is constant which means $\mathbf{r}$ is a plane curve lying on a plane with normal vector $\mathbf{B}$. □

We end this section by stating the fundamental theorem of space curves without proof.

Theorem 2.4.7 (Fundamental theorem of space curves)**.** *Let $\kappa(s), \tau(s) > 0$ be two positive functions. Then there exists unique, up to rigid transformation, space curve $\mathbf{r}(s)$ parametrized by arc length with curvature $\kappa(s)$ and torsion $\tau(s)$.*

Exercise 2

1. Write down a regular parametrization of the following curves in $\mathbb{R}^2$
   (a) The line segment joining $(1, -2)$ and $(-3, 2)$.
   (b) The circle of radius 5 centered at $(3, -1)$.
   (c) The ellipse with equation $\frac{x^2}{4} + \frac{y^2}{9} = 1$.
2. Find the arc-length of the following plane curves.
   (a) $y = \frac{x^4+3}{6x}$ from $x = 1$ to $x = 2$.
   (b) $y^2 = x^3$ from $(0, 0)$ to $(4, 8)$.
   (c) The astroid defined by $x^{\frac{2}{3}} + y^{\frac{2}{3}} = 1$.
   (d) The deltoid parametrized by $\mathbf{r}(\theta) = (2\cos\theta + \cos 2\theta, 2\sin\theta - \sin 2\theta)$, $0 \leq \theta \leq 2\pi$.
3. It is given that the following curves are parametrized by arc-length. Find the value of p where $p > 0$.
   (a) $\mathbf{r}(\theta) = (4\sin p\theta, -4\cos p\theta, 3p\theta)$
   (b) $\mathbf{r}(\theta) = (p\cos\theta, 2 + \sin\theta, 1 - \frac{\sqrt{3}}{2}\cos\theta)$, for $0 < \theta < 2\pi$.
   (c) $\mathbf{r}(t) = (\frac{1}{3}(1+t)^{\frac{3}{2}}, \frac{1}{3}(1-t)^{\frac{3}{2}}, pt)$ for $0 < t < 1$.
4. The logarithmic spiral is a curve defined by $r = e^\theta$ in polar coordinates.
   (a) Find the arc-length of the logarithmic spiral from $\theta = 0$ to $\theta = 2\pi$.
   (b) Find the curvature of the logarithmic spiral.

5. The tractrix is the curve parametrized by

$$\mathbf{r}(\theta) = \left(\sin\theta, \cos\theta + \ln\left(\tan\frac{\theta}{2}\right)\right), \text{ for } \theta \in (0, \frac{\pi}{2}).$$

   (a) Show that if the tangent to the tractrix at a point p meet the y-axis at q, then the distance between p and q is 1.

   (b) Show that $\|\mathbf{r}'(\theta)\| = \cot\theta$.

   (c) Show that the arc length of $\mathbf{r}(\theta)$ from $\theta = \alpha$ to $\theta = \frac{\pi}{2}$ is $-\ln\sin\alpha$.

   (d) Show that the curvature of the tractrix is given by $\kappa(\theta) = \tan\theta$.

6. Given a circle of radius R and it is rolling along a straight line (which may be assumed to be the x-axis). Let P be a point on the circumference of the circle of radius R. The curve travelled by the point P, i.e., the locus of P, is called a cycloid. Let θ be the angle between the vertical line (y-axis) and the radius from the center of the circle to P. The cycloid is parametrized by

$$\mathbf{r}(\theta) = (R(\theta - \sin\theta), R(1 - \cos\theta)), \text{ for } 0 \le \theta \le 2\pi$$

   (a) Show that $\mathbf{r}'$ is orthogonal to $\mathbf{r} - (R\theta, 0)$ and $\|\mathbf{r}'\| = \|\mathbf{r} - (R\theta, 0)\|$ for any $0 < \theta < 2\pi$.

   (b) Find the arc-length of $\mathbf{r}$ from $\theta = 0$ to $\theta = 2\pi$.

   (c) Find the curvature of $\mathbf{r}$ in terms of θ.

7. Consider the curve C given by the graph of the function $y = \ln\csc x$, $0 < x < \pi$, in rectangular coordinates.

   (a) Show that $\mathbf{r}(s) = (2\tan^{-1} e^s, \ln\cosh s)$, $s \in \mathbb{R}$ is an arc length parametrization of C.

   (b) Show that the curvature of the curve is

$$\kappa(s) = \frac{1}{\cosh s}$$

8. Prove that the curvature of the curve defined by $r = r(\theta)$ in polar coordinates is given by

$$\kappa(\theta) = \frac{|2r'^2 - rr'' + r^2|}{(r^2 + r'^2)^{\frac{3}{2}}}$$

9. Let $\mathbf{r}(t) = (x(t), y(t))$ be a regular parametrized curve. Suppose there is a differentiable function $\theta(t)$ such that $\tan\theta(t) = \frac{y'(t)}{x'(t)}$ for any t. Prove that
$$\frac{d\theta}{dt} = \frac{x'y'' - y'x''}{x'^2 + y'^2}$$

10. Let $\mathbf{r}(t)$ be a regular parametrized curve and $\kappa(t)$ be its curvature. Prove that if $\kappa(t) = 0$ for any t, then $\mathbf{r}(t)$ is a straight line.

11. Let $\mathbf{r}(s)$ be a regular arc length parametrized plane curve with curvature κ which is a constant.
    (a) Prove that $\dfrac{d}{ds}\left(\mathbf{r}(s) + \dfrac{1}{\kappa}\mathbf{N}(s)\right) = \mathbf{0}$ where $\mathbf{N}$ is the unit normal vector.
    (b) Hence show that $\mathbf{r}(s)$ lies on a circle.

12. Let $\mathbf{r}(s)$, $-1 < s < 1$, be an arc length parametrization of a simple closed curve with curvature $\kappa(s) = \dfrac{a}{1+s^2}$ where a is a constant. Find the value of a.

13. The **tractrix** is the curve parametrized by
$$\mathbf{r}(t) = (\text{sech} t, t - \tanh t),\ t > 0.$$
    (a) Suppose the tangent at $\mathbf{r}(t)$ intercept the y-axis at $\mathbf{r}_0(t)$. Prove that the distant between $\mathbf{r}_0(t)$ and $\mathbf{r}(t)$ is constantly equal to 1.
    (b) Find an arc length parametrization of the tractrix so that $s = 0$ corresponds to the point $(1, 0)$.
    (c) Show that the curvature of tractrix is
$$\kappa = \text{csch} t = \frac{1}{\sqrt{e^{2s} - 1}}.$$

14. Let $\mathbf{r}(t)$ be a regular parametrized plane curve with $\kappa(t) > 0$ for any t. Let $\lambda > 0$ be a constant. The parallel curve $\mathbf{r}_\lambda$ of $\mathbf{r}$ is defined by
$$\mathbf{r}_\lambda(t) = \mathbf{r}(t) - \lambda\mathbf{N}(t)$$
where $\mathbf{N}(t)$ is the unit normal vector at $\mathbf{N}$. Show that the curvature of $\mathbf{r}_\lambda(t)$ is $\dfrac{\kappa}{1+\lambda\kappa}$.

15. Find the curvature $\kappa(t)$ and torsion $\tau(t)$ of the following space curves $\mathbf{r}(t)$.

    (a) $\mathbf{r}(t) = (4\cos t, 4\sin t, 3t)$

    (b) $\mathbf{r}(t) = (\cosh t, \sinh t, t)$, $t \in \mathbb{R}$

    (c) $\mathbf{r}(t) = (\cos^3 t, \sin^3 t, \cos 2t)$, $0 < t < \dfrac{\pi}{2}$

16. Let $\mathbf{r}(t)$ be a regular parametrized space curve with $\kappa(t) > 0$ for any t. Suppose $\tau(t) = 0$ for any t, where $\tau(t)$ is the torsion at $\mathbf{r}(t)$. Prove that $\mathbf{r}(t)$ is contained in a plane.

17. Let $\mathbf{r}(s)$ be a regular space curve with arc length parametrization, $\mathbf{N}(s)$ and $\mathbf{B}(s)$ be the unit normal and unit binormal to the curve respectively. Let $\kappa(s)$ and $\tau(s)$ be the curvature and torsion of the curve. Suppose $\mathbf{r}(s)$ lies on the unit sphere for any s.

    (a) Prove that $\langle \mathbf{r}, \mathbf{N} \rangle = -\dfrac{1}{\kappa}$ for any s.

    (b) Prove that $\mathbf{r} = -\dfrac{1}{\kappa}\mathbf{N} + \dfrac{\kappa'}{\kappa^2\tau}\mathbf{B}$.

18. Let $\mathbf{r}(s)$ be a regular space curve with arc length parametrization, $\mathbf{T}(s)$ and $\mathbf{N}(s)$ be the unit tangent vector and unit normal vector respectively. Suppose $\kappa(s) > 0$ for any s and there exists a constant c and a constant unit vector $\mathbf{u}$ such that $\langle \mathbf{T}(s), \mathbf{u} \rangle = c$ for all s.

    (a) Show that $\mathbf{N}(s)$ and $\mathbf{u}$ are orthogonal for all s.

    (b) Using (a), show that there exists a constant θ such that

    $$\mathbf{u} = \cos\theta \mathbf{T}(s) + \sin\theta \mathbf{B}(s)$$

    for all s.

    (c) Using (b) and the Frenet formulas, or otherwise, prove that $\dfrac{\tau(s)}{\kappa(s)} = \cot\theta$.

3 Surfaces

3.1 Regular parametrized surfaces

In the last chapter, we study curves by parametrization which is a function from an open interval (a, b) to $\mathbb{R}^2$ or $\mathbb{R}^3$. We also require a parametrization $\mathbf{r}(t)$ to be regular, which means $\mathbf{r}'(t) \neq \mathbf{0}$, to ensure that the curve is sufficiently smooth. Similarly we consider regular parametrized surface.

Definition 3.1.1 (Regular parametrized surface)**.** *A* **regular parametrized surface** *is a differentiable function* $\mathbf{x} : D \to \mathbb{R}^3$*, where* $D \subset \mathbb{R}^2$ *is an open connected subset, such that* $\mathbf{x}_u \times \mathbf{x}_v \neq \mathbf{0}$*, for any* $(u, v) \in D \subset \mathbb{R}^2$*. The image* $S = \mathbf{x}(D) \subset \mathbb{R}^3$ *is called a* **regular surface***.*

For $\mathbf{x}(u, v)$, we denote $\mathbf{x}_u = \frac{\partial \mathbf{x}}{\partial u}$ and $\mathbf{x}_v = \frac{\partial \mathbf{x}}{\partial v}$ to be the partial derivatives of $\mathbf{x}$. Note that the condition that both $\mathbf{x}_u \neq \mathbf{0}$ and $\mathbf{x} \neq \mathbf{0}$ is not sufficient for $\mathbf{x}(u, v)$ to be regular. We require that $\mathbf{x}_u \times \mathbf{x}_v \neq \mathbf{0}$ which geometrically means that the vectors $\mathbf{x}_u$ and $\mathbf{x}_v$ span a nondegenerate parallelogram in $\mathbb{R}^3$.

A curve is an one dimensional object and there is only one tangent direction. A regular surface $\mathbf{x}(u, v)$ has infinitely many tangent directions which includes $\mathbf{x}_u$, $\mathbf{x}_v$ and all their linear combinations.

Definition 3.1.2 (Tangent space)**.** *Let S be a regular surface with parametrization* $\mathbf{x}(u, v)$*. The* **tangent space** *of S at* $p = \mathbf{x}(u, v)$ *is*

$$T_pS = \{\alpha \mathbf{x}_u + \beta \mathbf{x}_v : \alpha, \beta \in \mathbb{R}\} \subset \mathbb{R}^3.$$

We call it a 'space' because the tangent space is a vector space. In other words, the tangent T_pS satisfies the following condition.

For any $\mathbf{u}, \mathbf{v} \in T_pS$ and $\alpha, \beta \in \mathbb{R}$, we have $\alpha\mathbf{u} + \beta\mathbf{v} \in T_pS$.

Example 3.1.3.

1. *Sphere: Let $r > 0$ be a positive real number. The function*

   $$\mathbf{x}(\phi, \theta) = (r \sin\phi \cos\theta, r \sin\phi \sin\theta, r \cos\phi), \text{ for } (\phi, \theta) \in (0, \pi) \times (0, 2\pi)$$

   defines a **sphere** *of radius r centered at the origin.*

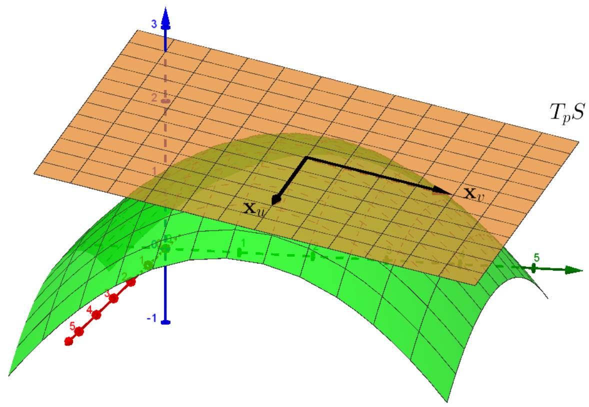

Figure 10: Tangent space

2. *Torus: Let* $R > r > 0$ *be positive real numbers. The function*

$$\mathbf{x}(\phi,\theta) = ((R+r\sin\phi)\cos\theta, (R+r\sin\phi)\sin\theta, r\cos\phi), \text{ for } \phi,\theta \in (0,2\pi)$$

defines a regular surface which is called **torus**.

3. *Helicoid: Let* $a > 0$ *be positive real numbers. The function*

$$\mathbf{x}(u,\theta) = (u\cos\theta, u\sin\theta, a\theta), \text{ for } u,\theta \in \mathbb{R}$$

defines a regular surface which is called **helicoid**.

3.2 First fundamental form and surface area

Analogue to the arc length of a curve is the surface area of a surface. To define surface area, we introduce the first fundamental form.

Definition 3.2.1 (First fundamental form)**.** *Let* $\mathbf{x}(u,v)$ *be a regular parametrized surface. The* **first fundamental form** *of* $\mathbf{x}$ *is the* 2×2 *matrix valued function*

$$I = \begin{pmatrix} E & F \\ F & G \end{pmatrix} = \begin{pmatrix} \langle \mathbf{x}_u, \mathbf{x}_u\rangle & \langle \mathbf{x}_u, \mathbf{x}_v\rangle \\ \langle \mathbf{x}_v, \mathbf{x}_u\rangle & \langle \mathbf{x}_v, \mathbf{x}_v\rangle \end{pmatrix}.$$

Here E, F, G are ordinary real valued functions and the first fundamental form I is a matrix valued function of u, v. Note that I is symmetric because

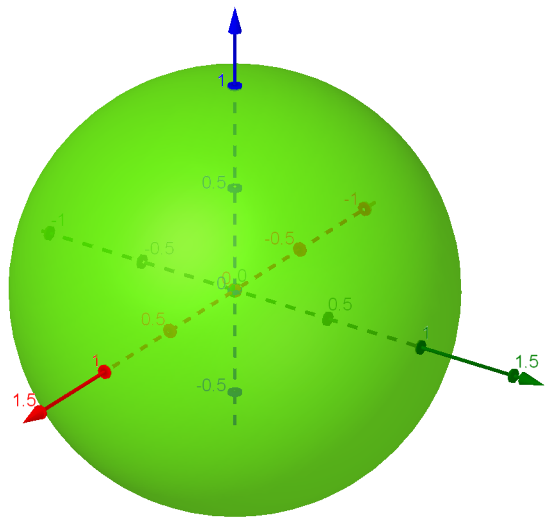

Figure 11: Sphere

$\langle \mathbf{x}_u, \mathbf{x}_v \rangle = \langle \mathbf{x}_v, \mathbf{x}_u \rangle$ by property of scalar product. The determinant of I has another interpretation.

Theorem 3.2.2. *Let $\mathbf{x}(u,v)$ be a regular parametrized surface and I be its first fundamental form. Then*

$$\det(I) = \|\mathbf{x}_u \times \mathbf{x}_v\|^2.$$

In particular, we have $\det(I) > 0$ for any u, v.

Proof. The first statement follows by Proposition 1.3.17. Since $\mathbf{x}(u,v)$ is regular, we have $\mathbf{x}_u \times \mathbf{x}_v \neq \mathbf{0}$ which implies $\det(I) = \|\mathbf{x}_u \times \mathbf{x}_v\|^2 > 0$ for any u, v. □

Now we give the definition of surface area of regular surface.

