MATH265H LECTURE NOTES

ZHIFEI ZHU

Contents

1. Preliminaries

Math265H covers multiple topics including vector spaces, multi-variable differentiation and integral, and vector calculus. To begin with, we must first ensure that we are on the same page with the notations.

1.1. Set theory notation.

In this course, we will be using the following basic set theory notations.

Definition 1.1. A ***set*** is defined to be a collection of objects. An object in a set is also called an element.

Example 1.2. We can consider a set "alphabet" which is defined to be

$$\text{alphabet} = \{a, b, c, ..., x, y, z\}.$$

In this notation, the set is represented by the name "alphabet". The letters $a, b, c, ..., x, y, z$ in $\{\}$ are the elements in this set. We use the symbol $\in$ to talk about when an element is in a set. For example, the letter a in the set alphabet can be written as

$$a \in \text{alphabet}.$$

We write "/" if we would like to say "not". For example,

$$\gamma \notin \text{alphabet}.$$

The above set alphabet is an example of ***finite set***, where the number of the elements in a set is finite. In some cases, we would like to describe an infinite set. We use ***set builder notation*** to describe it. Consider a set S and a proposition P on the set S. For each element $x \in S$, $P(x)$ can be either true or false. The notation

$$\{x \in S | P(x)\}$$

denotes the set of all elements in S such that P is true.

Example 1.3. The set $\{x \in \text{integers} | x \text{ odd}\}$ is the set of odd integers $\{..., -3, -1, 1, 3, 5, ...\}$.

Example 1.4. Some important infinite sets we will see in this course are the following.

- $\mathbb{N} = \{0, 1, 2, 3, ...\}$ is the set of natural numbers.
- $\mathbb{N}^* = \{1, 2, 3, ...\}$ is the set of positive natural numbers.
- $\mathbb{Z} = \{..., -2, -1, 0, 1, 2, ...\}$ is the set of integers.
- $\mathbb{Q} = \{\frac{a}{b} | a, b \in \mathbb{Z}, b \neq 0\}$ is the set of rational numbers.
- $\mathbb{R}$ is the set of the real numbers. (We will not discuss the definition of the real numbers in this course.)

Definition 1.5. If every element in a set A is also in a set B, then we say A is a subset of B. We denote $A \subseteq B$.

For example, $\mathbb{N}^* \subseteq \mathbb{N} \subseteq \mathbb{Z} \subseteq \mathbb{Q} \subseteq \mathbb{R}$.

Definition 1.6. In this course we will be using the following notations of set operations. Suppose that A and B are two subsets of a set S.

- Union

$$A \cup B = \{x \in S | x \in A \text{ or } x \in B\}$$

- Intersection
$$A \cap B = \{x \in S | x \in A \text{ and } x \in B\}$$
- Complement
$$A^c = \{x \in S | x \notin A\}$$
- Cartesian Product
$$A \times B = \{(x, y) | x \in A \text{ and } y \in B\}$$
Here (x, y) is an ordered tuple. In other words if x, y are distinct elements, then (x, y) and (y, x) represent different elements.

1.2. **Cartesian Coordinates.** (Ch12.1 textbook)

Example 1.7 (2-dimensional Cartesian Coordinates)**.** Let us consider the Cartesian product $\mathbb{R}^2 = \mathbb{R} \times \mathbb{R}$ of the set of the real numbers $\mathbb{R}$. The set $\mathbb{R}^2$ consists of tuples (x, y), where both x and y are real numbers.

One way to visualize $\mathbb{R}^2$ is to consider a ***plane*** with two perpendicular coordinate axes, which are called x-axis and y-axis. Each axis represents a copy of $\mathbb{R}$. The intersection of the two axes is defined to be the ***origin*** O.

Now we may assign the tuple $(0, 0)$ to the origin. Each tuple (x, y) corresponds to a point p on the plane, where the signed distance between p and the y-axis is the quantity x and the signed distance between p and the x-axis is the quantity y. In this way, $\mathbb{R}^2$ can be identified with the entire plane. We call the tuple (x, y) assigned to a point p the (2-dimensional) ***Cartesian coordinate*** of p.

One may generalize this to the identification between the $n-$tuples $\mathbb{R}^n = \underbrace{\mathbb{R} \times \mathbb{R} \times ... \times \mathbb{R}}_{n}$ with the "n-dimensional" space. In this course, we will be mainly studying the 3-dimensional space $\mathbb{R}^3$.

Example 1.8 (3-dimensional Cartesian Coordinates)**.** Consider a 3-dimensional space with three mutually perpendicular coordinate axes intersecting at a single point. Again, this point is defined to be the origin. By convention the axes are labeled with x, y, z-axis according to the "right-hand" rule.

We are going to identify this 3-dimensional space with the set $\mathbb{R}^3$. Note that the elements in $\mathbb{R}^3$ can be represented by 3-tuples (x, y, z). The tuple $(0, 0, 0)$ is identified with the origin O.

A point on the x-axis can be identified with $(x, 0, 0)$, where x is the distance between the point and the origin. Similarly, a point on the y or z axis can be identified with $(0, y, 0)$ or $(0, 0, z)$.

To identify a generic point p with the 3-tuple (x, y, z), let us first consider the following definition.

The x and y axis determined a unique plane in the space. This plane is called the coordinate plane $xy-$plane, or sometimes, $xOy-$plane. Note that this plane consists of two perpendicular axes and an origin. We may identify this plane with $\mathbb{R}^2$ as above. In other words, each point on this plane corresponds to some (x, y) in $\mathbb{R}^2$. Now as a plane in a 3-dimensional space, the points on $xy-$plane with coordinate (x, y) is identified with $(x, y, 0)$ in $\mathbb{R}^3$.

Similarly, one can define the $yz-$plane and the $xz-$plane. And the points in these planes can be identified with $(0, y, z)$ and $(x, 0, z)$.

Now a point p is identified with the tuple (x, y, z), if

- the signed distance between p and the yz-plane is x;
- the signed distance between p and the xz-plane is y;
- the signed distance between p and the xy-plane is z.

Food for thought. Here we are talking about the distance without defining it. Can you give a definition of the distance in the above context? One may use the identification between the real numbers and a straight line. Note that this definition should not depend on the "distance formular" in the next section.

Remark 1.9. Verify that this doesn't conflict with the identification of the points on the axes and the coordinate planes.

The 3-tuple (x, y, z) is called the ($3-$dimensinal) Cartesian coordinate of a point p in the three dimensional space. With this identification, sometimes we write a 3-dimensional space as $\mathbb{R}^3$.