Definition 3.2.3 (Surface area)**.** *Let S be a regular surface with parametrization $\mathbf{x}(u,v)$, $(u,v) \in D$. The* **surface area** *of S is defined by*

$$A = \iint_D \sqrt{\det(I)}\, dudv.$$

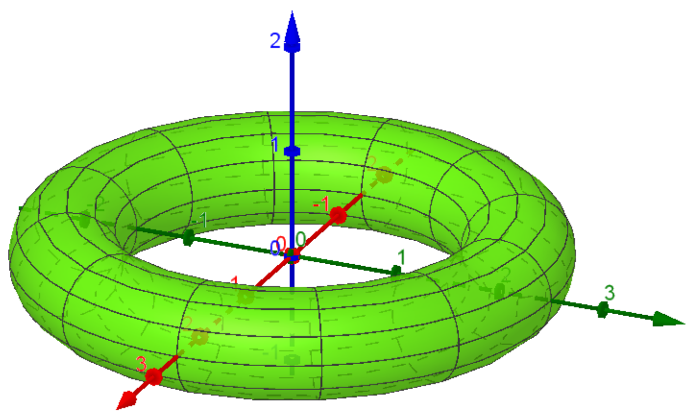

Figure 12: Torus

Note that by Theorem 3.2.2, the surface surface can also be expressed as

$$A = \iint_D \|\mathbf{x}_u \times \mathbf{x}_v\| du dv.$$

To see that this gives the surface area of the surface, one may cut the surface into small pieces which can be approximated by small parallelograms spanned by $\mathbf{x}(u+\Delta u, v) - \mathbf{x}(u,v) \approx \Delta u \mathbf{x}_u(u,v)$ and $\mathbf{x}(u, v+\Delta v) - \mathbf{x}(u,v) \approx \Delta v \mathbf{x}_v$. Then the area ΔA of each small piece can be approximated by the parallelogram and we have

$$\Delta A \approx \|\Delta u \mathbf{x}_u \times \Delta v \mathbf{x}_v\| = \|\mathbf{x}_u \times \mathbf{x}_v\| \Delta u \Delta v.$$

Therefore the surface area of the surface is

$$A = \lim \sum \|\mathbf{x}_u \times \mathbf{x}_v\| \Delta u \Delta v = \iint_D \|\mathbf{x}_u \times \mathbf{x}_v\| du dv.$$

Someone may ask why we write $\sqrt{\det(I)}$ instead of $\|\mathbf{x}_u \times \mathbf{x}_v\|$ in the definition of surface area. One reason is that cross product is defined only on $\mathbb{R}^3$ but scalar product can be calculated in any dimension which allows us to use Definition 3.2.3 to define surface area of regular surface in $\mathbb{R}^n$ for any $n \geq 3$.

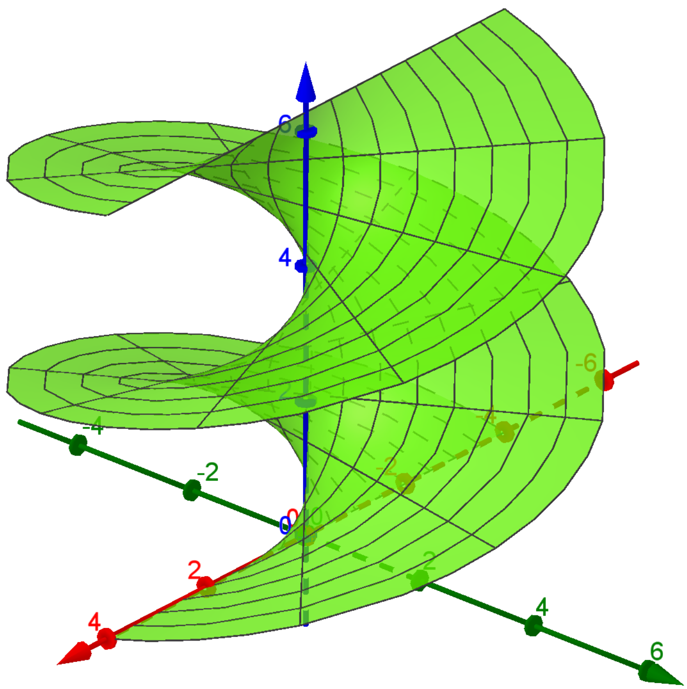

Figure 13: Helicoid

Example 3.2.4.

1. *Sphere: The function*

$$\mathbf{x}(\phi, \theta) = (r \sin \phi \cos \theta, r \sin \phi \sin \theta, r \cos \phi), \ 0 < \phi < \pi, 0 < \theta < 2\pi$$

parametrizes the sphere of radius r centered at the origin. We have

$$\begin{cases} \mathbf{x}_\phi = (r \cos \phi \cos \theta, r \cos \phi \sin \theta, -r \sin \phi) \\ \mathbf{x}_\theta = (-r \sin \phi \sin \theta, r \sin \phi \cos \theta, 0). \end{cases}$$

The first fundamental form is

$$\begin{aligned} I &= \begin{pmatrix} \langle \mathbf{x}_\phi, \mathbf{x}_\phi \rangle & \langle \mathbf{x}_\phi, \mathbf{x}_\theta \rangle \\ \langle \mathbf{x}_\theta, \mathbf{x}_\phi \rangle & \langle \mathbf{x}_\theta, \mathbf{x}_\theta \rangle \end{pmatrix} \\ &= \begin{pmatrix} r^2 & 0 \\ 0 & r^2 \sin^2 \phi \end{pmatrix}. \end{aligned}$$

Therefore the surface area of the sphere is

$$\begin{aligned} \int_0^{2\pi} \int_0^{\pi} \sqrt{r^2(r^2 \sin^2 \phi)} d\phi d\theta &= \int_0^{2\pi} \int_0^{\pi} r^2 \sin \phi d\phi d\theta \\ &= \int_0^{2\pi} [-r^2 \cos \phi]_0^{\pi} d\theta \\ &= \int_0^{2\pi} 2r^2 d\theta \\ &= 4\pi r^2. \end{aligned}$$

2. *Torus: The function*

$$\mathbf{x}(\phi, \theta) = ((R + r \sin \phi) \cos \theta, (R + r \sin \phi) \sin \theta, r \cos \phi), 0 < \phi, \theta < 2\pi$$

parametrizes a torus. We have

$$\begin{cases} \mathbf{x}_\phi = (r \cos \phi \cos \theta, r \cos \phi \sin \theta, -r \sin \phi) \\ \mathbf{x}_\theta = (-(R + r \sin \phi) \sin \theta, (R + r \sin \phi) \cos \theta, 0). \end{cases}$$

The first fundamental form is

$$I = \begin{pmatrix} r^2 & 0 \\ 0 & (R + r \sin \phi)^2 \end{pmatrix}.$$

Therefore the surface area of the torus is

$$\begin{aligned} \int_0^{2\pi} \int_0^{2\pi} \sqrt{r^2(R + r \sin \phi)^2} d\phi d\theta &= \int_0^{2\pi} \int_0^{2\pi} r(R + r \sin \phi) d\phi d\theta \\ &= \int_0^{2\pi} [r(R\phi - r \cos \phi)]_0^{2\pi} d\theta \\ &= \int_0^{2\pi} 2\pi r R d\theta \\ &= 4\pi^2 r R. \end{aligned}$$

Note that the surface area of the torus is the product of the circumferences of two circles of radius r and R.

Let's calculate the surface area of surfaces given by graphs of functions.

Theorem 3.2.5 (Surface area of graphs of functions)**.**

1. **Rectangular coordinates***: Let $z = f(x, y)$, $(x, y) \in D \subset \mathbb{R}^2$, be a differentiable function. The surface area of the graph of $z = f(x, y)$ in rectangular coordinates is*

$$A = \iint_D \sqrt{1 + f_x^2 + f_y^2}dxdy.$$

2. **Cylindrical coordinates***: Let $z = f(r, \theta)$, $(r, \theta) \in D \subset \mathbb{R}^+ \times (0, 2\pi)$, be a differentiable function. The surface area of the graph of $z = f(r, \theta)$ in cylindrical coordinates is*

$$A = \iint_D \sqrt{r^2 + r^2 f_r^2 + f_\theta^2}drd\theta.$$

Proof. 1. The surface is parametrized by $\mathbf{x}(x, y) = (x, y, f(x, y))$, $(x, y) \in D$. Then

$$\begin{cases} \mathbf{x}_x = (1, 0, f_x) \\ \mathbf{x}_y = (0, 1, f_y) \end{cases}$$

and the first fundamental form is

$$\begin{aligned} I &= \begin{pmatrix} \langle \mathbf{x}_u, \mathbf{x}_u \rangle & \langle \mathbf{x}_u, \mathbf{x}_v \rangle \\ \langle \mathbf{x}_v, \mathbf{x}_u \rangle & \langle \mathbf{x}_v, \mathbf{x}_v \rangle \end{pmatrix} \\ &= \begin{pmatrix} 1 + f_x^2 & f_x f_y \\ f_x f_y & 1 + f_y^2 \end{pmatrix}. \end{aligned}$$

Therefore the surface area is

$$\begin{aligned} A &= \iint_D \sqrt{\det(I)}dxdy \\ &= \iint_D \sqrt{(1 + f_x^2)(1 + f_y^2) - (f_x f_y)^2}dxdy \\ &= \iint_D \sqrt{1 + f_x^2 + f_y^2}dxdy \end{aligned}$$

2. Parametrize the graph of the function $z = f(r, \theta)$ in cylindrical coordinates by
$$\mathbf{x}(r, \theta) = (r\cos\theta, r\sin\theta, f(r, \theta)),\ (r, \theta) \in D.$$
The rest is left for the reader as exercise.

□

Theorem 3.2.6 (Surface area of surfaces of revolution)**.** *Let $f(z) > 0$, $z \in (a, b)$ be a positive differentiable function. The surface area of the surface obtained by rotating the graph of $x = f(z)$ in the xz-plane about the z axis is*

$$A = 2\pi \int_a^b f\sqrt{1 + f'^2}dz.$$

Proof. The surface is parametrized by $\mathbf{x}(\theta, z) = (f(z)\cos\theta, f(z)\sin\theta, z)$, $(\theta, z) \in (0, 2\pi) \times (a, b)$. The rest is left as exercise for the reader. □

3.3 Second fundamental form and Gaussian curvature

Recall that the vectors $\mathbf{x}_u$ and $\mathbf{x}_v$ are tangent to the parametrized surface $\mathbf{x}(u, v)$ in $\mathbb{R}^3$. A normal vector to the surface is a vector orthogonal to both $\mathbf{x}_u$ and $\mathbf{x}_v$ which can be obtained by taking the cross product of $\mathbf{x}_u$ and $\mathbf{x}_v$.

Definition 3.3.1 (Unit normal vector)**.** *Let $\mathbf{x}(u, v)$ be a regular parametrized surface. The* **unit normal vector** *to the surface is*

$$\mathbf{n} = \frac{\mathbf{x}_u \times \mathbf{x}_v}{\|\mathbf{x}_u \times \mathbf{x}_v\|}.$$

Note that there are two directions, namely $\mathbf{n}$ and $-\mathbf{n}$, normal to a surface in $\mathbb{R}^3$. Changing the order of parameters u, v would invert the direction of the unit normal vector $\mathbf{n}$. A vector is tangent to the surface if it is orthogonal to $\mathbf{n}$. This gives another description of tangent space to the surface.

Proposition 3.3.2. *Let S be a regular surface with parametrization $\mathbf{x}(u, v)$. Let T_pS be the tangent space to the surface at a point $p = \mathbf{x}(u, v)$. Then*

$$T_pS = \{\mathbf{v} \in \mathbb{R}^3 : \langle \mathbf{v}, \mathbf{n} \rangle = 0\}.$$

Next we define the second fundamental form which is, similar to the first fundamental form, a 2×2 matrix valued function of u, v.

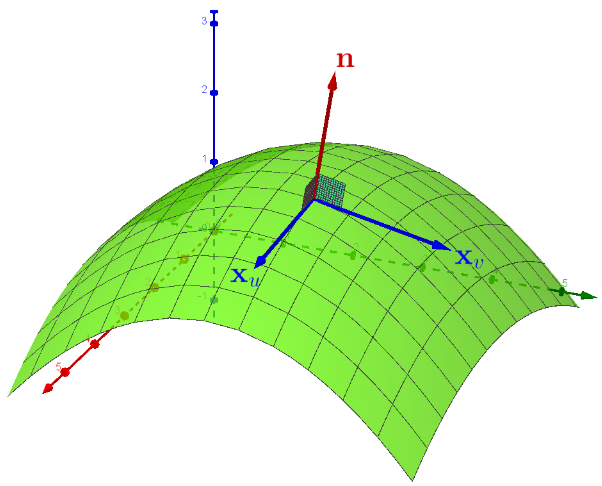

Figure 14: Unit normal vector

Definition 3.3.3 (Second fundamental form)**.** *Let* $\mathbf{x}(u,v)$ *be a regular parametrized surface which has continuous second derivatives. The* **second fundamental form** *is the* 2×2 *matrix valued function*

$$II = \begin{pmatrix} e & f \\ f & g \end{pmatrix} = \begin{pmatrix} \langle \mathbf{x}_{uu}, \mathbf{n}\rangle & \langle \mathbf{x}_{uv}, \mathbf{n}\rangle \\ \langle \mathbf{x}_{vu}, \mathbf{n}\rangle & \langle \mathbf{x}_{vv}, \mathbf{n}\rangle \end{pmatrix} = -\begin{pmatrix} \langle \mathbf{x}_{u}, \mathbf{n}_u\rangle & \langle \mathbf{x}_{u}, \mathbf{n}_v\rangle \\ \langle \mathbf{x}_{v}, \mathbf{n}_u\rangle & \langle \mathbf{x}_{v}, \mathbf{n}_v\rangle \end{pmatrix}.$$

Here we have given two equivalent formulas to calculate the second fundamental form. The two formulas give the same function for the following reason. Observe that $\langle \mathbf{x}_u, \mathbf{n}\rangle = 0$ for any u, v. Differentiating the equality with respect to v, we have

$$\begin{aligned} \frac{\partial}{\partial v}\langle \mathbf{x}_u, \mathbf{n}\rangle &= 0 \\ \langle \mathbf{x}_{uv}, \mathbf{n}\rangle + \langle \mathbf{x}_u, \mathbf{n}_v\rangle &= 0 \\ \langle \mathbf{x}_{uv}, \mathbf{n}\rangle &= -\langle \mathbf{x}_u, \mathbf{n}_v\rangle \end{aligned}$$

We may obtain, in a similar way, the equalities $\langle \mathbf{x}_{uu}, \mathbf{n}\rangle = -\langle \mathbf{x}_u, \mathbf{n}_u\rangle$ and $\langle \mathbf{x}_{vv}, \mathbf{n}\rangle = -\langle \mathbf{x}_v, \mathbf{n}_v\rangle$ and we see that the two formulas in Definition 3.3.3 give the same function. Here we have used an argument basically the same as the proof of Lemma 1.3.35. Note that, similar to first fundamental form I, the second fundamental form II is also a symmetric matrix. This follows from the standard fact in multivariables calculus that when calculating the second order derivative of a function, the order of differentiation does not matter if the function has continuous second derivatives. In particular, we have $\mathbf{x}_{uv} = \mathbf{x}_{vu}$ for any u, v and hence II is symmetric from the first formula in Definition 3.3.3. Now we are ready to introduce the important notion of Gaussian curvature in differential geometry.

Definition 3.3.4 (Gaussian curvature)**.** *Let* $\mathbf{x}(u, v)$ *be a regular parametrized surface which has continuous second derivatives. The* **Gaussian curvature** *of the surface is*

$$K = \frac{\det(II)}{\det(I)} = \frac{eg - f^2}{EG - F^2}$$

where I *is the first fundamental form and* II *is the second fundamental form of the surface.*

Example 3.3.5.

1. *Sphere: A sphere of radius* r *centered at the origin is parametrized by*

$$\mathbf{x}(\phi, \theta) = (r \sin\phi \cos\theta, r \sin\phi \sin\theta, r \cos\phi), \ \ 0 < \phi < \pi, 0 < \theta < 2\pi.$$

We have

$$\begin{cases} \mathbf{x}_\phi = (r \cos\phi \cos\theta, r \cos\phi \sin\theta, -r \sin\phi) \\ \mathbf{x}_\theta = (-r \sin\phi \sin\theta, r \sin\phi \cos\theta, 0) \end{cases}$$

and the first fundamental form is

$$I = \begin{pmatrix} r^2 & 0 \\ 0 & r^2 \sin^2\phi \end{pmatrix}.$$

Now

$$\mathbf{x}_\phi \times \mathbf{x}_\theta = (r^2 \sin^2\phi \cos\theta, r^2 \sin^2\phi \sin\theta, r^2 \sin\phi \cos\phi)$$

and the unit normal vector is

$$\mathbf{n} = \frac{\mathbf{x}_u \times \mathbf{x}_v}{\|\mathbf{x}_u \times \mathbf{x}_v\|} = (\sin\phi \cos\theta, \sin\phi \sin\theta, \cos\phi).$$

Thus

$$\begin{cases} \mathbf{n}_\phi = (\cos\phi \cos\theta, \cos\phi \sin\theta, -\sin\phi) \\ \mathbf{n}_\theta = (-\sin\phi \sin\theta, \sin\phi \cos\theta, 0) \end{cases}$$

and the second fundamental form is

$$\begin{aligned} II &= -\begin{pmatrix} \langle \mathbf{x}_u, \mathbf{n}_u \rangle & \langle \mathbf{x}_u, \mathbf{n}_v \rangle \\ \langle \mathbf{x}_v, \mathbf{n}_u \rangle & \langle \mathbf{x}_v, \mathbf{n}_v \rangle \end{pmatrix} \\ &= \begin{pmatrix} -r & 0 \\ 0 & -r\sin^2\phi \end{pmatrix}. \end{aligned}$$

Therefore the Gaussian curvature of the sphere is

$$K = \frac{\det(II)}{\det(I)} = \frac{r^2 \sin^2\phi}{r^4 \sin^2\phi} = \frac{1}{r^2}.$$

2. *Torus: Let $R > r > 0$ be constants. The function*

$$\mathbf{x}(\phi, \theta) = ((R + r\sin\phi)\cos\theta, (R + r\sin\phi)\sin\theta, r\cos\phi), \;\; 0 < \phi, \theta < 2\pi$$

parametrizes a torus. We have

$$\begin{cases} \mathbf{x}_\phi = (r\cos\phi\cos\theta, r\cos\phi\sin\theta, -r\sin\phi) \\ \mathbf{x}_\theta = (-(R + r\sin\phi)\sin\theta, (R + r\sin\phi)\cos\theta, 0) \end{cases}$$

and the first fundamental form is

$$I = \begin{pmatrix} r^2 & 0 \\ 0 & (R + r\sin\phi)^2 \end{pmatrix}.$$

Now the unit normal vector is

$$\mathbf{n} = \frac{\mathbf{x}_u \times \mathbf{x}_v}{\|\mathbf{x}_u \times \mathbf{x}_v\|} = (\sin\phi\cos\theta, \sin\phi\sin\theta, \cos\phi).$$

and the second derivatives of $\mathbf{x}$ *are*

$$\begin{cases} \mathbf{x}_{\phi\phi} = (-r\sin\phi\cos\theta, -r\sin\phi\sin\theta, -r\cos\phi) \\ \mathbf{x}_{\phi\theta} = (-r\cos\phi\sin\theta, r\cos\phi\cos\theta, 0) \\ \mathbf{x}_{\theta\theta} = (-(R + r\sin\phi)\cos\theta, -(R + r\sin\phi)\sin\theta, 0) \end{cases}$$

Thus the second fundamental form is

$$\begin{aligned} II &= \begin{pmatrix} \langle \mathbf{x}_{\phi\phi}, \mathbf{n} \rangle & \langle \mathbf{x}_{\phi\theta}, \mathbf{n} \rangle \\ \langle \mathbf{x}_{\theta\phi}, \mathbf{n} \rangle & \langle \mathbf{x}_{\theta\theta}, \mathbf{n} \rangle \end{pmatrix} \\ &= \begin{pmatrix} -r & 0 \\ 0 & -(R + r\sin\phi)\sin\phi \end{pmatrix} \end{aligned}$$

Therefore the Gaussian curvature of the torus is

$$K = \frac{\det(II)}{\det(I)} = \frac{r(R + r\sin\phi)\sin\phi}{r^2(R + r\sin\phi)^2} = \frac{\sin\phi}{r(R + r\sin\phi)}.$$

Note that for torus, we have $K > 0$ when $0 < \phi < \pi$, $K = 0$ when $\phi = 0, \pi$ and $K < 0$ when $\pi < \phi < 2\pi$.

Proposition 3.3.6 (Curvature of graphs of functions)**.**

1. *Let $f(x, y)$, $(x, y) \in D \subset \mathbb{R}^2$, be a function with continuous second derivatives. The Gaussian curvature of the graph of $z = f(x, y)$ in rectangular coordinates is*

$$K(x, y) = \frac{f_{xx}f_{yy} - f_{xy}^2}{(1 + f_x^2 + f_y^2)^2}.$$

2. *Let $f(r, \theta)$, $(r, \theta) \in D \subset \mathbb{R}^+ \times (0, 2\pi)$, be a function with continuous second derivatives. The Gaussian curvature of the graph of $z = f(r, \theta)$ in cylindrical coordinates is*

$$K(r, \theta) = \frac{r^2 f_{rr}(r f_r + f_{\theta\theta}) - (r f_{r\theta} - f_\theta)^2}{(r^2 + r^2 f_r^2 + f_\theta^2)^2}.$$

Proof. 1. Parametrize the graph of the function $z = f(x, y)$ in rectangular coordinates by

$$\mathbf{x}(u, v) = (u, v, f(u, v)).$$

We have

$$\begin{cases} \mathbf{x}_u = (1, 0, f_x) \\ \mathbf{x}_v = (0, 1, f_y) \end{cases}$$

and the first fundamental form is

$$I = \begin{pmatrix} 1 + f_x^2 & f_x f_y \\ f_x f_y & 1 + f_y^2 \end{pmatrix}.$$

Now the unit normal vector is

$$\mathbf{n} = \frac{\mathbf{x}_u \times \mathbf{x}_v}{\|\mathbf{x}_u \times \mathbf{x}_v\|} = (-\frac{f_x}{\sqrt{1 + f_x^2 + f_y^2}}, -\frac{f_y}{\sqrt{1 + f_x^2 + f_y^2}}, \frac{1}{\sqrt{1 + f_x^2 + f_y^2}}).$$

and the second derivatives of $\mathbf{x}$ are

$$\begin{cases} \mathbf{x}_{uu} = (0, 0, f_{xx}) \\ \mathbf{x}_{uv} = (0, 0, f_{xy}) \\ \mathbf{x}_{vv} = (0, 0, f_{yy}) \end{cases}$$

Thus the second fundamental form is

$$II = \begin{pmatrix} \dfrac{f_{xx}}{\sqrt{1 + f_x^2 + f_y^2}} & \dfrac{f_{xy}}{\sqrt{1 + f_x^2 + f_y^2}} \\ \dfrac{f_{xy}}{\sqrt{1 + f_x^2 + f_y^2}} & \dfrac{f_{yy}}{\sqrt{1 + f_x^2 + f_y^2}} \end{pmatrix}$$

Therefore the Gaussian curvature of the surface is

$$K = \frac{\det(II)}{\det(I)} = \frac{f_{xx} f_{yy} - f_{xy}^2}{(1 + f_x^2 + f_y^2)^2}.$$

2. Parametrize the graph of the function $z = f(r, \theta)$ in cylindrical coordinates by

$$\mathbf{x}(r, \theta) = (r\cos\theta, r\sin\theta, f(r, \theta)).$$

The rest is left for the reader as exercise.

□

Proposition 3.3.7 (Gaussian curvature of surfaces of revolution)**.**

1. **By graph of function**: *Let $f(z)$, $z \in (a, b)$, be a function with continuous second derivative. The Gaussian curvature of the surface obtained by rotating the graph of $x = f(z)$ on the xz-plane about the z axis is*

$$K(z) = -\frac{f''}{f(1 + f'^2)^2}.$$

2. **By parametrized curve**: *Let $(\varphi(u), \psi(u))$, $u \in (a, b)$, be a regular parametrized curve. The Gaussian curvature of the surface obtained by rotating the curve $(x, z) = (\varphi(u), \psi(u))$ on the xz-plane about the z axis is*

$$K(u) = \frac{\psi'(\varphi'\psi'' - \varphi''\psi')}{\varphi(\varphi'^2 + \psi'^2)^2}.$$

3. **By arc length parametrized curve**: *Let $(\varphi(s), \psi(s))$, $s \in (a,b)$, be an arc length parametrized curve. The Gaussian curvature of the surface obtained by rotating the curve $(x,z) = (\varphi(s), \psi(s))$ on the xz-plane about the z axis is*

$$K(s) = -\frac{\varphi''}{\varphi}.$$

Proof. 1. Parametrize the surface of revolution by graph of function $x = f(z)$ by

$$\mathbf{x}(u,\theta) = (f(u)\cos\theta, f(u)\sin\theta, u).$$

The derivatives of $\mathbf{x}$ are

$$\begin{cases} \mathbf{x}_u = (f'\cos\theta, f'\sin\theta, 1) \\ \mathbf{x}_\theta = (-f\sin\theta, f\cos\theta, 0) \\ \mathbf{x}_{uu} = (f''\cos\theta, f''\sin\theta, 0) \\ \mathbf{x}_{u\theta} = (-f'\sin\theta, f'\cos\theta, 0) \\ \mathbf{x}_{\theta\theta} = (-f\cos\theta, -f\sin\theta, 0) \end{cases}$$

The first fundamental form is

$$I = \begin{pmatrix} 1+f'^2 & 0 \\ 0 & f^2 \end{pmatrix}.$$

The unit normal vector is

$$\mathbf{n} = (-\frac{\cos\theta}{\sqrt{1+f'^2}}, -\frac{\sin\theta}{\sqrt{1+f'^2}}, \frac{f'}{\sqrt{1+f'^2}})$$

and the second fundamental form is

$$II = \begin{pmatrix} -\frac{f''}{\sqrt{1+f'^2}} & 0 \\ 0 & \frac{f}{\sqrt{1+f'^2}} \end{pmatrix}.$$

Therefore the Gaussian curvature is

$$K = \frac{\det(II)}{\det(I)} = -\frac{f''}{f(1+f'^2)^2}.$$

2. Parametrize the surface of revolution by parametrized curve $(x, z) = (\varphi(u), \psi(u))$ by

$$\mathbf{x}(u, \theta) = (\varphi(u)\cos\theta, \varphi(u)\sin\theta, \psi(u)).$$

The rest are left for the reader as exercise.

3. Exercise.

$\square$

Example 3.3.8 (Catenoid)**.** *Consider the surface obtained by rotating the catenary $x = f(z) = \cosh z$ in the xz-plane about the z axis which is called* **catenoid**. *The Gaussian curvature of catenoid is*

$$\begin{aligned} K(z) &= -\frac{f''}{f(1+f'^2)^2} \\ &= -\frac{\cosh z}{\cosh z(1+\sinh z^2)^2} \\ &= -\frac{1}{\cosh^4 z} \end{aligned}$$

Example 3.3.9 (Torus)**.** *Show that the Gaussian curvature of the torus obtained by rotating the arc length parametrized curve*

$$(x, z) = (\varphi(s), \psi(s)) = \left(R + r\sin\frac{s}{r}, r\cos\frac{s}{r}\right), \ s \in (0, 2\pi)$$

about the z-axis is

$$K = \frac{\sin\frac{s}{r}}{r(R + r\sin\frac{s}{r})}.$$

Proof. Observe that

$$\begin{cases} \varphi' = \cos\frac{s}{r}, \\ \varphi'' = -\frac{1}{r}\sin\frac{s}{r} \end{cases}$$

By Proposition 3.3.7, we have

$$\begin{aligned} K &= -\frac{\varphi''}{\varphi} \\ &= \frac{\sin\frac{s}{r}}{r(R + r\sin\frac{s}{r})} \end{aligned}$$

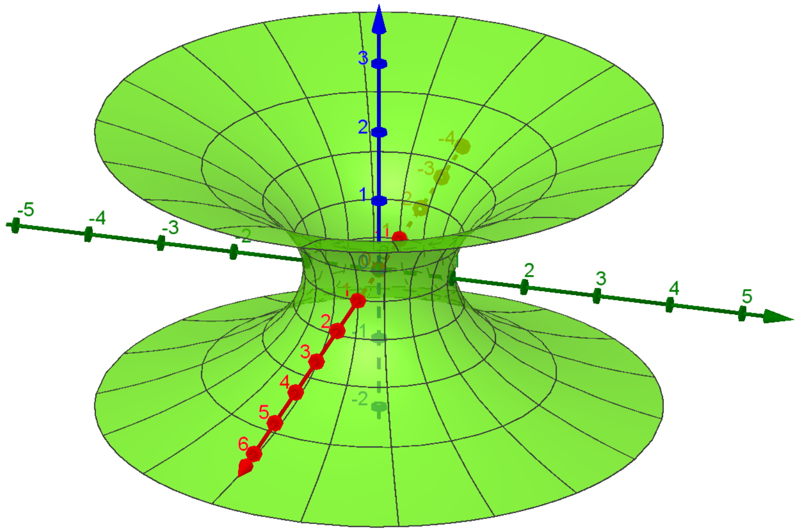

Figure 15: Catenoid

□

Example 3.3.10 (Pseudosphere)**.** *Consider the surface obtained by rotating the tractrix (Example 2.2.10)*

$$(x, z) = (\varphi(t), \psi(t)) = (\mathrm{sech} t, t - \tanh t)\,, \;\; t > 0$$

about the z-axis. This surface is called the **pseudosphere***. Show that the pseudosphere has constant Gaussian curvature equal to -1.*

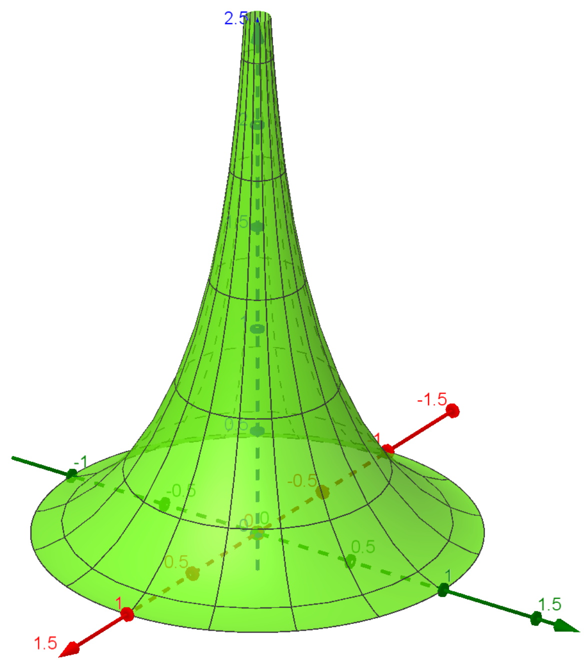

Figure 16: Pseudosphere

Proof. Observe that

$$
\begin{aligned}
\varphi' &= -\text{sech}t \tanh t \\
\varphi'' &= \text{sech}t \tanh^2 t - \text{sech}^3 t \\
&= \text{sech}t(\tanh^2 t - \text{sech}^2 t) \\
&= \text{sech}t(1 - 2\text{sech}^2 t) \\
\psi' &= 1 - \text{sech}^2 t \\
&= \tanh^2 t \\
\psi'' &= 2 \tanh t \text{sech}^2 t
\end{aligned}
$$

Note that

$$\varphi'^2 + \psi'^2 = \text{sech}^2 t \tanh^2 t + \tanh^4 t = \tanh^2 t(\text{sech}^2 t + \tanh^2 t) = \tanh^2 t.$$

By Proposition 3.3.7, we have

$$\begin{aligned} K &= \frac{\psi'(\varphi'\psi'' - \varphi''\psi')}{\varphi(\varphi'^2 + \psi'^2)^2} \\ &= \frac{\tanh^2 t(-\text{sech} t \tanh t(2\tanh t \text{sech}^2 t) - \text{sech} t(1 - 2\text{sech}^2 t)\tanh^2 t)}{\text{sech} t \tanh^4 t} \\ &= \frac{-\tanh^2 t(\text{sech} t \tanh^2 t)}{\text{sech} t \tanh^4 t} \\ &= -1. \end{aligned}$$

Alternative, we may use the arc length parametrization of the tractrix given by (Proposition 2.2.10)

$$(\varphi(s), \psi(s)) = (e^{-s}, \ln(e^s + \sqrt{e^{2s} - 1}) - \sqrt{1 - e^{-2s}}), \ s > 0.$$

Then the Gaussian curvature of the pseudosphere is

$$K = -\frac{\varphi''(s)}{\varphi(s)} = -\frac{e^{-s}}{e^{-s}} = -1.$$

□

We conclude this section by stating a formula for Gaussian curvature which involves only the first fundamental form and its derivatives but the the second fundamental form.