1.3. **Subset of $\mathbb{R}^3$.** (Ch12.1 textbook)

The identification between a points p in the space and a 3-tuple (x, y, z) allows us to describe the geometric objects (a collection of points) using a subset of $\mathbb{R}^3$ (a collection of the elements (x, y, z)).

For example, a straight line that is perpendicular to the $xy-$plane and intersecting the plane at the point $(1, 1, 0)$ can be identified with the subset

$$\{(x, y, z)|x = 1, y = 1\}$$

The plane that is perpendicular to the $y-$axis at $y = 3$ can be identified with

$$\{(x, y, z)|y = 3\}$$

The half-space consisting of the points above the xy-plane can be identified with

$$\{(x, y, z)|z \geq 0\}$$

These equations which determine the subsets are also called the equations of the geometric objects. For example, the equations $x = 1, y = 1$ above determines a subset of $\mathbb{R}^3$. Using the Cartesian coordinate system, we identified this subset with a specific line in the space. We may say the equation of this line is $x = 1, y = 1$.

Example 1.10. Which geometric object does the equation $x^2 + y^2 = 4$ determine in $\mathbb{R}^3$?

Solution. The equation determines a subset of $\mathbb{R}^3$ which is

$$\{(x, y, z)|x^2 + y^2 = 4\}$$

Note that for each fixed $z = z_0$, the points (x, y, z_0) (x,y may vary) determines a plane which is perpendicular to the $z-$axis at $z = z_0$. The equation $x^2 + y^2 = 4$ determines a circle on this plane. Now if we let z_0 vary, these circles will sweep-out a cylinder in the space.

Therefore, the equation $x^2 + y^2 = 4$ can be identified with a cylinder in the space.

Example 1.11. Find out the equation of a circle of radius 2 which lies in the plane $z = -1$ and centered at $(0, 0, -1)$.

Solution. From the previous example, we know that on the $z = -1$ plane, the circle can be identified with $x^2 + y^2 = 4$. Therefore, the equation of the circle is $x^2 + y^2 = 4, z = -1$.

Remark 1.12. There are several (equivalent) ways of writing this circle as a subset of $\mathbb{R}^3$.

$$\{(x, y, z)|x^2 + y^2 = 4, z = -1\} = \{(x, y, -1)|x^2 + y^2 = 4\}$$

In some cases, we simply write the first one $\{(x, y, z)|x^2 + y^2 = 4, z = -1\}$ as $\{x^2 + y^2 = 4, z = -1\}$. We omit a generic for of element (x, y, z).

However, we **cannot** say $\{(x, y, -1)|x^2 + y^2 = 4\} = \{x^2 + y^2 = 4\}$, as $(x, y, -1)$ is not generic. Note that $\{x^2 + y^2 = 4\}$ represents the cylinder in the previous example.

In the above examples, we are trying to identify a geometric object in 3-dimensional space with a subset of $\mathbb{R}^3$. As a result, this object is identified with, in the sense of set theory, the proposition P (the equations) which is used to describe this subset. This identification does not only apply in the case of objects, but also the relation between the objects. For example, we will see in the next week that, "two lines are perpendicular" can be identified with their equations satisfies certain relation.

Below is an application in the case the distance function, and in the future we will study more these kinds of identifications.

Example 1.13 (Triangle law)**.** In $\mathbb{R}^2$, the distance between the points $(x, 0)$ and $(0, y)$ can be computed using the triangle law since the coordinate axes are perpendicular to each other. This distance is in fact $\sqrt{x^2 + y^2}$. More generally, consider two points $p_1 = (x_1, y_1)$ and $p_2 = (x_2, y_2)$ on the plane. One may form a right triangle in the following way (see figure). The triangle law:

$$\text{square of the length of hypotenuse} = \text{the sum of square of the length of sides},$$

can be translated into:

$$\text{square of the distance between } p_1 \text{ and } p_2 = (x_2 - x_1)^2 + (y_2 - y_1)^2,$$

in other words,

$$\text{the distance between } p_1 \text{ and } p_2 = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}.$$

In $\mathbb{R}^3$ this can be used to compute the distance between two points. The distance between two points is the length of the segment connecting the points. Suppose the two points are identified with the coordinates (x_1, y_1, z_1) and (x_2, y_2, z_2), then using the law of triangle, the length of the segment connecting two points is given by

$$\sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2}.$$

This is known as the ***distance formula***.

Example 1.14 (Spheres)**.** With the distance formula, we are able to identify the equation of the spheres. In geometry, a sphere is a collection of the points where the distance between the points and a fixed point is a constant called the radius.

Suppose that the fixed point has coordinate (x_1, y_1, z_1) and the constant distance, i.e., radius, is r. Then a point on the sphere, if we identify it with (x, y, z), satisfies

$$\sqrt{(x - x_1)^2 + (y - y_1)^2 + (z - z_1)^2} = r,$$

or equivalently,

$$(x - x_1)^2 + (y - y_1)^2 + (z - z_1)^2 = r^2,$$

Therefore, as a subset of $\mathbb{R}^3$, the sphere can be identified with

$$\{(x,y,z)|(x-x_1)^2+(y-y_1)^2+(z-z_1)^2=r^2\}$$

The equation $(x-x_1)^2+(y-y_1)^2+(z-z_1)^2=r^2$ is called the ***standard equation*** of the sphere centered at (x_1,y_1,z_1) of radius r.

Example 1.15. Find the radius and center of the sphere $x^2+y^2-2y+z^2+z=1$.

Solution. In order to find the center and the radius of the sphere, we write the equation in standard form. By completing the square,

$$x^2+y^2-2y+z^2+z=1 \Rightarrow x^2+(y-1)^2+(z+\frac{1}{2})^2=\frac{9}{4}$$

Therefore, the center is $(0,1,-\frac{1}{2})$ and the radius is $\frac{3}{2}$.

Example 1.16. Write down the following region as a subset of $\mathbb{R}^3$.

1. The first octant.
   Solution.
   $$\{(x,y,z)|x\geq 0, y\geq 0, z\geq 0\}.$$
2. The interior of a sphere of radius 3 centered at the origin.
   Solution.
   $$\{(x,y,z)|x^2+y^2+z^2<9\}.$$
3. The upper hemisphere cut from the sphere $x^2+y^2+z^2=1$ by the $xy-$plane.
   Solution.
   $$\{(x,y,z)|x^2+y^2+z^2=1, z\geq 0\}.$$

1.4. **Cylindrical and Spherical Coordinates.** (Ch15.7, Definition only.)

In addition to the Cartesian Coordinate, there are two other type of coordinates in the three dimensional space which are very useful in this course.