Theorem 3.3.11. *Let* $\mathbf{x}(u, v)$ *be a regular parametrized surface. Suppose* $F = 0$, *i.e., the first fundamental form of* $\mathbf{x}(u, v)$ *is*

$$I = \begin{pmatrix} E & 0 \\ 0 & G \end{pmatrix}.$$

Then the Gaussian curvature of $\mathbf{x}(u, v)$ *is*

$$K = -\frac{1}{2\sqrt{EG}}\left[\left(\frac{E_v}{\sqrt{EG}}\right)_v + \left(\frac{G_u}{\sqrt{EG}}\right)_u\right].$$

Example 3.3.12 (Helicoid)**.** *Show that the Gaussian curvature of the* **helicoid** *parametrized by*

$$\mathbf{x}(u,\theta) = (u\cos\theta, u\sin\theta, \theta),\ \ u,\theta \in \mathbb{R},$$

is

$$K = -\frac{1}{(1+u^2)^2}.$$

Proof. The first derivatives of $\mathbf{x}$ are

$$\begin{cases} \mathbf{x}_u = (\cos\theta, \sin\theta, 0), \\ \mathbf{x}_\theta = (-u\sin\theta, u\cos\theta, 1). \end{cases}$$

Thus the first fundamental form is

$$I = \begin{pmatrix} 1 & 0 \\ 0 & 1+u^2 \end{pmatrix}$$

Now

$$\begin{cases} E_\theta = 0 \\ G_u = \dfrac{\partial}{\partial u}(1+u^2) = 2u, \end{cases}$$

Therefore by Theorem 3.3.11, the Gaussian curvature is

$$\begin{aligned} K &= -\frac{1}{2\sqrt{EG}}\left[\left(\frac{E_\theta}{\sqrt{EG}}\right)_\theta + \left(\frac{G_u}{\sqrt{EG}}\right)_u\right] \\ &= -\frac{1}{2\sqrt{1+u^2}}\left(\frac{2u}{\sqrt{1+u^2}}\right)_u \\ &= -\frac{1}{\sqrt{1+u^2}}\left(\frac{\sqrt{1+u^2} - u(\frac{u}{\sqrt{1+u^2}})}{1+u^2}\right) \\ &= -\frac{1}{(1+u^2)^2} \end{aligned}$$

□

We summarize the formulas for Gaussian curvature in the following proposition.

Proposition 3.3.13 (Formulas for Gaussian curvature)**.** *Let S be a regular surface and K be its Gaussian curvature.*

1. *Definition:*

$$K = \frac{\det(II)}{\det(I)} = \frac{eg - f^2}{EG - F^2}$$

2. *Graph of functions:*

   (a) *Rectangular coordinates: For $z = f(x, y)$,*

   $$K = \frac{f_{xx} f_{yy} - f_{xy}^2}{(1 + f_x^2 + f_y^2)^2}.$$

   (b) *Cylindrical coordinates: For $z = f(r, \theta)$,*

   $$K = \frac{r^2 f_{rr}(r f_r + f_{\theta\theta}) - (r f_{r\theta} - f_\theta)^2}{(r^2 + r^2 f_r^2 + f_\theta^2)^2}.$$

3. *Surface of revolution:*

   (a) *By graph of function: For $\mathbf{x}(u, \theta) = (f(u)\cos\theta, f(u)\sin\theta, u)$,*

   $$K(u, \theta) = K(u) = -\frac{f''}{f(1 + f'^2)^2}.$$

   (b) *By parametrized curve: For $\mathbf{x}(u, \theta) = (\varphi(u)\cos\theta, \varphi(u)\sin\theta, \psi(u))$,*

   $$K(u, \theta) = K(u) = \frac{\psi'(\varphi'\psi'' - \varphi''\psi')}{\varphi(\varphi'^2 + \psi'^2)^2}.$$

   (iii) *By arc length parametrized curve: For $\mathbf{x}(u, \theta) = (\varphi(u)\cos\theta, \varphi(u)\sin\theta, \psi(u))$, with $\varphi'^2 + \psi'^2 = 1$,*

   $$K(u, \theta) = K(u) = -\frac{\varphi''}{\varphi}$$

4. *Parametrized surface with $F = 0$:*

$$K(u, v) = -\frac{1}{2\sqrt{EG}}\left[\left(\frac{E_v}{\sqrt{EG}}\right)_v + \left(\frac{G_u}{\sqrt{EG}}\right)_u\right].$$

3.4 Gauss map and its differential

To understand the geometric meaning of the Gaussian curvature, we introduce the Gauss map which is defined simply by the unit normal vector.

Definition 3.4.1 (Gauss map)**.** *Let S be a regular surface in $\mathbb{R}^3$ with regular parametrization $\mathbf{x}(u,v)$. For each $p = \mathbf{x}(u,v)$, we associate the unit normal vector $\mathbf{n}(p)$ to p. This defines a map $\mathbf{n} : S \to S^2$ from the surface S to the unit sphere $S^2 = \{(x,y,z) : x^2 + y^2 + z^2 = 1\}$ and is called the* **Gauss map** *of S.*

The Gauss map has the following distinguish properties.

Proposition 3.4.2. *Let S be a regular surface with regular parametrization $\mathbf{x}(u,v)$ and $\mathbf{n} : S \to S^2$ be the Gauss map which sends a point $p \in S$ to the unit normal vector $\mathbf{n} = \mathbf{n}(p)$ which is a point on the unit sphere S^2. Let $p \in S$ be any point on the surface S. Then the following statements hold.*

1. *The unit normal vector $\mathbf{n} = \mathbf{n}(p)$ to the surface S is a unit normal vector to the unit sphere S^2 at $\mathbf{n}$.*

2. *The tangent space to the unit sphere S^2 at $\mathbf{n} = \mathbf{n}(p)$ is equal to the tangent space to the surface S at p. In other words,*

$$T_{\mathbf{n}}S^2 = T_pS.$$

3. *The vectors $\mathbf{n}_u(p)$ and $\mathbf{n}_v(p)$ are tangent to S at p. In other words,*

$$\mathbf{n}_u, \mathbf{n}_v \in T_pS$$

which means both $\mathbf{n}_u$, $\mathbf{n}_v$ can be written as linear combinations of $\mathbf{x}_u$ and $\mathbf{x}_v$.

Proof. 1. By writing down a regular parametrization of S^2, one can prove easily that the unit normal vector to S^2 at any point $\mathbf{v} \in S^2$ is $\mathbf{v}$ itself.

2. Observe that

$$T_pS = \{\mathbf{v} \in \mathbb{R}^3 : \langle \mathbf{v}, \mathbf{n} \rangle\},$$

it suffice to show that for any $\mathbf{n} \in S^2$, we have

$$T_{\mathbf{n}}S^2 = \{\mathbf{v} \in \mathbb{R}^3 : \langle \mathbf{v}, \mathbf{n} \rangle\}.$$

Let $\mathbf{n}(\theta, \phi)$ be a parametrization of S^2 which is regular at $\mathbf{n}$. By taking derivatives of the constant function $\|\mathbf{n}(\theta, \phi)\|^2 = 1$, we have

$$\langle \mathbf{n}_\theta, \mathbf{n} \rangle = \langle \mathbf{n}_\phi, \mathbf{n} \rangle = 0.$$

It follows that $\mathbf{n}$ is normal to S^2 at $\mathbf{n}$ and therefore

$$T_{\mathbf{n}} S^2 = \{\mathbf{v} \in \mathbb{R}^3 : \langle \mathbf{v}, \mathbf{n} \rangle\} = T_p S.$$

3. Using an argument (see Lemma 1.3.35, proof of Theorem 2.4.4 and the exposition after Definition 3.3.3) which has been used for many times, we see that

$$\langle \mathbf{n}_u, \mathbf{n} \rangle = \langle \mathbf{n}_v, \mathbf{n} \rangle = 0$$

are constantly equal to zero. Therefore we have $\mathbf{n}_u, \mathbf{n}_v \in T_p S$.

□

A consequence of the above proposition is that since both $\mathbf{n}_u$ and $\mathbf{n}_v$ are orthogonal to $\mathbf{n}$, their cross product $\mathbf{n}_u \times \mathbf{n}_v$ is normal to the surface S and thus is a scalar multiple of $\mathbf{x}_u \times \mathbf{x}_v$. This multiple is exactly the Gaussian curvature.

Theorem 3.4.3. *Let $\mathbf{x}(u, v)$ be a regular parametrized surface and $\mathbf{n}(u, v)$ be the unit normal vector at $\mathbf{x}(u, v)$. Then*

$$\mathbf{n}_u \times \mathbf{n}_v = K \mathbf{x}_u \times \mathbf{x}_v$$

where K is the Gaussian curvature of the surface.

Proof. Since $\mathbf{n}_u, \mathbf{n}_v \in T_p S$, we have $\mathbf{n}_u \times \mathbf{n}_v$ is normal to the surface S and thus

$$\mathbf{n}_u \times \mathbf{n}_v = c \mathbf{x}_u \times \mathbf{x}_v$$

for some real number c which is a function on S. By Proposition 1.3.17, we have

$$\det(I) = \langle \mathbf{x}_u \times \mathbf{x}_v, \mathbf{x}_u \times \mathbf{x}_v \rangle$$

and

$$\begin{aligned} \det(II) &= \langle \mathbf{x}_u \times \mathbf{x}_v, \mathbf{n}_u \times \mathbf{n}_v \rangle \\ &= \langle \mathbf{x}_u \times \mathbf{x}_v, c\mathbf{x}_u \times \mathbf{x}_v \rangle \\ &= c \langle \mathbf{x}_u \times \mathbf{x}_v, \mathbf{x}_u \times \mathbf{x}_v \rangle \\ &= c \det(I) \end{aligned}$$

Note that $\det(I) > 0$ and we obtain

$$c = \frac{\det(II)}{\det(I)} = K$$

where K is the Gaussian curvature of the surface. □

The Gaussian curvature has a geometric interpretation as follows. Let S be a regular surface parametrized by $\mathbf{x}(u, v)$, $(u, v) \in D$. Suppose $\Omega \subset S$ is a region on S which is an open connected subset of S. We define $A(\Omega)$ as the surface area of $\Omega \subset S$ and $\sigma(\Omega)$ as the surface area of the image $\mathbf{n}(\Omega) \subset S^2$ of Ω under the Gauss map. Now consider the small region

$$\Omega = \{\mathbf{x}(s, t) : u < s < u + \Delta u, v < t < \Delta v\} \subset S$$

on S which is the image of a small rectangle $(u, u + \Delta u) \times (v + \Delta v) \subset D$. We would like to compare the area of this small region $\Omega \subset S$ and the signed area[10] of its image $\mathbf{n}(\Omega) \subset S^2$ under the Gauss map. The area of Ω can be approximated by the parallelogram spanned by $\Delta u \mathbf{x}_u$ and $\Delta v \mathbf{x}_v$ which has area

$$\Delta A \approx \|\mathbf{x}_u \times \mathbf{x}_v\| \Delta u \Delta v.$$

On the other hand, since $\mathbf{n}_u, \mathbf{n}_v$ are tangent to S, we have

$$\mathbf{n}_u \times \mathbf{n}_v = \langle \mathbf{n}_u \times \mathbf{n}_v, \mathbf{n} \rangle \mathbf{n} = \pm \|\mathbf{n}_u \times \mathbf{n}_v\| \mathbf{n}.$$

Here the sign is positive if $\mathbf{n}$ preserves the orientation at $p = \mathbf{x}(u, v)$ or equivalently the Gaussian curvature is positive at p and the sign is negative if $\mathbf{n}$ reserves the orientation at p or equivalently the Gaussian curvature is negative at p. Thus the signed area of $\mathbf{n}(\Omega)$ can be approximated by

$$\Delta \sigma \approx \langle \Delta u \mathbf{n}_u \times \Delta v \mathbf{n}_v, \mathbf{n} \rangle = \langle \mathbf{n}_u \times \mathbf{n}_v, \mathbf{n} \rangle \Delta u \Delta v.$$

[10]The signed area of $\mathbf{n}(\Omega)$ is positive if $\mathbf{n}$ preserves orientation and is negative if $\mathbf{n}$ reverses orientation.

Now if we let $\Delta u, \Delta v$ go to zero, the ratio of these two areas would be

$$\begin{aligned}
\frac{d\sigma}{dA} &= \lim_{\Delta u, \Delta v \to 0} \frac{\Delta \sigma}{\Delta A} \\
&= \lim_{\Delta u, \Delta v \to 0} \frac{\langle \mathbf{n}_u \times \mathbf{n}_v, \mathbf{n} \rangle \Delta u \Delta v}{\|\mathbf{x}_u \times \mathbf{x}_v\| \Delta u \Delta v} \\
&= \frac{\langle \mathbf{n}_u \times \mathbf{n}_v, \|\mathbf{x}_u \times \mathbf{x}_v\| \mathbf{n} \rangle}{\|\mathbf{x}_u \times \mathbf{x}_v\|^2} \\
&= \frac{\langle \mathbf{n}_u \times \mathbf{n}_v, \mathbf{x}_u \times \mathbf{x}_v \rangle}{\|\mathbf{x}_u \times \mathbf{x}_v\|^2} \\
&= \frac{\det(II)}{\det(I)} \quad \text{(Proposition 1.3.17)} \\
&= K.
\end{aligned}$$

So we have the following geometric interpretation of Gaussian curvature which can be thought of as an analogue of Proposition 2.3.12 for surface.

Proposition 3.4.4. *Let S be a regular surface with parametrization $\mathbf{x}(u, v)$, $(u, v) \in D$. Let A and σ be the signed surface area function on S and S^2 respectively. Then we have*

$$\frac{d\sigma}{dA} = K$$

where K is the Gaussian curvature.

We may also understand the Gaussian curvature through the differential of Gauss map, which is a linear operator on the tangent space T_pS induced naturally by the Gauss map. Let $f : S_1 \to S_2$ be a differentiable map from regular surface S_1 to regular surface S_2. Let $\mathbf{x}_1(u, v)$ be a regular parametrization of S_1. Then $\mathbf{x}_2(u, v) = f(\mathbf{x}_1(u, v)$ gives a regular parametrization of S_2. For each $p \in S_1$, we define a function $df_p : T_pS_1 \to T_{f(p)}S_2$ by

$$df_p \left(\alpha \frac{\partial \mathbf{x}_1}{\partial u} + \beta \frac{\partial \mathbf{x}_1}{\partial v} \right) = \alpha \frac{\partial \mathbf{x}_2}{\partial u} + \beta \frac{\partial \mathbf{x}_2}{\partial v}$$

which is a linear transformation from T_pS_1 to $T_{f(p)}S_2$ and is called the differential of f at p. One can show that d_pf does not depends on the parametrization $x_1(u, v)$ of S_1. For if $\mathbf{x}_1(s, t)$ is another regular parametrization of S_1, by chain rule in multivariable calculus, we have

$$\begin{cases} \dfrac{\partial \mathbf{x}_1}{\partial s} = \dfrac{\partial u}{\partial s}\dfrac{\partial \mathbf{x}_1}{\partial u} + \dfrac{\partial v}{\partial s}\dfrac{\partial \mathbf{x}_1}{\partial v} \\ \dfrac{\partial \mathbf{x}_1}{\partial t} = \dfrac{\partial u}{\partial t}\dfrac{\partial \mathbf{x}_1}{\partial u} + \dfrac{\partial v}{\partial t}\dfrac{\partial \mathbf{x}_1}{\partial v} \end{cases}$$

which implies

$$\begin{aligned} df_p\left(\frac{\partial \mathbf{x}_1}{\partial s}\right) &= df_p\left(\frac{\partial u}{\partial s}\frac{\partial \mathbf{x}_1}{\partial u} + \frac{\partial v}{\partial s}\frac{\partial \mathbf{x}_1}{\partial v}\right) \\ &= \frac{\partial u}{\partial s}\frac{\partial \mathbf{x}_2}{\partial u} + \frac{\partial v}{\partial s}\frac{\partial \mathbf{x}_2}{\partial v} \\ &= \frac{\partial \mathbf{x}_2}{\partial s} \end{aligned}$$

and similarly $df_p\left(\frac{\partial \mathbf{x}_1}{\partial t}\right) = \frac{\partial \mathbf{x}_2}{\partial t}$.

Now for the case of Gauss map $\mathbf{n} : S \to S^2$, we have $T_{\mathbf{n}(p)}S^2 = T_pS$ for any $p \in S$ (Proposition 3.4.2). Therefore the differential $d\mathbf{n}_p : T_pS \to T_pS$ of Gauss map at p is a linear transformation from T_pS to itself, in other words, a linear operator on T_pS.

Definition 3.4.5 (Differential of Gauss map)**.** *Let S be a regular surface in $\mathbb{R}^3$ with regular parametrization $\mathbf{x}(u,v)$. For each $p \in S$, define $d\mathbf{n}_p : T_pS \to T_pS$ called the* **differential of Gauss map** *by*

$$d\mathbf{n}_p(\alpha\mathbf{x}_u + \beta\mathbf{x}_v) = \alpha\mathbf{n}_u + \beta\mathbf{n}_v$$

for any real numbers $\alpha, \beta \in \mathbb{R}$.

The differential of Gauss map measures how rigorously the Gauss map, that is the unit normal vector, bends near p along different directions. There are two special directions which somehow determine the local geometry of the surface.