Definition 1.17 (Cylindrical coordinates)**.** The cylindrical coordinate of a point p in the space is a 3-tuple (r,θ,z), where

- (r,θ) is the polar coordinate of the projection of the point p onto the $xy-$plane.
- z is the z-coordinate of the Cartesian coordinate of p.

Remark 1.18. If we view the 3-dimensional space with the Cartesian coordinate as a product $\mathbb{R}^2\times\mathbb{R}$, where the first $\mathbb{R}^2$ gives us the Cartesian coordinate of the $xy-$plane, then the cylindrical coordinate is $\mathbb{R}^2\times\mathbb{R}$, where the plane is equipped with the polar coordinate.

Therefore the relation between the cylindrical coordinate and the Cartesian coordinate is the following.

Proposition 1.19. *Let p be a point in $\mathbb{R}^3$ with Cartesian coordinate (x,y,z). Suppose that (r,θ,z) is the cylindrical coordinate of the point p, then*

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

$$r^2=x^2+y^2, \tan\theta=y/x.$$

Note that the cylindrical coordinate is not a "one-to-one" correspondence between the points in the space and an element (r,θ,z) in $\mathbb{R}^3$.

Definition 1.20 (Spherical coordinates)**.** The spherical coordinate of a point p in the space is a 3-tuple (ρ, ϕ, θ), where

- $\rho \geq 0$ is the distance between the point p and the origin.
- ϕ is the angle between the segment OP and the positive $z-$axis. $0 \leq \phi \leq \pi$.
- θ is the angle in the cylindrical coordinate.

Proposition 1.21. *Let p be a point in $\mathbb{R}^3$ with cylindrical coordinate (r, θ, z) and spherical coordinate (ρ, ϕ, θ), then*

$$z = \rho \cos \phi, r = \rho \sin \phi, \theta = \theta.$$

Proposition 1.22. *Let p be a point in $\mathbb{R}^3$ with Cartesian coordinate (x, y, z) and spherical coordinate (ρ, ϕ, θ), then*

$$x = \rho \sin \phi \cos \theta, y = \rho \sin \phi \sin \theta, z = \rho \cos \phi.$$

$$\rho^2 = x^2 + y^2 + z^2.$$

Similar to the Cartesian coordinates, one can represent regions in the 3-dimensional space using both cylindrical coordinates and spherical coordinates.

Example 1.23. In cylindrical coordinates, the Cartesian coordinate planes can be represented by the following.

- $xy-$plane: $z = 0$.
- $yz-$plane: $\theta = \pi/2$.
- $xz-$plane: $\theta = 0$.

Example 1.24. Find out the spherical coordinate of the sphere $x^2 + y^2 + (z-1)^2 = 1$.

Solution. To find the equation, we simply substitute using Proposition 1.22.

$$\begin{aligned} x^2 + (y+1)^2 + (z-1)^2 &= 1 \\ (\rho \sin \phi \cos \theta)^2 + (\rho \sin \phi \sin \theta)^2 + (\rho \cos \phi - 1)^2 &= 1 \\ \rho^2 \sin^2 \phi + \rho^2 \cos^2 \phi + 1 - 2\rho \cos \phi &= 1 \\ \rho^2 - 2\rho \cos \phi &= 0 \\ \rho &= 2 \cos \phi. \end{aligned}$$

2. Vectors and the geometry of space

2.1. **Vectors.** (Ch 12.2)

A ***geometric vector*** is a directed line segment. Suppose that $\vec{v} = \overrightarrow{AB}$ is a (geometric) vector pointing from a point A to B. A is called the initial point and B the terminal point. The length of the vector is the length of the segment $\overrightarrow{AB}$, which is denoted by $|\vec{v}|$ or $|\overrightarrow{AB}|$.

Definition 2.1. Two vectors are equal if they have the same length and direction.

Remark 2.2. Note that this is saying two vectors can be equal regardless of the initial points. A vector, up to vector equivalent, is determined by its direction and length. We may also determine a vector using its initial and terminal point.

Definition 2.3. There are two operations on vectors which are called ***vector addition*** and ***scalar multiplication***.

For geometric vectors, the vector addition is achieved by the ***parallelogram law*** (or ***triangle law***). Suppose that $\vec{v}$ and $\vec{u}$ are two vectors. We place the initial point of the vector $\vec{u}$ at the terminal point of $\vec{v}$. In this case, the vector $\vec{v}+\vec{u}$ is defined to be a vector whose initial point is the initial point of $\vec{v}$ and the terminal point is the terminal point of $\vec{u}$.

Suppose that k is a scalar. If $k \geq 0$, then the scalar multiplication $k\vec{v}$ is a vector in the same direction of $\vec{v}$ such that the length $|k\vec{v}|$ is $k \cdot |\vec{v}|$. For $k < 0$, then $k\vec{v}$ is in the opposite direction of $\vec{v}$ with length $|k||\vec{v}|$.

Remark 2.4. Note that the above definition of the vector addition and scalar multiplication works for both vectors in a two or three dimensional space.

The vector addition and scalar multiplication satisfies the following basic properties.

Proposition 2.5. *Let $\vec{u}, \vec{v}, \vec{w}$ be vectors and a, b be scalars.*

- $\vec{u}+\vec{v}=\vec{v}+\vec{u}$
- $(\vec{u}+\vec{v})+\vec{w}=\vec{u}+(\vec{v}+\vec{w})$
- $\vec{u}+\vec{0}=\vec{u}$
- $\vec{u}+(-\vec{u})=\vec{0}$
- $1\vec{u}=\vec{u}$
- $a(b\vec{u})=(ab)\vec{u}$
- $a(\vec{u}+\vec{v})=a\vec{u}+a\vec{v}$
- $(a+b)\vec{u}=a\vec{u}+b\vec{u}$

Definition 2.6 (Vector space)**.** Let V be the collection of all vectors in a two or three dimensional space. We say that V is a vector space, if V is equipped with two operations vector addition and scalar multiplication, and these two operations satisfies Proposition 2.5

Remark 2.7. In general, we say a set S is a vector space, if S is equipped with these two operations which satisfies the above proposition. In this case, the operations can be abstract and the element in S is called an abstract vector, or simply just vector.

In a two or three dimensional space, we may identify the points with $\mathbb{R}^2$ or $\mathbb{R}^3$ using the Cartesian coordinate. This identification allows us to represent a geometric vector using the coordinates.

Definition 2.8 (Coordinate representation of geometric vectors)**.** Let $\vec{v}=\overrightarrow{AB}$ be a geometric vector in a 3-dimensional space with initial point A and terminal point B. If the Cartesian coordinates of $A=(x_1, y_1, z_1)$ and $B=(x_2, y_2, z_2)$, then the coordinate of the vector $\vec{v}$ is defined to be

$$\mathbf{v}=\langle x_2-x_1, y_2-y_1, z_2-z_1\rangle,$$

which is also called the component form of $\vec{v}$. Each number in the component form is called a component.