Definition 3.4.6 (Principal curvatures and principal directions)**.** *Let S be a regular surface and $p \in S$. Let $\boldsymbol{\xi}_1, \boldsymbol{\xi}_2 \in T_pS$ be two linearly independent eigenvectors of the differential $d\mathbf{n}_p : T_pS \to T_pS$ of Gauss map at p and κ_1, κ_2 be negative of the associated eigenvalues respectively. In other words,*

$$\begin{cases} d\mathbf{n}_p(\boldsymbol{\xi}_1) = -\kappa_1\boldsymbol{\xi}_1 \\ d\mathbf{n}_p(\boldsymbol{\xi}_2) = -\kappa_2\boldsymbol{\xi}_2 \end{cases}.$$

Then we say that κ_1, κ_2 *are the* **principal curvatures** *of* S *at* p*, and* $\boldsymbol{\xi}_1, \boldsymbol{\xi}_2$ *are the corresponding* **principal directions**.

One may wonder whether we can always find two distinct principal directions, in other words, two linearly independent eigenvectors for $d\mathbf{n}_p$ at any $p \in S$. In fact, once we prove that $d\mathbf{n}_p$ is self-adjoint, it will follow that there exists two orthogonal principal directions.

Theorem 3.4.7 (Self-adjointness of differential of Gauss map)**.** *The differential of Gauss map* $d\mathbf{n}_p : T_pS \to T_pS$ *is self-adjoint. In other words, for any* $\mathbf{u}, \mathbf{v} \in T_pS$*, we have*

$$\langle d\mathbf{n}_p(\mathbf{u}), \mathbf{v}\rangle = \langle \mathbf{u}, d\mathbf{n}_p(\mathbf{v})\rangle.$$

Proof. It suffices to check that

$$\begin{aligned} \langle d\mathbf{n}_p(\mathbf{x}_u), \mathbf{x}_v\rangle &= \langle \mathbf{n}_u, \mathbf{x}_v\rangle \\ &= -\langle \mathbf{n}, \mathbf{x}_{vu}\rangle \quad \text{(Lemma 1.3.35)} \\ &= -\langle \mathbf{n}, \mathbf{x}_{uv}\rangle \\ &= \langle \mathbf{n}_v, \mathbf{x}_u\rangle \quad \text{(Lemma 1.3.35)} \\ &= \langle d\mathbf{n}_p(\mathbf{x}_v), \mathbf{x}_u\rangle \\ &= \langle \mathbf{x}_u, d\mathbf{n}_p(\mathbf{x}_v)\rangle \end{aligned}$$

□

Now applying the spectral theorem for self-adjoint operator (Theorem 1.6.15) to $d\mathbf{n}_p$, we obtain the following theorem.

Theorem 3.4.8. *Let* S *be a regular surface in* $\mathbb{R}^3$ *and* $p \in S$*. Then there exists principal directions* $\boldsymbol{\xi}_1, \boldsymbol{\xi}_2 \in T_pS$ *which constitute an orthonormal basis for* T_pS*.*

Next we find a matrix representation (Definition 1.6.3) of the linear operator $d\mathbf{n}_p$ which can be expressed in terms of the first and second fundamental forms.

Proposition 3.4.9. *The matrix representation of* $d\mathbf{n}_p$ *with respect to basis* $\mathbf{x}_u, \mathbf{x}_v$ *is*

$$-(II)(I^{-1}) = -\frac{1}{EG - F^2}\begin{pmatrix} eG - fF & fE - eF \\ fG - gF & gE - fF \end{pmatrix}.$$

In other words, we have

$$\begin{cases} d\mathbf{n}_p(\mathbf{x}_u) = a\mathbf{x}_u + b\mathbf{x}_v \\ d\mathbf{n}_p(\mathbf{x}_v) = c\mathbf{x}_u + d\mathbf{x}_v \end{cases}$$

where

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} = -\frac{1}{EG - F^2} \begin{pmatrix} eG - fF & fE - eF \\ fG - gF & gE - fF \end{pmatrix}.$$

Proof. Let a, b, c, d be real numbers such that

$$\begin{cases} d\mathbf{n}_p(\mathbf{x}_u) = \mathbf{n}_u = a\mathbf{x}_u + b\mathbf{x}_v \\ d\mathbf{n}_p(\mathbf{x}_v) = \mathbf{n}_v = c\mathbf{x}_u + d\mathbf{x}_v \end{cases}.$$

It follows that

$$\begin{aligned} \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} \mathbf{x}_u \\ \mathbf{x}_v \end{pmatrix} &= \begin{pmatrix} \mathbf{n}_u \\ \mathbf{n}_v \end{pmatrix} \\ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \langle \begin{pmatrix} \mathbf{x}_u \\ \mathbf{x}_v \end{pmatrix}, \begin{pmatrix} \mathbf{x}_u & \mathbf{x}_v \end{pmatrix} \rangle &= \langle \begin{pmatrix} \mathbf{n}_u \\ \mathbf{n}_v \end{pmatrix}, \begin{pmatrix} \mathbf{x}_u & \mathbf{x}_v \end{pmatrix} \rangle \\ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} \langle \mathbf{x}_u, \mathbf{x}_u \rangle & \langle \mathbf{x}_u, \mathbf{x}_v \rangle \\ \langle \mathbf{x}_v, \mathbf{x}_u \rangle & \langle \mathbf{x}_v, \mathbf{x}_v \rangle \end{pmatrix} &= \begin{pmatrix} \langle \mathbf{n}_u, \mathbf{x}_u \rangle & \langle \mathbf{n}_u, \mathbf{x}_v \rangle \\ \langle \mathbf{n}_v, \mathbf{x}_u \rangle & \langle \mathbf{n}_v, \mathbf{x}_v \rangle \end{pmatrix} \\ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} E & F \\ F & G \end{pmatrix} &= -\begin{pmatrix} e & f \\ f & g \end{pmatrix} \end{aligned}$$

Therefore

$$\begin{aligned} \begin{pmatrix} a & b \\ c & d \end{pmatrix} &= -\begin{pmatrix} e & f \\ f & g \end{pmatrix} \begin{pmatrix} E & F \\ F & G \end{pmatrix}^{-1} \\ &= -\begin{pmatrix} e & f \\ f & g \end{pmatrix} \frac{1}{EG - F^2} \begin{pmatrix} G & -F \\ -F & E \end{pmatrix} \\ &= -\frac{1}{EG - F^2} \begin{pmatrix} eG - fF & fE - eF \\ fG - gF & gE - fF \end{pmatrix} \end{aligned}$$

□

Now we can express Gaussian curvature in terms of principal curvatures.

Theorem 3.4.10. *Let S be a regular surface and K be the Gaussian curvature of S. Then for any $p \in S$,*

$$K(p) = \det(d\mathbf{n}_p) = \kappa_1 \kappa_2$$

where $\det(d\mathbf{n}_p)$ is the determinant (Definition 1.6.6) of $d\mathbf{n}_p$ and κ_1, κ_2 are the principal curvatures of S at p.

Proof. Since $d\mathbf{n}_p$ is represented by the matrix $-(II)(I^{-1})$, we have

$$\det(d\mathbf{n}_p) = \det(-(II)(I^{-1})) = \frac{\det(II)}{\det(I)} = K.$$

Since κ_1, κ_2 are the eigenvalues of $d\mathbf{n}_p$, we have $K = \det(d\mathbf{n}_p) = \kappa_1 \kappa_2$. □

Since the Gauss map $\mathbf{n}$ does not depend on the parametrization $\mathbf{x}(u, v)$ of the surface, we see that the Gaussian curvature also does not depend on parametrization.

Another geometric quantity that comes out naturally from $d\mathbf{n}_p$ is the mean curvature.

Definition 3.4.11 (Mean curvature)**.** *Let S be a regular surface and $d\mathbf{n}_p$ be the differential of Gauss map at $p \in S$. The* **mean curvature** *of S at p is*

$$H = -\frac{1}{2}\mathrm{tr}(d\mathbf{n}_p) = \frac{1}{2}(\kappa_1 + \kappa_2) = \frac{1}{2}\mathrm{tr}((II)(I^{-1})) = \frac{1}{2}\left(\frac{gE - 2fF + eG}{EG - F^2}\right).$$

where $\mathrm{tr}(d\mathbf{n}_p)$ is the trace (Definition 1.6.6) of $d\mathbf{n}_p$ and κ_1, κ_2 are the principal curvatures of S at p.

Note that if we reverse the direction of the unit vector $\mathbf{n}$, that is, reserving the order of the parameters u, v, there will be a change of sign of the mean curvature but the Gaussian curvature would remain unchanged. So the sign of mean curvature does not matter. A surface with mean curvature zero is called a minimal surface.

Definition 3.4.12 (Minimal surface)**.** *Let S be a regular surface in $\mathbb{R}^3$ and H be the mean curvature of S. We say that S is a* **minimal surface** *if $H = 0$ at every point of S.*

Minimal surfaces have a distinguish property which can be considered as a two dimensional analogue of the arc length minimizing property (Theorem 2.2.11) of straight lines. A straight line segment is a curve of minimum arc length among all curves with fixed end points. Similarly a minimal surface is a surface of minimal surface area among all surfaces with fixed boundary and hence its name.

Theorem 3.4.13. *Let S be a minimal surface with parametrization $\mathbf{x} : D \to \mathbb{R}^3$ such that $\mathbf{x}$ can be extended continuously to the boundary. Then S has the minimum surface area among all surfaces with the same boundary of S.*

Example 3.4.14. *Show that the catenoid parametrized by*

$$\mathbf{x}(\theta, v) = (\cosh v \cos\theta, \cosh v \sin\theta, v),\ 1 < \theta < 2\pi, v \in \mathbb{R},$$

is a minimal surface.

Proof. We have

$$\begin{aligned}
\mathbf{x}_\theta &= (-\cosh v \sin\theta, \cosh v \cos\theta, 0) \\
\mathbf{x}_v &= (\sinh v \cos\theta, \sinh v \sin\theta, 1) \\
\mathbf{x}_\theta \times \mathbf{x}_v &= (\cosh v \cos\theta, \cosh v \sin\theta, -\cosh v \sinh v) \\
\|\mathbf{x}_\theta \times \mathbf{x}_v\|^2 &= \cosh^2 v + \cosh^2 v \sinh^2 v = \cosh^2 v(1 + \sinh^2 v) = \cosh^4 v \\
\mathbf{n} &= (\text{sech} v \cos\theta, \text{sech} v \sin\theta, \tanh v) \\
\mathbf{x}_{\theta\theta} &= (-\cosh v \cos\theta, -\cosh v \sin\theta, 0) \\
\mathbf{x}_{\theta v} &= (-\sinh v \sin\theta, \sinh v \cos\theta, 0) \\
\mathbf{x}_{vv} &= (\cosh v \cos\theta, \cosh v \sin\theta, 0).
\end{aligned}$$

Then the first and second fundamental forms are

$$\begin{aligned}
I &= \begin{pmatrix} E & F \\ F & G \end{pmatrix} = \begin{pmatrix} \cosh^2 v & 0 \\ 0 & \cosh^2 v \end{pmatrix} \\
II &= \begin{pmatrix} e & f \\ f & g \end{pmatrix} = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}
\end{aligned}$$

Thus the mean curvature is

$$H = \frac{1}{2}\left(\frac{gE - 2fF + eG}{EG - F^2}\right) = \frac{1}{2}\left(\frac{\cosh^2 v - \cosh^2 v}{\cosh^4 v}\right) = 0.$$

Therefore the catenoid is a minimal surface. □

To understand the geometric meaning of Gaussian curvature and mean curvature, let's take a closer look at the principal curvatures. A natural way of studying the curvature of a surface is to examine the curvature of curves on the surface. Let S be a regular surface and $\mathbf{v} \in T_pS$ be a unit vector tangent to S at p. Suppose C is a curve lying on S passing through p and is tangent to $\mathbf{v}$ at p. In other words, the unit tangent vector $\mathbf{T}$ of C at p satisfies $\mathbf{T} = \mathbf{v}$. We would like to understand the curvature of S from the curvature of the curve C. It turns out that the curvature of C at p depends only on the unit tangent $\mathbf{T} \in T_pS$ and the angle between the unit normal $\mathbf{N}$ of C and the unit normal vector of $\mathbf{n}$ at p.

Theorem 3.4.15. *Let S be a regular surface and $p \in S$ be a point on S. Let C be a regular parametrized curve passing through p. Then we have*

$$\kappa \cos\phi = -\langle \mathbf{T}, d\mathbf{n}_p(\mathbf{T})\rangle$$

where $\mathbf{T}, \kappa$ are the unit tangent vector, signed curvature of C at p respectively, $d\mathbf{n}_p$ is the differential of Gauss map of S at p and ϕ is the angle between the unit normal vector $\mathbf{N}$ of C and the unit normal vector $\mathbf{n}$ of S at p. Furthermore if $\mathbf{T} = \alpha\mathbf{x}_u + \beta\mathbf{x}_v \in T_pS$, then we have

$$\kappa \cos\phi = (\alpha \quad \beta)\, II \begin{pmatrix} \alpha \\ \beta \end{pmatrix}$$

where II is the second fundamental form.

Proof. Let $\mathbf{x}(u, v)$ be a regular parametrization of S. Since C lies on S, C has an arc length parametrization $\mathbf{r}(s)$ such that $\mathbf{r}(s) = \mathbf{x}(u(s), v(s))$ for some functions $u(s), v(s)$ with $\mathbf{r}(0) = p$ and $\mathbf{r}'(0) = \mathbf{T}$. By chain rule in multivariable calculus, we have

$$\begin{aligned} \mathbf{r}'(s) &= \frac{d}{ds}\mathbf{x}(u(s), v(s)) \\ &= u'(s)\mathbf{x}_u + v'(s)\mathbf{x}_v \end{aligned}$$

and similarly

$$\begin{aligned} \mathbf{n}'(s) &= \frac{d}{ds}\mathbf{n}(s) \\ &= u'(s)\mathbf{n}_u + v'(s)\mathbf{n}_v \end{aligned}$$

Observe that $\cos\phi = \langle \mathbf{n}, \mathbf{N}\rangle$ and we have

$$\begin{aligned}
\kappa\cos\phi &= \kappa\langle \mathbf{n}, \mathbf{N}\rangle \\
&= \langle \kappa\mathbf{N}, \mathbf{n}\rangle \\
&= \langle \mathbf{T}'(0), \mathbf{n}\rangle \quad \text{(Theorem 2.3.6)} \\
&= -\langle \mathbf{T}(0), \mathbf{n}'(0)\rangle \quad \text{(Lemma 1.3.35)} \\
&= -\langle \mathbf{T}, u'(0)\mathbf{n}_u + v'(0)\mathbf{n}_v\rangle \\
&= -\langle \mathbf{T}, u'(0)d\mathbf{n}_p(\mathbf{x}_u) + v'(0)d\mathbf{n}_p(\mathbf{x}_u)\rangle \quad \text{(Definition 3.4.5)} \\
&= -\langle \mathbf{T}, d\mathbf{n}_p(u'(0)\mathbf{x}_u + v'(0)\mathbf{x}_u)\rangle \\
&= -\langle \mathbf{T}, d\mathbf{n}_p(\mathbf{r}'(0))\rangle \\
&= -\langle \mathbf{T}, d\mathbf{n}_p(\mathbf{T})\rangle.
\end{aligned}$$

Furthermore if $\mathbf{T} = \alpha\mathbf{x}_u + \beta\mathbf{x}_v \in T_pS$, then $u'(0) = \alpha$, $v'(0) = \beta$ since $\mathbf{T} = \mathbf{r}'(0) = u'(0)\mathbf{x}_u + v'(0)\mathbf{x}_v$ and we have

$$\begin{aligned}
\kappa\cos\phi &= -\langle \mathbf{T}, u'(0)\mathbf{n}_u + v'(0)\mathbf{n}_v\rangle \\
&= -\langle \alpha\mathbf{x}_u + \beta\mathbf{x}_v, \alpha\mathbf{n}_u + \beta\mathbf{n}_v\rangle \\
&= -\begin{pmatrix}\alpha & \beta\end{pmatrix}\begin{pmatrix}\langle \mathbf{x}_u, \mathbf{n}_u\rangle & \langle \mathbf{x}_u, \mathbf{n}_v\rangle \\ \langle \mathbf{x}_v, \mathbf{n}_u\rangle & \langle \mathbf{x}_v, \mathbf{n}_v\rangle\end{pmatrix}\begin{pmatrix}\alpha \\ \beta\end{pmatrix} \\
&= \begin{pmatrix}\alpha & \beta\end{pmatrix} II \begin{pmatrix}\alpha \\ \beta\end{pmatrix}
\end{aligned}$$

where II is the second fundamental form. □

In particular if $\phi = 0$, then the curvature of C depends only on the tangent direction $\mathbf{T}$ and is called the normal curvature of S along $\mathbf{T}$.

Definition 3.4.16 (Normal curvature)**.** *Let S be a regular surface and p be a point on S. Let $\mathbf{v} \in T_pS$ be a unit vector tangent to the surface S at p. The* **normal curvature** *of S at p along $\mathbf{v}$ is*

$$\kappa_n(\mathbf{v}) = \kappa\cos\phi = -\langle \mathbf{v}, d\mathbf{n}_p(\mathbf{v})\rangle$$

where κ is the curvature of a curve C which passes through p and has unit tangent vector equals to $\mathbf{v}$, and ϕ is the angle between the unit normal vectors $\mathbf{N}$ and $\mathbf{n}$ of C and S at p respectively.