Note that is the initial point $A=O$ is the origin, then the component form of the vector is the same as the coordinate of the terminal point B.

Proposition 2.9. *Two vectors are equal if and only if they have the same component form.*

Proof. Indeed, by definition, we may place the vector so that the initial point is the origin. In this case, two vectors are equal if and only if they have the same terminal points whose coordinate is the component form of the vectors. ☐

Using the coordinate representation or the component form of the vectors, we may write down the two vector operations.

Proposition 2.10. *Let* $\mathbf{u} = \langle u_1, u_2, u_3 \rangle$ *and* $\mathbf{v} = \langle v_1, v_2, v_3 \rangle$ *be the component forms of two vectors* $\vec{u}$ *and* $\vec{v}$*, and* k *a scalar. Then the component form of the vector* $\vec{u} + \vec{v}$ *and* $k\vec{u}$ *is given by the following.*

$$\mathbf{u} + \mathbf{v} = \langle u_1 + v_1, u_2 + v_2, u_3 + v_3 \rangle$$
$$k\mathbf{u} = \langle ku_1, ku_2, ku_3 \rangle$$

This is also called component-wise addition and scalar multiplication.

Definition 2.11 (Vector space structure on $\mathbb{R}^3$)**.** Let V be the geometric vector space of the vectors in a 3-dimensional space. Using the component form of the vectors, we may identify the element in V with the element in $\mathbb{R}^3$, namely,

$$\mathbf{v} = \langle v_1, v_2, v_3 \rangle \leftrightarrow (v_1, v_2, v_3) \in \mathbb{R}^3$$

And the component-wise addition and scalar multiplication naturally extends to $\mathbb{R}^3$. One may verify that these two operations on $\mathbb{R}^3$ satisfies Proposition 2.5. And thus, defines a vector space structure on $\mathbb{R}^3$.

Under the above identification, sometimes, we do not differ the following:

$$\text{Geometric vector space with vector equivalent}$$
$$\updownarrow$$
$$\text{Component form of vectors}$$
$$\updownarrow$$
$$\mathbb{R}^3$$

To determine a vector, we may use the initial point and the terminal point. Another way to determine a vector is through its length and direction.

The length of a vector can be obtained using the distance formula. If $\vec{v}$ is a vector from a initial point $A = (x_1, y_1, z_1)$ to a terminal point $B = (x_2, y_2, z_2)$. Then

$$|\vec{v}| = |\mathbf{v}| = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2}.$$

Sometimes the length is also called the ***magnitude*** of a vector. The only vector of length 0 is the zero vector $\mathbf{0} = \langle 0, 0, 0 \rangle$.

The direction of a vector can be represented by the ***unit vector***, i.e., a vector of length 1. If $\mathbf{v} = \langle v_1, v_2, v_3 \rangle$, then the unit vector in the direction of $\mathbf{v}$ can be computed by

$$\mathbf{u} = \frac{\langle v_1, v_2, v_3 \rangle}{|\mathbf{v}|}.$$

Example 2.12. Find the magnitude of the vector $\langle 1, 1, 1 \rangle$.

Solution. We simply apply the formula for magnitude.

$$|\langle 1, 1, 1 \rangle| = \sqrt{1 + 1 + 1} = \sqrt{3}.$$

Example 2.13. The unit vector in the positive axes direction can be written as

$$\mathbf{i} = \langle 1, 0, 0\rangle, \mathbf{j} = \langle 0, 1, 0\rangle, \mathbf{k} = \langle 0, 0, 1\rangle$$

Therefore, a vector $\mathbf{v} = \langle v_1, v_2, v_3\rangle$ can be also written as

$$\mathbf{v} = v_1\mathbf{i} + v_2\mathbf{j} + v_3\mathbf{k}.$$

For example,

$$\langle 2, 3, -1\rangle = 2\mathbf{i} + 3\mathbf{j} - \mathbf{k}.$$

Example 2.14. Find a vector of magnitude 5 in the direction of $\langle 2, 1, -1\rangle$.

Solution. To find this vector, let us first compute the unit vector in this direction.

$$\mathbf{u} = \frac{\langle 2, 1, -1\rangle}{\sqrt{4+1+1}} = \langle \frac{2}{\sqrt{6}}, \frac{1}{\sqrt{6}}, \frac{-1}{\sqrt{6}}\rangle.$$

Now a vector in the direction of $\mathbf{u}$ of magnitude 5 is $5\mathbf{u}$. In other words,

$$5 \cdot \langle \frac{2}{\sqrt{6}}, \frac{1}{\sqrt{6}}, \frac{-1}{\sqrt{6}}\rangle = \langle \frac{10}{\sqrt{6}}, \frac{5}{\sqrt{6}}, \frac{-5}{\sqrt{6}}\rangle..$$

2.2. **Dot and cross product.** (Ch12.3,12.4)

There are two important products between vectors, the dot product and the cross product. The dot product can be used to measure the angles between vectors.

Consider a triangle ABC. let θ be the angle between AB and AC. Suppose that $\overrightarrow{AB} = \mathbf{u} = \langle u_1, u_2, u_3\rangle$, $\overrightarrow{AC} = \mathbf{v} = \langle v_1, v_2, v_3\rangle$ and $\overrightarrow{BC} = \mathbf{w} = \langle w_1, w_2, w_3\rangle$. Then

$$\mathbf{w} = \mathbf{v} - \mathbf{u} = \langle v_1 - u_1, v_2 - u_2, v_3 - u_3\rangle.$$

Applying the law of cosine, we obtain that

$$\cos\theta = \frac{|\mathbf{u}|^2 + |\mathbf{v}|^2 - |\mathbf{w}|^2}{2|\mathbf{u}||\mathbf{v}|}.$$

And the numerator

$$|\mathbf{u}|^2 + |\mathbf{v}|^2 - |\mathbf{w}|^2 = 2u_1v_1 + 2u_2v_2 + 2u_3v_3.$$

Therefore,

$$\cos\theta = \frac{u_1v_1 + u_2v_2 + u_3v_3}{|\mathbf{u}||\mathbf{v}|}.$$

We define the ***dot product*** between the vector $\mathbf{u}$ and $\mathbf{v}$ to be

$$\mathbf{u} \cdot \mathbf{v} = u_1v_1 + u_2v_2 + u_3v_3.$$

And then

$$\cos\theta = \frac{\mathbf{u} \cdot \mathbf{v}}{|\mathbf{u}||\mathbf{v}|}.$$

This immediately tells us that if two vectors $\mathbf{u}$ and $\mathbf{v}$ are perpendicular, then $\cos\theta = 0$ and therefore,

Proposition 2.15. $\mathbf{u} \perp \mathbf{v}$ *if and only if* $\mathbf{u} \cdot \mathbf{v} = 0$. *We also say that* $\mathbf{u}$ *and* $\mathbf{v}$ *are orthogonal.*

Example 2.16. Find the angle between two vectors $\mathbf{u} = \langle 1, 1, 1\rangle$ and $\mathbf{v} = \langle -1, 2, 3\rangle$.

Solution. We first compute the dot product

$$\mathbf{u} \cdot \mathbf{v} = 1 \cdot -1 + 1 \cdot 2 + 1 \cdot 3 = 4.$$

And the magnitude of the vectors can be found by

$$|\mathbf{u}| = \sqrt{1+1+1} = \sqrt{3}$$
$$|\mathbf{v}| = \sqrt{1+4+9} = \sqrt{14}$$

Therefore,

$$\cos\theta = \frac{4}{\sqrt{3} \cdot \sqrt{14}} = \frac{4}{\sqrt{42}}.$$

And the angle

$$\theta = \arccos \frac{4}{\sqrt{42}}.$$

One may check that the dot product satisfies the following basic properties.

Proposition 2.17. *Suppose* $\mathbf{u}$, $\mathbf{v}$ *and* $\mathbf{w}$ *are vectors and* k *a scalar.*

- $\mathbf{u} \cdot \mathbf{v} = \mathbf{v} \cdot \mathbf{u}$ *(symmetric)*
- $(\mathbf{u} + \mathbf{v}) \cdot \mathbf{w} = \mathbf{u} \cdot \mathbf{w} + \mathbf{v} \cdot \mathbf{w}$
  $(k\mathbf{u}) \cdot \mathbf{v} = k(\mathbf{u} \cdot \mathbf{v})$ *(linearity)*
- $\mathbf{u} \cdot \mathbf{u} = |\mathbf{u}|^2 \geq 0$ *(positive definite)*
  The equality holds, in other words $\mathbf{u} \cdot \mathbf{u} = 0$, *if and only if* $\mathbf{u} = \mathbf{0}$.

The dot product allows us to compute the ***vector projection***. The vector projection of a vector $\mathbf{u}$ on to a vector $\mathbf{v}$, which is denoted by $\operatorname{proj}_{\mathbf{v}} \mathbf{u}$, is the vector obtained by dropping a perpendicular from the vector $\mathbf{u}$ to the vector $\mathbf{v}$. If θ is the angle between the vectors $\mathbf{u}$ and $\mathbf{v}$, then

$$\operatorname{proj}_{\mathbf{v}} \mathbf{u} = (|\mathbf{u}| \cos\theta) \frac{\mathbf{v}}{|\mathbf{v}|}.$$

Using the dot product formula of $\cos\theta$, we obtain that

$$\operatorname{proj}_{\mathbf{v}} \mathbf{u} = \left(\frac{\mathbf{u} \cdot \mathbf{v}}{|\mathbf{v}|}\right) \frac{\mathbf{v}}{|\mathbf{v}|} = \left(\frac{\mathbf{u} \cdot \mathbf{v}}{|\mathbf{v}|^2}\right) \mathbf{v}.$$

If $\mathbf{n}$ is the unit vector in the direction of $\mathbf{v}$, then

$$\operatorname{proj}_{\mathbf{v}} \mathbf{u} = (\mathbf{u} \cdot \mathbf{n})\mathbf{n}.$$

The magnitude of $\operatorname{proj}_{\mathbf{v}} \mathbf{u}$ is $\left|\frac{\mathbf{u} \cdot \mathbf{v}}{|\mathbf{v}|}\right|$ and $\frac{\mathbf{u} \cdot \mathbf{v}}{|\mathbf{v}|}$ is called the scalar component of $\operatorname{proj}_{\mathbf{v}} \mathbf{u}$.

Example 2.18. Find the vector projection of $\mathbf{u} = \langle 1, 2, 3\rangle$ onto $\mathbf{v} = \langle 1, 0, -1\rangle$.

Solution. Using the projection formula,

$$\operatorname{proj}_{\mathbf{v}} \mathbf{u} = \left(\frac{\mathbf{u} \cdot \mathbf{v}}{|\mathbf{v}|^2}\right) \mathbf{v}.$$

And

$$\mathbf{u} \cdot \mathbf{v} = 1 + 0 - 3 = -2$$
$$|\mathbf{v}| = \sqrt{1+1} = \sqrt{2}.$$

Therefore,

$$\operatorname{proj}_{\mathbf{v}} \mathbf{u} = \frac{-2}{2}\langle 1, 0, -1\rangle = \langle -1, 0, 1\rangle.$$

Example 2.19. Verify that $\mathbf{u} - \operatorname{proj}_{\mathbf{v}} \mathbf{u}$ is perpendicular to $\operatorname{proj}_{\mathbf{v}} \mathbf{u}$.

Solution. To show that two vectors are perpendicular, it suffices to show that their dot product is 0.

$$\begin{aligned}(\mathbf{u} - \operatorname{proj}_{\mathbf{v}} \mathbf{u}) \cdot \operatorname{proj}_{\mathbf{v}} \mathbf{u} &= \left(\frac{\mathbf{u}\cdot\mathbf{v}}{|\mathbf{v}|^2}\right)(\mathbf{u}\cdot\mathbf{v}) - \left(\frac{\mathbf{u}\cdot\mathbf{v}}{|\mathbf{v}|^2}\right)^2(\mathbf{v}\cdot\mathbf{v}) \\ &= \frac{(\mathbf{u}\cdot\mathbf{v})^2}{|\mathbf{v}|^2} - \frac{(\mathbf{u}\cdot\mathbf{v})^2}{|\mathbf{v}|^2} \\ &= 0.\end{aligned}$$

Therefore, the vector $\mathbf{u} - \operatorname{proj}_{\mathbf{v}} \mathbf{u}$ is perpendicular to $\operatorname{proj}_{\mathbf{v}} \mathbf{u}$.