To visualize the normal curvature of a surface S at p along unit vector $\mathbf{v} \in T_pS$, one may cut the surface using a plane which passes through p and tangent to $\mathbf{v}$ and the normal vector $\mathbf{n}$ of S. Then the cross section, that is the intersection of the plane and the surface S, is a curve with normal vector $\mathbf{n}$ or $-\mathbf{n}$ and has curvature equals $\pm\kappa_n(\mathbf{v})$, where κ_n is the normal curvature.

Note that if the choice of direction of $\mathbf{n}$ is reversed, the normal curvature would have a change in sign. So the sign of normal curvature is not important.

Theorem 3.4.17. *Let S be a regular surface and $p \in S$ be a point on S. Let $\boldsymbol{\xi}_1, \boldsymbol{\xi}_2$ be the principal directions which constitute an orthonormal basis for T_pS and κ_1, κ_2 be the associated principal curvatures at p respectively. Let $\mathbf{v} \in T_pS$ be a unit vector tangent to S at p with $\mathbf{v} = \cos\theta\boldsymbol{\xi}_1 + \sin\theta\boldsymbol{\xi}_2$ where θ is the angle between $\mathbf{v}$ and $\boldsymbol{\xi}_1$. Then the normal curvature of S at p along $\mathbf{v}$ is*

$$\kappa_n(\mathbf{v}) = \kappa_1 \cos^2\theta + \kappa_2 \sin^2\theta.$$

Proof. By Theorem 3.4.15, the normal curvature along $\mathbf{v}$ is

$$\begin{aligned}
\kappa_n(\mathbf{v}) &= -\langle \mathbf{v}, d\mathbf{n}_p(\mathbf{v})\rangle \\
&= -\langle \cos\theta\boldsymbol{\xi}_1 + \sin\theta\boldsymbol{\xi}_2, d\mathbf{n}_p(\cos\theta\boldsymbol{\xi}_1 + \sin\theta\boldsymbol{\xi}_2)\rangle \\
&= -\langle \cos\theta\boldsymbol{\xi}_1 + \sin\theta\boldsymbol{\xi}_2, -\kappa_1\cos\theta\boldsymbol{\xi}_1 - \kappa_2\sin\theta\boldsymbol{\xi}_2)\rangle \\
&= \kappa_1\cos^2\theta + \kappa_2\sin^2\theta.
\end{aligned}$$

□

A direct consequence of the above theorem is that the normal curvature attains its maximum and minimum along the two orthogonal principal directions.

Theorem 3.4.18. *Let S be a regular surface and $p \in S$. Let $\kappa_1 \leq \kappa_2$ be the principal curvatures of S at p which associate with two orthogonal principal directions. Then for any unit vector $\mathbf{v} \in T_pS$ tangent to S at p, the normal curvature $\kappa_n(\mathbf{v})$ along $\mathbf{v}$ satisfies*

$$\kappa_1 \leq \kappa_n(\mathbf{v}) \leq \kappa_2.$$

Let us summarize the properties of Gaussian curvature we have discussed in the following theorem.

Theorem 3.4.19. *Let S be a regular surface parametrized by $\mathbf{x}(u, v)$ and K be the Gaussian curvature of S.*

1.

$$K = \frac{\det(II)}{\det(I)}$$

where I and II are the first fundamental forms of S.

2.

$$\mathbf{n}_u \times \mathbf{n}_v = K\mathbf{x}_u \times \mathbf{x}_v$$

where $\mathbf{n}$ is the unit normal vector of S.

3.

$$K = \frac{d\sigma}{dA}$$

where A and σ are the signed area function on S and S^2 respectively.

4.

$$K = \kappa_1\kappa_2$$

where κ_1, κ_2 are the principal curvatures associated with two orthogonal principal directions.

3.5 Theorema egregium

One may find that the Gaussian curvature of a surface somehow describe the change of normal vector along the surface. When Gauss introduced the notion of Gaussian curvature, he noticed already that one does not need to use normal vector to calculate the curvature. Say it in another way, the Gaussian curvature depends only on the mensuration on the surface but not how the surface is put into $\mathbb{R}^3$. This property of Gaussian curvature is so important and elegant that Gauss named his result 'Theorema Egregium' which are Latin meaning remarkable theorem. The theorem laid the foundation and inspired the development of the theory of differential geometry. Before we state the theorem, we introduce the notion of isometry.

Let S_1 be a regular surface and $f : S_1 \to S_2$ be a differentiable bijective map from S_1 to another regular surface S_2. Then any regular parametrization $\mathbf{x}_1(u, v)$ of S_1 induces a parametrization of S_2 by $\mathbf{x}_2(u, v) = f \circ \mathbf{x}_1(u, v) = f(\mathbf{x}_1(u, v))$. Furthermore the first fundamental forms $I_1(u, v)$ and $I_2(u, v)$ on S_1 and S_2 with respect to $\mathbf{x}_1(u, v)$ and $\mathbf{x}_2(u, v)$ can both be considered as matrix valued functions of u, v. We say that $f : S_1 \to S_2$ is an isometry if $I_1(u, v) = I_2(u, v)$ for any u, v.

Definition 3.5.1 (Isometry)**.** *Let S_1 and S_2 be regular surfaces. Let $f : S_1 \to S_2$ be a differentiable bijective map from S_1 to S_2. We say that a map $f : S_1 \to S_2$ is an* **isometry** *if $I_1(u, v) = I_2(u, v)$ for any u, v, where $I_1(u, v)$ is the first fundamental form of S_1 and $I_2(u, v)$ is the first fundamental form of S_2 induced by I_1. We say that S_1 and S_2 are* **isometric** *if there exists an isometry between S_1 and S_2.*

Roughly speaking, two regular surfaces S_1 and S_2 are isometric if they have the same first fundamental form. Intuitively, it means that one may get S_2 from S_1 by bending S_1 without stretching it. In this case, the mensuration on S_1 and S_2 would be the same. A curve on S_1 would have the same arc length as its image in S_2. A region on S_1 would have the same surface area as its image in S_2 and two curves on S_1 would intersect at the same angle as their image in S_2. Gauss' groundbreaking result asserts that two isometric surfaces must have identical Gaussian curvature.

Theorem 3.5.2 (Theorema egregium)**.** *Let S_1 and S_2 be two regular surfaces. Suppose S_1 and S_2 are isometric, that is, there exists isometry $f : S_1 \to S_2$ between S_1 and S_2. Then for any $p \in S_1$, the Gaussian curvature of S_1 at p is equal to the Gaussian curvature of S_2 at $f(p)$. In other words,*

$$K(f(p)) = K(p)$$

for any $p \in S_1$.

Proof. The proof of the theorem is complete if one can find a formula for Gaussian curvature which involves only first fundamental form but not second fundamental form. We will provide such a formula (Theorem 3.5.4) and give a proof of it at the end of this section. □

For example, one can get a cylindrical or conical surface by rolling up a plane which has Gaussian curvature zero everywhere. The theorem then implies that the Gaussian curvature of a cylindrical or conical surface must also be identically zero because the Gaussian curvature of a plane is zero. Another consequence of the theorem is that one cannot bend a plane into a spherical surface without stretching the plane because a spherical surface has nonzero Gaussian curvature. Thus it is impossible to draw a map for the earth surface with uniform scale. The following example is less obvious.

Example 3.5.3 (Isometry between catenoid and helicoid)**.** *Let S_1 be the catenoid which is parametrized by*

$$\mathbf{x}_1(\theta, v) = (\cosh v \cos\theta, \cosh v \sin\theta, v), \ \ (\theta, v) \in (0, 2\pi) \times \mathbb{R}$$

and S_2 be the helicoid which is parametrized by

$$\mathbf{x}_2(\theta, v) = (\sinh v \cos\theta, \sinh v \sin\theta, \theta), \ \ (\theta, v) \in (0, 2\pi) \times \mathbb{R}.$$

The first fundamental forms of them are the same and is equal to

$$I_1(\theta, v) = I_2(\theta, v) = \begin{pmatrix} \cosh^2 v & 0 \\ 0 & \cosh^2 v \end{pmatrix}$$

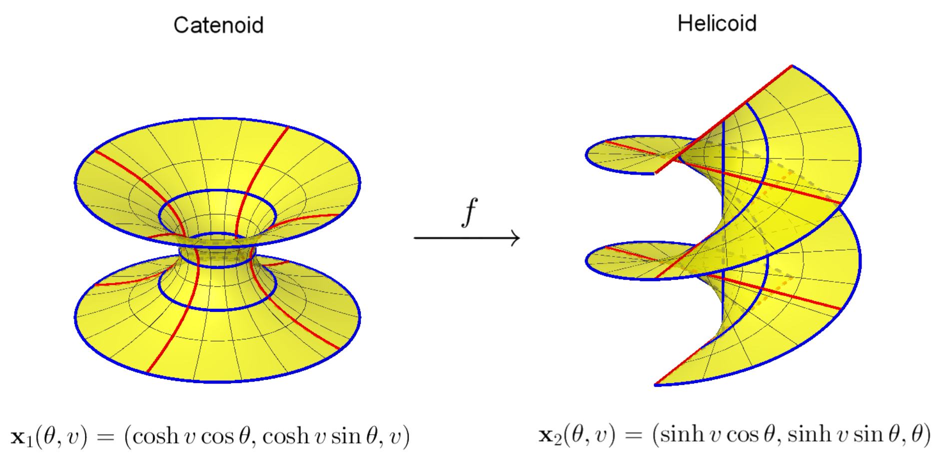

Figure 17: Isometry between catenoid and helicoid

By theorema egregium (Theorem 3.5.2), the two surfaces have the identical Gaussian curvature (See Example 3.3.8 and Example 3.3.12) which is equal to

$$K = -\frac{1}{\cosh^4 v}.$$

Catenoid and Helicoid are both minimal surface and thus have mean curvature identically zero. However, the mean curvature of two isometric surfaces may not be identical. For example, a cylindrical surface and a plane are isometric but a cylindrical surface has nonzero mean curvature while that of a plane is zero.

We conclude this section by providing a formula for Gaussian curvature which involve only first fundamental form but not second fundamental form as promised.

Theorem 3.5.4. *Let* $\mathbf{x}(u, v)$ *be a regular parametrized surface. Then*

$$K = \frac{1}{4(EG - F^2)^2}\left(\begin{vmatrix} -E_{vv} + 2F_{uv} - G_{uu} & E_u & 2F_u - E_v \\ 2F_v - G_u & E & F \\ G_v & F & G \end{vmatrix} - \begin{vmatrix} 0 & E_v & G_u \\ E_v & E & F \\ G_u & F & G \end{vmatrix} \right).$$

In particular, if $F = 0$ *is identically zero, then*

$$K = -\frac{1}{2\sqrt{EG}}\left[\left(\frac{E_v}{\sqrt{EG}} \right)_v + \left(\frac{G_u}{\sqrt{EG}} \right)_u \right].$$

Proof. Since $\det(I) = EG - F^2 = \|\mathbf{x}_u \times \mathbf{x}_v\|^2$ and $\|\mathbf{x}_u \times \mathbf{x}_v\|\mathbf{n} = \mathbf{x}_u \times \mathbf{x}_v$,

$$\begin{aligned}
& K(EG - F^2)^2 \\
= \ & \det(I)\det(II) \\
= \ & \|\mathbf{x}_u \times \mathbf{x}_v\|^2 \begin{vmatrix} \langle \mathbf{x}_{uu}, \mathbf{n} \rangle & \langle \mathbf{x}_{uv}, \mathbf{n} \rangle \\ \langle \mathbf{x}_{vu}, \mathbf{n} \rangle & \langle \mathbf{x}_{vv}, \mathbf{n} \rangle \end{vmatrix} \\
= \ & \begin{vmatrix} \langle \mathbf{x}_{uu}, \|\mathbf{x}_u \times \mathbf{x}_v\|\mathbf{n} \rangle & \langle \mathbf{x}_{uv}, \|\mathbf{x}_u \times \mathbf{x}_v\|\mathbf{n} \rangle \\ \langle \mathbf{x}_{vu}, \|\mathbf{x}_u \times \mathbf{x}_v\|\mathbf{n} \rangle & \langle \mathbf{x}_{vv}, \|\mathbf{x}_u \times \mathbf{x}_v\|\mathbf{n} \rangle \end{vmatrix} \\
= \ & \begin{vmatrix} \langle \mathbf{x}_{uu}, \mathbf{x}_u \times \mathbf{x}_v \rangle & \langle \mathbf{x}_{uv}, \mathbf{x}_u \times \mathbf{x}_v \rangle \\ \langle \mathbf{x}_{vu}, \mathbf{x}_u \times \mathbf{x}_v \rangle & \langle \mathbf{x}_{vv}, \mathbf{x}_u \times \mathbf{x}_v \rangle \end{vmatrix} \\
= \ & \begin{vmatrix} \langle \mathbf{x}_{uu}, \mathbf{x}_{vv} \rangle - \langle \mathbf{x}_{uv}, \mathbf{x}_{uv} \rangle & \langle \mathbf{x}_{uu}, \mathbf{x}_u \rangle & \langle \mathbf{x}_{uu}, \mathbf{x}_v \rangle \\ \langle \mathbf{x}_{vv}, \mathbf{x}_u \rangle & \langle \mathbf{x}_u, \mathbf{x}_u \rangle & \langle \mathbf{x}_u, \mathbf{x}_v \rangle \\ \langle \mathbf{x}_{vv}, \mathbf{x}_v \rangle & \langle \mathbf{x}_v, \mathbf{x}_u \rangle & \langle \mathbf{x}_v, \mathbf{x}_v \rangle \end{vmatrix} \\
& - \begin{vmatrix} 0 & \langle \mathbf{x}_{uv}, \mathbf{x}_u \rangle & \langle \mathbf{x}_{uv}, \mathbf{x}_v \rangle \\ \langle \mathbf{x}_{vu}, \mathbf{x}_u \rangle & \langle \mathbf{x}_u, \mathbf{x}_u \rangle & \langle \mathbf{x}_u, \mathbf{x}_v \rangle \\ \langle \mathbf{x}_{vu}, \mathbf{x}_v \rangle & \langle \mathbf{x}_v, \mathbf{x}_u \rangle & \langle \mathbf{x}_v, \mathbf{x}_v \rangle \end{vmatrix}. \text{ (Proposition 1.3.17)}
\end{aligned}$$

Observe that by product rule (Proposition 1.3.34),

$$\begin{aligned}\begin{pmatrix} E_u & F_u \\ F_u & G_u \end{pmatrix} &= \frac{\partial}{\partial u}\begin{pmatrix} E & F \\ F & G \end{pmatrix} \\ &= \frac{\partial}{\partial u}\begin{pmatrix} \langle \mathbf{x}_u, \mathbf{x}_u\rangle & \langle \mathbf{x}_u, \mathbf{x}_v\rangle \\ \langle \mathbf{x}_v, \mathbf{x}_u\rangle & \langle \mathbf{x}_v, \mathbf{x}_v\rangle \end{pmatrix} \\ &= \begin{pmatrix} \langle \mathbf{x}_{uu}, \mathbf{x}_u\rangle & \langle \mathbf{x}_{uu}, \mathbf{x}_v\rangle \\ \langle \mathbf{x}_{vu}, \mathbf{x}_u\rangle & \langle \mathbf{x}_{vu}, \mathbf{x}_v\rangle \end{pmatrix} + \begin{pmatrix} \langle \mathbf{x}_u, \mathbf{x}_{uu}\rangle & \langle \mathbf{x}_u, \mathbf{x}_{vu}\rangle \\ \langle \mathbf{x}_v, \mathbf{x}_{uu}\rangle & \langle \mathbf{x}_v, \mathbf{x}_{vu}\rangle \end{pmatrix} \\ &= \begin{pmatrix} 2\langle \mathbf{x}_{uu}, \mathbf{x}_u\rangle & \langle \mathbf{x}_{uu}, \mathbf{x}_v\rangle + \langle \mathbf{x}_{uv}, \mathbf{x}_u\rangle \\ \langle \mathbf{x}_{uu}, \mathbf{x}_v\rangle + \langle \mathbf{x}_{uv}, \mathbf{x}_u\rangle & 2\langle \mathbf{x}_{vu}, \mathbf{x}_v\rangle \end{pmatrix}.\end{aligned}$$

Similarly

$$\begin{pmatrix} E_v & F_v \\ F_v & G_v \end{pmatrix} = \begin{pmatrix} 2\langle \mathbf{x}_{uv}, \mathbf{x}_u\rangle & \langle \mathbf{x}_{vv}, \mathbf{x}_u\rangle + \langle \mathbf{x}_{uv}, \mathbf{x}_v\rangle \\ \langle \mathbf{x}_{vv}, \mathbf{x}_u\rangle + \langle \mathbf{x}_{uv}, \mathbf{x}_v\rangle & 2\langle \mathbf{x}_{vv}, \mathbf{x}_v\rangle \end{pmatrix}.$$