Given two vectors $\mathbf{u}$ and $\mathbf{v}$ which are non-parallel, let $\mathbf{n}$ be a unit vector which is perpendicular to both $\mathbf{u}$ and $\mathbf{v}$. Note that there are two possible unit vectors in this case. We fix $\mathbf{n}$ to be the vector determined by the right-hand rule with respect to $\mathbf{u}$ and $\mathbf{v}$. Let θ be the angle between $\mathbf{u}$ and $\mathbf{v}$. Then the ***cross product*** is the vector

$$\mathbf{u} \times \mathbf{v} = (|\mathbf{u}||\mathbf{v}| \sin\theta)\mathbf{n}$$

If $\mathbf{u} \parallel \mathbf{v}$, then since $\theta = 0$,

$$\mathbf{u} \times \mathbf{v} = \mathbf{0}.$$

The cross product satisfies the following properties.

Proposition 2.20. *Suppose* $\mathbf{u}$, $\mathbf{v}$ *and* $\mathbf{w}$ *are vectors and* k *a scalar.*

- $\mathbf{u} \times \mathbf{v} = -\mathbf{v} \times \mathbf{u}$ *(skew-symmetric)*
- $(\mathbf{u} + \mathbf{v}) \times \mathbf{w} = \mathbf{u} \times \mathbf{w} + \mathbf{v} \times \mathbf{w}$
  $(k\mathbf{u}) \times \mathbf{v} = k(\mathbf{u} \times \mathbf{v})$ *(linearity)*

Note that using the skew-symmetric property, the linearity is also true for the second vector.

To compute the component form of the cross product $\mathbf{u}\times\mathbf{v}$, first note that by the definition of the cross product, if we apply it on the unit vector in the coordinate axes direction

$$\mathbf{i} = \langle 1, 0, 0\rangle, \mathbf{j} = \langle 0, 1, 0\rangle, \mathbf{k} = \langle 0, 0, 1\rangle$$

Then we obtain the following relations

$$\begin{aligned}\mathbf{i} \times \mathbf{j} &= \mathbf{k} \\ \mathbf{j} \times \mathbf{k} &= \mathbf{i} \\ \mathbf{k} \times \mathbf{i} &= \mathbf{j}.\end{aligned}$$

Now, suppose that two vectors $\mathbf{u} = \langle u_1, u_2, u_3\rangle$ and $\mathbf{v} = \langle v_1, v_2, v_3\rangle$. Then

$$\begin{aligned}\mathbf{u} &= u_1\mathbf{i} + u_2\mathbf{j} + u_3\mathbf{k} \\ \mathbf{v} &= v_1\mathbf{i} + v_2\mathbf{j} + v_3\mathbf{k}\end{aligned}$$

Therefore, using Proposition 2.20, in component form,

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

This happens to be the ***determinant*** of the following matrix in linear algebra.

$$\det\begin{pmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ u_1 & u_2 & u_3 \\ v_1 & v_2 & v_3 \end{pmatrix}$$

Remark 2.21. Recall that the determinant of a 3×3 matrix can be computed by first copy the matrix twice, then take the sum of the product of the diagonal entries and subtract the sum of the product of the anti-diagonal entries.

Example 2.22. Find out a unit vector which is perpendicular to both $\langle 1, 0, 1\rangle$ and $\langle 1, -2, 2\rangle$.

Solution. By the definition of the cross product, we simply have to find the unit vector in the direction of $\langle 1, 0, 1\rangle \times \langle 1, -2, 2\rangle$. We compute that

$$\begin{aligned} \langle 1, 0, 1\rangle \times \langle 1, -2, 2\rangle &= \det\begin{pmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ 1 & 0 & 1 \\ 1 & -2 & 2 \end{pmatrix} \\ &= (0+2)\mathbf{i} - (2-1)\mathbf{j} + (-2-0)\mathbf{k} \\ &= \langle 2, -1, -2\rangle. \end{aligned}$$

Therefore, one unit vector in the direction of $\langle 2, -1, -2\rangle$ is

$$\frac{\langle 2, -1, -2\rangle}{|\langle 2, -1, -2\rangle|} = \frac{1}{\sqrt{9}}\langle 2, -1, -2\rangle = \langle \frac{2}{3}, -\frac{1}{3}, -\frac{2}{3}\rangle.$$

Example 2.23. Find out the cross product of the vectors $\mathbf{u} = \langle 1, 2, 3\rangle$ and $\mathbf{v} = \langle 1, 0, -1\rangle$.

Solution. Using the component formula of the cross product, we obtain that

$$\begin{aligned} \mathbf{u} \times \mathbf{v} &= \det\begin{pmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ 1 & 2 & 3 \\ 1 & 0 & -1 \end{pmatrix} \\ &= (-2-0)\mathbf{i} - (-1-3)\mathbf{j} + (0-2)\mathbf{k} \\ &= \langle -2, 4, -2\rangle. \end{aligned}$$

Example 2.24. Find out the area of the triangle spanned by $\mathbf{u} = \langle 1, 2, 3\rangle$ and $\mathbf{v} = \langle 1, 0, -1\rangle$.

Solution. Let θ be the angle between $\mathbf{u}$ and $\mathbf{v}$. Then the area is

$$\text{Area} = \frac{1}{2}|\mathbf{u}|\sin\theta|\mathbf{v}|.$$

Therefore, using the definition of the cross product,

$$\text{Area} = \frac{1}{2}|\mathbf{u} \times \mathbf{v}|,$$

where $|\mathbf{u} \times \mathbf{v}|$ is the magnitude of $\mathbf{u} \times \mathbf{v}$. Now from the previous example, we have computed that

$$\mathbf{u} \times \mathbf{v} = \langle -2, 4, -2\rangle.$$

Thus, the area of the triangle is

$$\text{Area} = \frac{1}{2}|\mathbf{u} \times \mathbf{v}| = \frac{1}{2}\sqrt{4+16+4} = \frac{\sqrt{24}}{2} = \sqrt{6}.$$

From the above example, we see that in fact the magnitude of the cross product $|\mathbf{u} \times \mathbf{v}|$ represents the area of the parallelogram spanned by the vectors $\mathbf{u}$ and $\mathbf{v}$. If we generalize this to the 3-dimensional case, consider three vectors $\mathbf{u}$, $\mathbf{v}$ and $\mathbf{w}$. Let θ be the angle between $\mathbf{w}$ and $\mathbf{u} \times \mathbf{v}$. Then the signed volume (depending on the sign of $\cos\theta$) of the parallelepiped spanned by $\mathbf{u}$, $\mathbf{v}$ and $\mathbf{w}$ is

$$|\mathbf{u} \times \mathbf{v}||\mathbf{w}| \cos\theta = (\mathbf{u} \times \mathbf{v}) \cdot \mathbf{w}.$$

We define the ***triple product*** between $\mathbf{u}$, $\mathbf{v}$ and $\mathbf{w}$ to be

$$(\mathbf{u} \times \mathbf{v}) \cdot \mathbf{w}$$

Note that the sign of $(\mathbf{u} \times \mathbf{v}) \cdot \mathbf{w}$ only depends on whether $\mathbf{u}$, $\mathbf{v}$ and $\mathbf{w}$ satisfies the right-hand or left-hand rule relation, and in particular, if $\mathbf{n}$ is the unit vector in the definition of $\mathbf{u} \times \mathbf{v}$, then the sign of $(\mathbf{u} \times \mathbf{v}) \cdot \mathbf{n}$ is $+1$.

2.3. **Equations of lines and planes.** (Ch12.5)

In this section we will study how to represent a line or a plane in the 3-dimensional space using vectors and other type of equations.

A line in a 3-dimensional space is determined by two points $P = (x_1, y_1, z_1)$ and $Q = (x_2, y_2, z_2)$, or equivalently, a point P and a vector in the direction of the line, say $\mathbf{v} = \langle x_2 - x_1, y_2 - y_1, z_2 - z_1\rangle$.

Suppose that $R = (x, y, z)$ is a point on the line. Consider the vector $\mathbf{r} = \overrightarrow{OR} = \langle x, y, z\rangle$. Because R and P are both on the line, the vector $\overrightarrow{PR} = \overrightarrow{OR} - \overrightarrow{OP}$ must be parallel to $\mathbf{v}$. If we denote the vector $\overrightarrow{OP}$ by $\mathbf{r_0}$, then the above relation suggests that there is a scalar t such that

$$\mathbf{r} - \mathbf{r_0} = t\mathbf{v}$$

This is called the ***vector equation*** of the line. The vector $\mathbf{v}$ in the equation is called a direction vector of the line. In fact, when the scalar t changes between $(-\infty, \infty)$, the terminal point R of the vector $\mathbf{r}$ will sweep out the entire line.

Component wise, if we assume that $\mathbf{v} = \langle v_1, v_2, v_3\rangle$, then

$$\langle x, y, z\rangle - \langle x_1, y_1, z_1\rangle = t\langle v_1, v_2, v_3\rangle.$$

In other words, we obtain that

$$\begin{aligned} x =& x_1 + tv_1 \\ y =& y_1 + tv_2 \\ z =& z_1 + tv_3 \end{aligned}$$

These equations are called ***parametric equations*** of the line and t is called a parameter.

Example 2.25. Find out the parametric equation of a line passing through the points $(1, 0, 0)$ and $(2, -1, 3)$.

Solution. We may choose $P = (1, 0, 0)$ and the vector $\mathbf{v} = \langle 2-1, -1-0, 3-0\rangle = \langle 1, -1, 3\rangle$. Suppose that $\mathbf{r} = \langle x, y, z\rangle$, then

$$\begin{aligned} x =& 1 + t \\ y =& -t \\ z =& 3t. \end{aligned}$$

Note that in the above example, the solution is not unique. We may also pick $(2, -1, 3)$ to be the point P, which will give us a different expression of the vector $\mathbf{v}$ as well as the equation of the line. In fact, recall that in the case of dimension 2, a line, say $y = x + 2$ can be also written as $2y = 2x + 2$ or any other constant multiple of the equation. We may ask if there is a similar expression in the case of dimension 3. Using the parametric equation, we see that if $v_1, v_2, v_3 \neq 0$, then

$$\begin{aligned} t =& \frac{x - x_1}{v_1} \\ t =& \frac{y - y_1}{v_2} \\ t =& \frac{z - z_1}{v_3} \end{aligned}$$

And therefore if we equating the results, we obtain that

$$\frac{x - x_1}{v_1} = \frac{y - y_1}{v_2} = \frac{z - z_1}{v_3}.$$

These equations are called ***symmetric equation*** of the line. In the case when v_1 or v_2 or $v_3 = 0$, we may write, for example, when $v_1 = 0$,

$$x = x_1, \frac{y - y_1}{v_2} = \frac{z - z_1}{v_3}.$$

Example 2.26. The symmetric equation of the line in Example 2.25 is

$$\frac{x - 1}{1} = \frac{y}{-1} = \frac{z}{3},$$

or when simplified,

$$x - 1 = -y = \frac{z}{3}.$$

In the symmetric equation of the line, we have two identities. We will see that each of these two identities represents a plane in the space and thus, a line can be viewed as the intersection of the planes in a 3-dimensional space.

To describe a plane in the space, suppose that we know a point $P = (x_1, y_1, z_1)$ is in the plane. A single vector in the plane is not enough to determine the direction of the entire plane. However, if we are given a vector $\mathbf{n}$ which is perpendicular to the plane, then with the point P, we are able to determine the plane. We call $\mathbf{n}$ a ***normal vector*** of the plane.

Remark 2.27. A normal vector $\mathbf{n}$ **does not** have to be a unit vector.

In this case, let $R = (x, y, z)$ be a generic point on the plane. Then the vector $\overrightarrow{PR}$ must be perpendicular to $\mathbf{n}$. In other words,

$$\overrightarrow{PR} \cdot \mathbf{n} = 0.$$

If we denote the vector $\overrightarrow{OR} = \mathbf{r}$ and $\overrightarrow{OP} = \mathbf{r_0}$, where O is the origin, then

$$(\mathbf{r} - \mathbf{r_0}) \cdot \mathbf{n} = 0.$$

This is called the ***vector equation*** of the plane.

Now suppose that $\mathbf{n} = \langle a, b, c \rangle$, the above equation becomes

$$\langle x - x_1, y - y_1, z - z_1 \rangle \cdot \langle a, b, c \rangle = 0.$$

or

$$a(x - x_1) + b(y - y_1) + c(z - z_1) = 0.$$

This is called the ***scalar equation*** of the plane. If we simplify the equation by letting $d = -ax_1 - by_1 - cz_1$, then the above equation becomes

$$ax + by + cz + d = 0.$$

Note that the coefficients a, b, c of x, y, z are the coordinates of a normal vector to the plane.

Example 2.28. The symmetric equation $x - 1 = -y = \frac{z}{3}$ in Example 2.26 consists of several planes

$$x + y - 1 = 0, x - \frac{z}{3} - 1 = 0 \text{ and } y + \frac{z}{3} = 0.$$

The line can be viewed as the intersection of any two of these three planes.

Example 2.29. Find an equation of the plane that passes through the points $P = (1, 3, 2)$, $Q = (1, 1, 0)$ and $R = (2, 5, 0)$.