Combining the above two equalities, we obtain

$$\begin{cases} \langle \mathbf{x}_{uu}, \mathbf{x}_u\rangle = \dfrac{E_u}{2}, \\ \langle \mathbf{x}_{uv}, \mathbf{x}_u\rangle = \dfrac{E_v}{2}, \\ \langle \mathbf{x}_{uv}, \mathbf{x}_v\rangle = \dfrac{G_u}{2}, \\ \langle \mathbf{x}_{vv}, \mathbf{x}_v\rangle = \dfrac{G_v}{2}, \\ \langle \mathbf{x}_{uu}, \mathbf{x}_v\rangle = F_u - \langle \mathbf{x}_{uv}, \mathbf{x}_u\rangle = F_u - \dfrac{E_v}{2}, \\ \langle \mathbf{x}_{vv}, \mathbf{x}_u\rangle = F_v - \langle \mathbf{x}_{uv}, \mathbf{x}_v\rangle = F_v - \dfrac{G_u}{2}. \end{cases}$$

Moreover by considering the second derivative of $F = \langle \mathbf{x}_u, \mathbf{x}_v\rangle$ with respect to u, v, we have

$$\begin{aligned}\langle \mathbf{x}_{uu}, \mathbf{x}_{vv}\rangle &= \frac{\partial}{\partial u}\langle \mathbf{x}_{vv}, \mathbf{x}_u\rangle - \langle \mathbf{x}_{vvu}, \mathbf{x}_u\rangle \\ &= \frac{\partial}{\partial u}\left(F_v - \frac{G_u}{2}\right) - \left(\frac{\partial}{\partial v}\langle \mathbf{x}_{uv}, \mathbf{x}_u\rangle - \langle \mathbf{x}_{uv}, \mathbf{x}_{uv}\rangle\right) \\ &= F_{uv} - \frac{G_{uu}}{2} - \left(\frac{\partial}{\partial v}\frac{E_v}{2} - \langle \mathbf{x}_{uv}, \mathbf{x}_{uv}\rangle\right) \\ &= -\frac{E_{vv}}{2} + F_{uv} - \frac{G_{uu}}{2} + \langle \mathbf{x}_{uv}, \mathbf{x}_{uv}\rangle\end{aligned}$$

which implies

$$\langle \mathbf{x}_{uu}, \mathbf{x}_{vv} \rangle - \langle \mathbf{x}_{uv}, \mathbf{x}_{uv} \rangle = -\frac{E_{vv}}{2} + F_{uv} - \frac{G_{uu}}{2}.$$

Therefore

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

as desire. If particular, if $F = 0$, then

$$\begin{aligned} K &= \frac{1}{4E^2G^2} \left(\begin{vmatrix} -E_{vv} - G_{uu} & E_u & -E_v \\ -G_u & E & 0 \\ G_v & 0 & G \end{vmatrix} - \begin{vmatrix} 0 & E_v & G_u \\ E_v & E & 0 \\ G_u & 0 & G \end{vmatrix} \right) \\ &= \frac{1}{4E^2G^2} \left(-EGE_{vv} - EGG_{uu} + GE_uG_u + EE_vG_v + GE_v^2 + EG_u^2 \right) \\ &= -\frac{E_{vv}}{4EG} - \frac{G_{uu}}{4EG} + \frac{E_uG_u}{4E^2G} + \frac{E_vG_v}{4EG^2} + \frac{E_v^2}{4E^2G} + \frac{G_u^2}{4EG^2}. \end{aligned}$$

Observe that

$$\begin{cases} \left(\frac{E_v}{\sqrt{EG}} \right)_v = \frac{E_{vv}}{\sqrt{EG}} - \frac{E_v^2}{2E\sqrt{EG}} - \frac{E_vG_v}{2G\sqrt{EG}} \\ \left(\frac{G_u}{\sqrt{EG}} \right)_u = \frac{G_{uu}}{\sqrt{EG}} - \frac{G_u^2}{2G\sqrt{EG}} - \frac{E_uG_u}{2E\sqrt{EG}}. \end{cases}$$

Hence

$$\begin{aligned} & \left(\frac{E_v}{\sqrt{EG}} \right)_v + \left(\frac{G_u}{\sqrt{EG}} \right)_u \\ = & \frac{E_{vv}}{\sqrt{EG}} - \frac{E_v^2}{2E\sqrt{EG}} - \frac{E_vG_v}{2G\sqrt{EG}} + \frac{G_{uu}}{\sqrt{EG}} - \frac{G_u^2}{2G\sqrt{EG}} - \frac{E_uG_u}{2E\sqrt{EG}} \\ = & -2\sqrt{EG} \left(-\frac{E_{vv}}{4EG} + \frac{E_v^2}{4E^2G} + \frac{E_vG_v}{4EG^2} - \frac{G_{uu}}{4EG} + \frac{G_u^2}{4EG^2} + \frac{E_uG_u}{4E^2G} \right) \\ = & -2K\sqrt{EG} \end{aligned}$$

and the result follows. □

Note that for any regular surface, there is always a parametrization with $F = 0$ identically zero.

One may ask whether the converse of theorema egregium holds. The answer is negative. For example, it is known that there exists constant Gaussian curvature surface which is not isometric to any part of a sphere.

3.6 Gauss-Bonnet theorem

In this section, we explain the Gauss-Bonnet theorem. The theorem is important because it relates a local quantity, the Gaussian curvature, with a global quantity, the Euler characteristic, of a surface. It can also be interpreted as an analogue of Theorem 2.3.16 for surfaces. To state the theorem, we introduce the notion of Euler characteristic. For a closed surface S, a polyhedron modeled on S is a polyhedron whose vertices, edges, faces are points, curves, regions on the surface S.

Definition 3.6.1 (Euler characteristic)**.** *The* **Euler characteristic** *of a closed surface S is*

$$\chi(S) = v - e + f$$

where v, e and f are the number of vertices, edges and faces of a polyhedron modeled on S.

Given a closed surface S, one can find many different polyhedrons modeled on S but it can be proved that $\chi(S)$ does not depend on the choice of models. Two surfaces have the same Euler characteristic if one can deform the surface to another without stretching. Thus Euler characteristic is a topological invariant[11]. For example, a sphere S^2 has Euler characteristic $\chi(S^2) = 2$ which means any polyhedron modeled on S^2 would have $v - e + f = 2$. Before we prove this fact, we derive a formula for area of polygons on the unit sphere.

Theorem 3.6.2 (Area of polygon on unit sphere)**.** *Let α, β, γ be the interior angles of a triangle, with edges being great circular arcs[12], on the unit sphere and A be the area of the triangle. Then*

$$\alpha + \beta + \gamma = A + \pi.$$

[11] More precisely if S_1 and S_2 are homeomorphic, which means there exists bijective map $f : S_1 \to S_2$ such that both f and f^{-1} are continuous, then $\chi(S_1) = \chi(S_2)$.

[12] A great circle on the unit sphere is a circle on the sphere with radius 1.

More generally, Let $\alpha_1, \alpha_2, \ldots, \alpha_n$ be the interior angles of a polygon with n edges, which are great circular arcs, on the unit sphere and A be the area of the polygon. Then

$$\alpha_1 + \alpha_2 + \cdots + \alpha_n = A + (n-2)\pi.$$

Proof. The second statement follows readily from the first by a standard argument of cutting the polygon into $n-2$ triangles. To prove the first statement, consider a region on the unit sphere which is bounded by 2 great semicircles with both interior angles equal α. The region occupied $\alpha/2\pi$ of the surface of the sphere and thus has an area of

$$\frac{\alpha}{2\pi} \times 4\pi = 2\alpha.$$

We will call such a region a **biangle** with interior angle α.

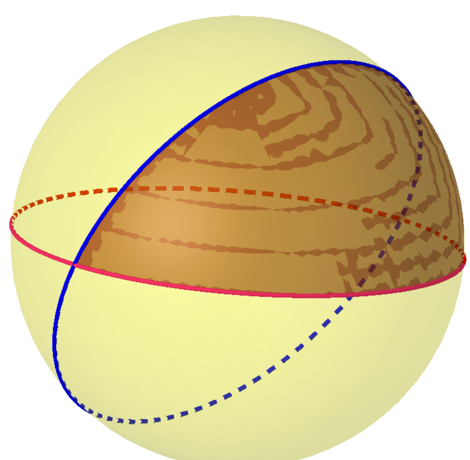

Figure 18: Biangle on sphere

Now extend the 3 edges of the triangle to great circles on the sphere which cut the sphere into 8 regions. One may use 2 of the 8 regions to form a biangle with interior α and two such biangles can be obtained. Similarly, we get two biangles with interior angle β and two biangles with interior angle γ. All 6 biangles obtained in this way cover (See Figure 19) the unit sphere with 4 extra triangles with interior angles α, β, γ.

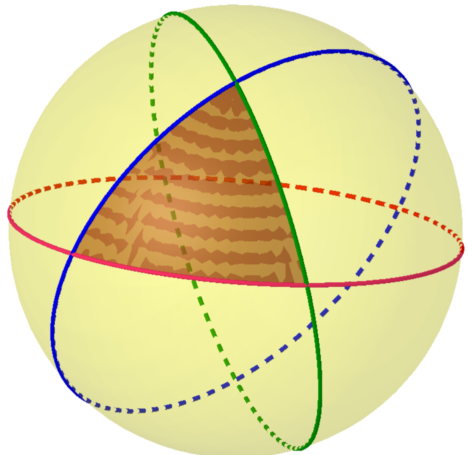

Figure 19: Triangle on sphere

By considering the total area of them, we get

$$2 \times 2\alpha + 2 \times 2\beta + 2 \times 2\gamma = 4\pi + 4A$$

where A is the area of the triangle with interior angles α, β, γ which implies

$$\alpha + \beta + \gamma = \pi + A.$$

□

Now we prove that the Euler characteristic of a sphere is 2.

Theorem 3.6.3 (Euler characteristic of sphere)**.** *A polyhedron which is modeled on a sphere has Euler characteristic* $\chi = 2$*.*

Proof. Consider a polyhedron modeled on the unit sphere. By deforming the edges, we may assume that the edges are great circular arcs on the unit sphere. Let v, e and f be the number of vertices, edges and faces of the polyhedron. Suppose the k-th face, $k = 1, 2, \ldots, f$, is a polygon with e_k edges, e_k interior angles $\alpha_{k_1}, \alpha_{k_2}, \ldots, \alpha_{k_{e_k}}$ and has area equal to A_k. By Theorem 3.6.2, we have

$$\sum_{i=1}^{e_k} \alpha_{k_i} = (e_k - 2)\pi + A_k.$$

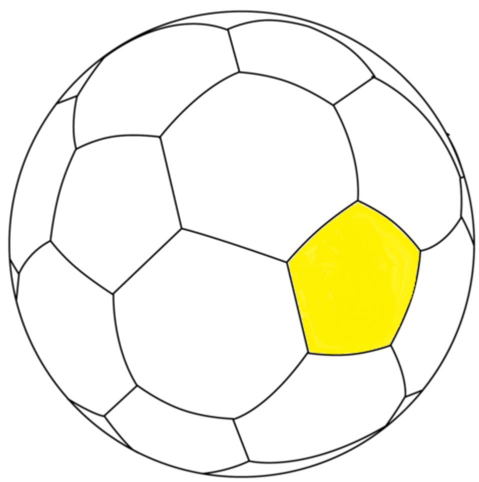

Figure 20: Polygon on sphere

Summing up the above equalities for $k = 1, 2, \ldots, f$, we have

$$\sum_{k=1}^{f} \sum_{i=1}^{e_k} \alpha_{k_i} = \sum_{k=1}^{f} e_k \pi - 2 \sum_{k=1}^{f} \pi + \sum_{k=1}^{f} A_k.$$

Now the sum of all interior angles of all faces is equal to 2π times the number of vertices v which gives

$$\sum_{k=1}^{f} \sum_{i=1}^{e_k} \alpha_{k_i} = 2\pi v.$$

The sum of all e_k, $k = 1, 2, \ldots, f$, is equal to 2 times the total number of edges e of the polyhedron and we obtain

$$\sum_{k=1}^{f} e_k \pi = 2\pi e.$$

Furthermore, the sum of the area of all faces is equal to the area of the unit sphere and we have

$$\sum_{k=1}^{f} A_k = 4\pi.$$

Combining the above equalities, we have

$$\begin{aligned} 2\pi v &= 2\pi e - 2\pi f + 4\pi \\ v - e + f &= 2 \end{aligned}$$

□

By the classification theorem of closed surfaces, simple closed surfaces in $\mathbb{R}^3$ are completely classified by its **genus** g. Intuitively, speaking the genus of a closed surface is the number of 'holes' of the surface. For example, a sphere has genus 0, a torus has genus 1 and one obtains a surface of genus g by gluing g tori together. The Euler characteristic of a closed surface in $\mathbb{R}$ can be determined by its genus.

Theorem 3.6.4 (Euler characteristic of simple closed surface)**.** *Let S be a simple closed surface of genus g. Then the Euler characteristic of S is*

$$\chi(S) = 2 - 2g.$$

Proof. We have proved that the sphere S^2, which has genus 0, has Euler characteristic $\chi(S^2) = 2$ (Theorem 3.6.3). Now we calculate the Euler characteristic of a torus T which has genus 1. One may construct a polyhedron modeled on a torus with 9 vertices, 18 edges and 9 faces. Therefore

$$\chi(T) = 9 - 18 + 9 = 0.$$

Next we observe the change in Euler characteristics when gluing two surfaces. Let S_1 and S_2 be two closed surfaces. We may remove a circular region from each surface and glue the two surfaces together along the boundaries of the two regions. We denote the new surface obtained in this way by $S_1\#S_2$ and call it the connected sum of S_1 and S_2. Let v_1, e_1, f_1 and v_2, e_2, f_2 be the number of vertices, edges, faces of polyhedrons modeled on S_1 and S_2 respectively. One may find such polyhedrons so that the regions removed on the two surfaces are polygons with k edges. When gluing the two surfaces, k vertices, k edges and 2 surfaces have been removed. Thus the resulting polyhedron modeled on $S_1\#S_2$ has v_1+v_2-k vertices, e_1+e_2-k edges and f_1+f_2-2 faces. Hence the Euler characteristic of $S_1\#S_2$ is

$$\begin{aligned}\chi(S_1\#S_2) &= (v_1+v_2-k)-(e_1+e_2-k)+(f_1+f_2-2)\\ &= v_1-e_1+f_1+v_2-e_2+f_2-2\\ &= \chi(S_1)+\chi(S_2)-2.\end{aligned}$$

Now a closed surface S_g in $\mathbb{R}^3$ of genus g is obtained by gluing $g-1$ tori to a torus. Every time we glue one torus to a surface, the Euler characteristic is decreased by 2. Therefore the Euler characteristic os S_g is

$$\chi(S_g) = 0 - 2(g-1) = 2-2g.$$

□

To prove the Gauss-Bonnet theorem, we introduce one more definition. Let S_1 and S_2 be two simple closed surface in $\mathbb{R}^3$. Let $f: S_1 \to S_2$ be a continuous map from S_1 to S_2. For $q \in S_2$, we define the degree of f at q to be the integer

$$\deg(f,q) = \begin{array}{l}\text{number of preimages of } q \text{ preserving orientation}\\ -\text{number of preimages of } q \text{ reversing orientation}\end{array}.$$

It can be proved that this integer are the same for almost all points $q \in S_2$. We call it the **degree** of f and denote it by $\deg(f)$. Intuitively, if the degree of $f: S_1 \to S_2$ is k, the first surface covers the second surface k times via f. Now let S be a simple closed regular surface in $\mathbb{R}^3$ and $\mathbf{n}: S \to S^2$ be its Gauss map. To calculate the degree of Gauss map, it is useful to note that for any $p \in S$, the Gauss map $\mathbf{n}$ is orientation preserving at p if the Gaussian curvature at p is positive and is orientation reversing at p if the Gaussian curvature at p is negative. Thus one needs to find the number of points with positive and negative Gaussian curvature with a given normal direction. It turns out that the degree $\deg(\mathbf{n})$ of the Gauss map depends only on the genus g of S. To see this, let S_g be another simple closed surface of genus g. One may always deform S continuously to obtain S_g and the degree of Gauss map in the process would remain constant. This is because the degree changes continuously when one deforms the surface continuously and degree takes only integer values. This implies that the degree of Gauss map must be constant when the surface is being deformed. The degree of Gauss map depends on the genus g in the following way.

Theorem 3.6.5 (Degree of Gauss map of simple closed regular surface)**.** *Let S be a simple closed surface of genus g. The the degree of Gauss map of S is*

$$\deg(\mathbf{n}) = 1 - g.$$

Proof. It is not difficult to see that there exists a surface S_g of genus g such that there are exactly $g+1$ points on S_g with unit normal vector $(0,0,1)$, where g of them are orientation reversing, that is, having negative Gaussian curvature, and the remaining 1 of them is orientation preserving, that is, having positive Gaussian curvature. Now the degree of Gauss map of S is equal to that of S_g which is equal to $1-g$. □

We are ready to state and prove the Gauss-Bonnet theorem.

Theorem 3.6.6 (Gauss-Bonnet theorem)**.** *Let S be a simple closed regular surface in $\mathbb{R}^3$. Then*

$$\iint_S K dA = 2\pi\chi(S)$$

where K is the Gaussian curvature, $\chi(S)$ is the Euler characteristic of S and $dA = \sqrt{\det(I)}dudv$ is the surface area element. In particular, if S is homeomorphic[13] *to the sphere S^2, then $\chi(S) = 2$ and*

$$\iint_S K dA = 4\pi.$$

[13]That means there exists a bijective map $f : S \to S^2$ from S to the sphere S^2 such that both f and f^{-1} are continuous.

Proof. We have

$$\begin{aligned}
\iint_S K dA &= \iint_S \frac{d\sigma}{dA} dA \quad \text{(Proposition 3.4.4)} \\
&= \iint_S d\sigma \\
&= \deg(\mathbf{n}) \iint_{S^2} d\sigma \\
&= (1-g)(4\pi) \quad \text{(Theorem 3.6.5)} \\
&= 2\pi(2-2g) \\
&= 2\pi\chi(S) \quad \text{(Theorem 3.6.4)}
\end{aligned}$$

□

Exercise 3

1. Prove that a regular parametrized surface in $\mathbb{R}^3$ is contained in a plane if and only if the unit normal vector $\mathbf{n}$ is constant.

2. Prove that a regular parametrized surface in $\mathbb{R}^3$ is part of a sphere if and only if all normal vectors pass through a fixed point.

3. Find the first fundamental form and the surface area of the following parametrized surface.

   (a) $\mathbf{x}(u,\theta) = (u\cos\theta, u\sin\theta, u^2)$, $u \in (0,1)$, $\theta \in (0, 2\pi)$.

   (b) $\mathbf{x}(u,\theta) = (u^3\cos\theta, u^3\sin\theta, u)$, $u \in (0,1)$, $\theta \in (0, 2\pi)$.

   (c) $\mathbf{x}(u,\theta) = (u\cos\theta, u\sin\theta, \theta)$, $u \in (-1,1)$, $\theta \in (0, 2\pi)$.
   (You may use $\int \sqrt{x^2+1}\, dx = \frac{1}{2}(x\sqrt{x^2+1} + \ln(x+\sqrt{x^2+1})) + C$ directly.)

4. Prove that the area of the surface define by $z = f(r,\theta)$, $(r,\theta) \in D$, where (r,θ) is the polar coordinates on the xy-plane such that $(x,y) = (r\cos\theta, r\sin\theta)$, is given by

$$\iint_D \sqrt{r^2 + r^2 f_r^2 + f_\theta^2}\, dr d\theta$$

5. Let $\mathbf{x}(u,v)$ be a regular parametrized surface. Let $\mathbf{r}(t) = \mathbf{x}(u(t), v(t))$, $a < t < b$, be a curve lying on the surface. Prove that the arc length of $\mathbf{r}(t)$ is

$$\int_a^b \sqrt{\begin{pmatrix} \dot{u} & \dot{v} \end{pmatrix} I \begin{pmatrix} \dot{u} \\ \dot{v} \end{pmatrix}}\, dt$$

where I is the first fundamental form of $\mathbf{x}(u,v)$.

6. Find the second fundamental form and the Gaussian curvature of the following parametrized surface.

   (a) $\mathbf{x}(u,v) = (u^2 - v^2, 2uv, u^2 + v^2)$, $u \in \mathbb{R}$, $v > 0$.

   (b) (Enneper surface) $\mathbf{x}(u,v) = (u - \frac{u^3}{3} + uv^2, -v + \frac{v^3}{3} - u^2 v, u^2 - v^2)$, $u, v \in \mathbb{R}$.

   (c) (Torus) $\mathbf{x}(\phi,\theta) = ((R + r\sin\phi)\cos\theta, (R + r\sin\phi)\sin\theta, r\cos\phi)$, $\phi, \theta \in (0, 2\pi)$, where R, r are constants.

7. Prove that the Gaussian curvature of the surface defined by $z = f(x,y)$ is

$$K(x,y) = \frac{f_{xx} f_{yy} - f_{xy}^2}{(1 + f_x^2 + f_y^2)^2}$$

8. Prove that the Gaussian curvature of the surface defined by $z = f(r,\theta)$ in the cylindrical coordinates, where (r,θ) is the polar coordinates in the xy-plane such that $(x,y) = (r\cos\theta, r\sin\theta)$, is given by

$$K(r,\theta) = \frac{r^2 f_{rr}(r f_r + f_{\theta\theta}) - (r f_{r\theta} - f_\theta)^2}{(r^2 + r^2 f_r^2 + f_\theta^2)^2}$$

9. Let $\mathbf{r}(s)$ be an arc-length parametrized space curve. The tangent developable surface of $\mathbf{r}$ is the surface parametrized by $\mathbf{x}(s,t) = \mathbf{r}(s) + t\mathbf{T}(s)$ where $\mathbf{T}(s)$ is the unit tangent vector. Prove that the Gaussian curvature of a tangent developable surface is always zero.

10. Let $\mathbf{r}(t) = (x(t), y(t))$ be a regular parametrized curve on the xy-plane. The conical surface spanned by the curve $\mathbf{r}(t)$ is the surface parametrized by $\mathbf{x}(u,v) = (vx(u), vy(u), v)$, $v \in (0, +\infty)$. Prove that the Gaussian curvature of the conical surface is 0.

11. Let $f(u)$ be a second differentiable function and $\mathbf{x}(u,\theta) = (u\cos\theta, u\sin\theta, f(u))$, $u > 0$, $\theta \in (0, 2\pi)$, be a parametrized surface.

    (a) Prove that the Gaussian curvature of the surface is

    $$K(u) = \frac{f'f''}{u(1+f'^2)^2}$$

    (b) Prove that

    $$-\frac{1}{2u}\frac{d}{du}\left(\frac{1}{1+f'^2}\right) = K(u)$$

    (c) Suppose $f(u)$ is a function such that

    $$f'(u) = \sqrt{\frac{9+u^2}{16-u^2}}$$

    Prove that K is a constant and find the constant.

    (This exercise shows that a surface with constant positive Gaussian curvature may not necessarily be a sphere.)

12. Find the Gaussian curvature of the parametrized surface $\mathbf{x}(u,v)$ with the following first fundamental form.

    (a) $I = \begin{pmatrix} \frac{1}{u^2} & 0 \\ 0 & \frac{1}{u^2} \end{pmatrix}$

    (b) $I = \begin{pmatrix} \frac{1}{u^2+v^2+1} & 0 \\ 0 & \frac{1}{u^2+v^2+1} \end{pmatrix}$

    (c) $I = \begin{pmatrix} 1 & 0 \\ 0 & \cosh^2 u \end{pmatrix}$

13. Suppose the first fundamental form of a parametrized surface $\mathbf{x}(u,v)$ is

    $$I = \begin{pmatrix} f^2 & 0 \\ 0 & f^2 \end{pmatrix}$$

    where $f = f(u,v) > 0$ is a second differentiable function. Show that the Gaussian curvature of the surface is

    $$K = -\frac{1}{f^2}\left(\frac{\partial^2}{\partial u^2} + \frac{\partial^2}{\partial v^2}\right)\ln f$$

14. Find the mean curvature of the following parametrized surface.

    (a) $\mathbf{x}(u,v) = (u, v, uv)$, $u, v \in \mathbb{R}$.

    (b) (Torus) $\mathbf{x}(\phi,\theta) = ((R + r\sin\phi)\cos\theta, (R + r\sin\phi)\sin\theta, r\cos\phi)$, $\phi, \theta \in (0, 2\pi)$, where $R > r > 0$ are constants.

    (c) (Helicoid) $\mathbf{x}(u,\theta) = (au\cos\theta, au\sin\theta, b\theta)$, $u, \theta \in \mathbb{R}$ where $a, b > 0$ are constants.

15. Prove that the surface defined by $z = f(x,y)$ is a minimal surface if and only if

$$(1 + f_x^2)f_{yy} - 2f_x f_y f_{xy} + (1 + f_y^2)f_{xx} = 0$$

16. Let $\mathbf{x}(u,v)$ be a regular parametrized surface. Let $\mathbf{r}(s) = \mathbf{x}(u(s), v(s))$ be a curve lying on the surface parametrized by arc length. Prove that

$$\kappa\langle \mathbf{N}(s), \mathbf{n}(s)\rangle = \begin{pmatrix} \dot{u} & \dot{v} \end{pmatrix} II \begin{pmatrix} \dot{u} \\ \dot{v} \end{pmatrix}$$

    where κ is the curvature of $\mathbf{r}(s)$, $\mathbf{N}$ is the unit normal vector to the curve, $\mathbf{n}$ is the unit normal vector to the surface, II is the second fundamental form of the surface, $\dot{u}$ and $\dot{v}$ are the derivatives of u and v with respect to s respectively.

17. Consider the surface obtained by rotating the curve on the xz-plane defined by $x = f(z)$, $a < z < b$, along the z-axis.

    (a) Prove that the area of the surface is given by

$$2\pi \int_a^b f\sqrt{1 + f'^2}\, dz$$

    (b) Prove that the Gaussian curvature of the surface is

$$K = -\frac{f''}{f(1 + f'^2)^2}$$

    (c) Prove that the mean curvature of the surface is

$$H = \frac{ff'' - f'^2 - 1}{2f(1 + f'^2)^{\frac{3}{2}}}$$

18. For each of the surface S, calculate $\iint_S K dA$, where K is the Gaussian curvature and dA is the area element of S.

    (a) (Helicoid) $\mathbf{x}(u,\theta) = (u\cos\theta, u\sin\theta, \theta)$, $u \in (0,1)$, $\theta \in (0, 2\pi)$.

    (b) (Ellipsoid) $\mathbf{x}(\varphi,\theta) = (a\sin\varphi\cos\theta, a\sin\varphi\sin\theta, b\cos\varphi)$, $\varphi \in (0, \frac{\pi}{2})$, $\theta \in (0, 2\pi)$.
    (Hint: $\displaystyle\int \frac{\sin\varphi}{(a^2\cos^2\varphi + b^2\sin^2\varphi)^{\frac{3}{2}}} d\varphi = -\frac{\cos\varphi}{b^2\sqrt{a^2\cos^2\varphi + b^2\sin^2\varphi}} + C$.)

19. Consider the surface obtained by rotating the arc length parametrized curve $(x,z) = (\varphi(s), \psi(s))$, $s \in (0,l)$, $\varphi(s) > 0$, on the xz-plane, along the z-axis with parametrization

$$\mathbf{x}(s,\theta) = (\varphi(s)\cos\theta, \varphi(s)\sin\theta, \psi(s)), \text{ for } s \in (0,l), \theta \in (0, 2\pi)$$

    (a) Find the second fundamental form of the surface.

    (b) Prove that the Gaussian curvature of the surface is given by

$$K = -\frac{\varphi''}{\varphi}$$

    (c) Prove that the mean curvature of the surface is given by

$$H = \frac{\varphi\psi'' + \varphi'\psi'}{2\varphi\varphi'}$$

    (d) Prove that the surface is a minimal surface if and only if $\varphi\psi'$ is constant.

    (e) Suppose the surface is a minimal surface and $\psi(s) = \sinh^{-1} s = \ln(s + \sqrt{s^2+1})$. Find $\varphi(s)$.

20. Let $\mathbf{x}(u,v)$ be a regular parametrized surface and $\mathbf{n} = \frac{\mathbf{x}_u \times \mathbf{x}_v}{\|\mathbf{x}_u \times \mathbf{x}_v\|}$ be the unit normal vector. Let

$$\begin{aligned} I &= \begin{pmatrix} E & F \\ F & G \end{pmatrix} = \begin{pmatrix} \langle \mathbf{x}_u, \mathbf{x}_u \rangle & \langle \mathbf{x}_u, \mathbf{x}_v \rangle \\ \langle \mathbf{x}_v, \mathbf{x}_u \rangle & \langle \mathbf{x}_v, \mathbf{x}_v \rangle \end{pmatrix} \\ II &= \begin{pmatrix} e & f \\ f & g \end{pmatrix} = \begin{pmatrix} \langle \mathbf{n}, \mathbf{x}_{uu} \rangle & \langle \mathbf{n}, \mathbf{x}_{vu} \rangle \\ \langle \mathbf{n}, \mathbf{x}_{uv} \rangle & \langle \mathbf{n}, \mathbf{x}_{vv} \rangle \end{pmatrix} = \begin{pmatrix} -\langle \mathbf{n}_u, \mathbf{x}_u \rangle & -\langle \mathbf{n}_u, \mathbf{x}_v \rangle \\ -\langle \mathbf{n}_v, \mathbf{x}_u \rangle & -\langle \mathbf{n}_v, \mathbf{x}_v \rangle \end{pmatrix} \end{aligned}$$

be the first and second fundamental form respectively. Suppose

$$\begin{aligned} \mathbf{n}_u &= a_{11}\mathbf{x}_u + a_{12}\mathbf{x}_v \\ \mathbf{n}_v &= a_{21}\mathbf{x}_u + a_{22}\mathbf{x}_v \end{aligned}$$

(a) Prove that

$$II = -\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} I$$

(b) Prove that

$$\mathbf{n}_u \times \mathbf{n}_v = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} \mathbf{x}_u \times \mathbf{x}_v = \frac{eg - f^2}{EG - F^2} \mathbf{x}_u \times \mathbf{x}_v$$

(This exercise shows that $\mathbf{n}_u \times \mathbf{n}_v = K\mathbf{x}_u \times \mathbf{x}_v$.)

(c) Prove that

$$\mathbf{x}_u \times \mathbf{n}_v + \mathbf{n}_u \times \mathbf{x}_v = (a_{11} + a_{22})\mathbf{x}_u \times \mathbf{x}_v = -\left(\frac{eG - 2fF + gE}{EG - F^2}\right) \mathbf{x}_u \times \mathbf{x}_v$$

(This exercise shows that $\mathbf{x}_u \times \mathbf{n}_v + \mathbf{n}_u \times \mathbf{x}_v = -2H\mathbf{x}_u \times \mathbf{x}_v$.)

21. Let $\mathbf{x}(u, v)$ be a regular parametrized surface. A parallel surface of $\mathbf{x}$ is a surface parametrized by

$$\mathbf{y}(u, v) = \mathbf{x}(u, v) + a\mathbf{n}(u, v)$$

where $\mathbf{n} = \frac{\mathbf{x}_u \times \mathbf{x}_v}{\|\mathbf{x}_u \times \mathbf{x}_v\|}$ is the unit normal vector of $\mathbf{x}(u, v)$ and a is a constant.

(a) Prove that

$$\mathbf{y}_u \times \mathbf{y}_v = (1 - 2Ha + Ka^2)\mathbf{x}_u \times \mathbf{x}_v$$

where K and H are the Gaussian and mean curvature of $\mathbf{x}$ respectively.

(b) Prove that the unit normal vector to $\mathbf{y}$ is $\mathbf{n}$.

(c) Prove that at a regular point, the Gaussian curvature of $\mathbf{y}$ is

$$\frac{K}{1 - 2Ha + Ka^2}$$

(d) Prove that at a regular point, the mean curvature of $\mathbf{y}$ is

$$\frac{H - Ka}{1 - 2Ha + Ka^2}$$

(e) Prove that if the mean curvature H of $\mathbf{x}$ is a nonzero constant, then there exists a such that $\mathbf{y}$ has constant Gaussian curvature.

22. Let $\mathbf{r}(s)$, $s \in [0, l]$ be a regular simple closed space curve parametrized by arc length. A tubular surface is a surface S parametrized by

$$\mathbf{x}(s, \theta) = \mathbf{r}(s) + a \cos\theta \mathbf{N}(s) + a \sin\theta \mathbf{B}(s)$$

where $\mathbf{N}(s)$ and $\mathbf{B}(s)$ are the unit normal and binormal to the curve at $\mathbf{r}(s)$ respectively, and a is a constant.

(a) Prove that $\mathbf{x}$ is regular if $a\kappa(s) < 1$ for any s, where $\kappa(s)$ is the curvature of the curve at $\mathbf{r}(s)$.

(b) Prove that the Gaussian curvature of the surface is given by

$$K(s, \theta) = -\frac{\kappa(s) \cos\theta}{a(1 - a\kappa \cos\theta)}$$

(c) Find

$$\iint_S K dA = \int_0^{2\pi} \int_0^l K(s, \theta) \|\mathbf{x}_s \times \mathbf{x}_\theta\| ds d\theta$$

(d) Find the Euler's characteristic of the tubular surface S.

Index