Solution. To write the equation of the plane, we need a point from the plane and the coordinate of a normal vector. Recall that similar to the Example 2.22, we may form two vectors $\overrightarrow{PQ} = \langle 0, -2, -2\rangle$ and $\overrightarrow{PR} = \langle 1, 2, -2\rangle$ using the three given points. Therefore, a normal vector $\mathbf{n}$ can be found by taking the cross product $\overrightarrow{PQ} \times \overrightarrow{PR}$.

$$\begin{aligned}
\mathbf{n} = \langle 0, -2, -2\rangle \times \langle 1, 2, -2\rangle &= \det \begin{pmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ 0 & -2 & -2 \\ 1 & 2 & -2 \end{pmatrix} \\
&= (4 + 4)\mathbf{i} - (0 + 2)\mathbf{j} + (0 + 2)\mathbf{k} \\
&= \langle 8, -2, 2\rangle.
\end{aligned}$$

Therefore, if we pick $P = (1, 3, 2)$ to plug in to the equation, an equation of the plane can be computed by

$$8(x - 1) - 2(y - 3) + 2(z - 2) = 0,$$

or when simplified,

$$4x - y + z - 3 = 0.$$

Remark 2.30. Note that if we pick a different point, for example, $Q = (1, 1, 0)$, on the plane, then

$$8(x - 1) - 2(y - 1) + 2(z - 0) = 0$$

will still give us the same simplified equation

$$4x - y + z - 3 = 0.$$

By choosing a different normal vector and points on the plane, the equation we obtain may be differed by a constant multiplication.

Example 2.31. Find the parametric equation of the line of intersection of the planes $x + y + z = 1$ and $x - 2y + 3z = 1$.

Solution. We will use two different methods to solve this problem.

1. We may first find the symmetric equation of the line using the equations of the plane. Given that
$$x + y + z = 1, x - 2y + 3z = 1.$$
By eliminating the variable y, we get
$$3x + 5z = 3 \Rightarrow \frac{x-1}{-5} = \frac{z}{3}.$$
By eliminating the variable x, we get
$$3y - 2z = 0 \Rightarrow \frac{y}{2} = \frac{z}{3}.$$
Therefore,
$$\frac{x-1}{-5} = \frac{y}{2} = \frac{z}{3}.$$
If we let $\frac{x-1}{-5} = \frac{y}{2} = \frac{z}{3} = t$, then
$$x = -5t + 1$$
$$y = 2t$$
$$z = 3t$$

2. Another way to find the equation of the line of the intersection is the following. We know that to determine the equation, we need a point on the line and a directional vector. Because a point P on the line satisfies both of the equations
$$x + y + z = 1, x - 2y + 3z = 1,$$
we may pick $P = (1, 0, 0)$. Indeed,
$$1 + 0 + 0 = 1, 1 - 2 \cdot 0 + 3 \cdot 0 = 1.$$
Now we compute a directional vector. Note that a directional vector $\mathbf{v}$ of the line must be perpendicular to both normal vectors of the planes. Therefore, $\mathbf{v}$ can be computed using the cross product. Let $\mathbf{n_1} = \langle 1, 1, 1\rangle$ and $\mathbf{n_2} = \langle 1, -2, 3\rangle$. Then
$$\begin{aligned}\mathbf{v} = \mathbf{n_1} \times \mathbf{n_2} &= \det \begin{pmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ 1 & 1 & 1 \\ 1 & -2 & 3 \end{pmatrix} \\ &= \langle 5, -2, -3\rangle.\end{aligned}$$
Now the parametric equation of the line can be written as
$$x = 5t + 1$$
$$y = -2t$$
$$z = -3t$$

Remark 2.32. Although this is slightly different from Method 1., by taking $s = -t$, we will get the same equation. In fact, one can see that the symmetric equation we obtained in two methods are differed by a multiplication of -1.

Our last topic in this section is the distance between a point and a line or a plane. Let $Q = (x_0, y_0, z_0)$ be a point in the space and consider a line passes through $P = (x_1, y_1, z_1)$ with a directional vector $\mathbf{v} = \langle v_1, v_2, v_3 \rangle$. Suppose that the angle between $\overrightarrow{PQ}$ and $\mathbf{v}$ is θ. Then ***the distance between the point Q and the line*** is

$$d = |\overrightarrow{PQ}| \sin\theta.$$

Using the magnitude of the cross product, this can be computed as

$$d = \frac{|\overrightarrow{PQ}||\mathbf{v}| \sin\theta}{|\mathbf{v}|} = \frac{|\overrightarrow{PQ} \times \mathbf{v}|}{|\mathbf{v}|}.$$

Consider a plane $ax + by + cz + d = 0$ in the space. Then $\mathbf{n} = \langle a, b, c \rangle$ is a normal vector to the plane. Let $R = (x_2, y_2, z_2)$ be a point on the plane. Then ***the distance between Q and the plane*** is the magnitude of the projection of the vector $\overrightarrow{RQ}$ onto the normal direction $\mathbf{n}$. In other words,

$$d = |\operatorname{proj}_{\mathbf{n}} \overrightarrow{RQ}| = \frac{|\mathbf{n} \cdot \overrightarrow{RQ}|}{|\mathbf{n}|}$$

Note that

$$\mathbf{n} \cdot \overrightarrow{RQ} = a(x_0 - x_2) + b(y_0 - y_2) + c(z_0 - z_2)$$

and

$$ax_2 + by_2 + cz_2 + d = 0.$$

Therefore, in coordinates,

$$d = \frac{|ax_0 + by_0 + cz_0 + d|}{\sqrt{a^2 + b^2 + c^2}}.$$

Example 2.33. Find the distance between the point $(2, 1, 0)$ and $x + y - z = 2$.

Solution. Using the distance formula, note that the plane is $x + y - z - 2 = 0$

$$d = \frac{1 \cdot 2 + 1 \cdot 1 - 1 \cdot 0 - 2}{\sqrt{1 + 1 + 1}} = \frac{1}{\sqrt{3}}.$$

2.4. **Quadric surfaces.** (Ch12.6)

In the previous section, we learned that the linear equation represents planes in the space. In this section we will investigate quadratic equations and see which type of surfaces do these equations represents.

A quadratic equation of three variables in general can be expressed by the following

$$Ax^2 + By^2 + Cz^2 + Dxy + Eyz + Fxz + Gx + Hy + Iz + J = 0.$$

Up to some rotation, translation and symmetry in the space, we may consider the following two types of equations.

$$Ax^2 + By^2 + Cz^2 + J = 0, Ax^2 + By^2 + Iz = 0.$$

Example 2.34. Consider the equation $2x^2 + y^2 - 4x + z = 3$. By completing the square, this can be written as

$$2(x - 2)^2 + y^2 + (z - 7) = 0.$$

In the case when A, B or C vanish, the surface usually is a ***cylinder***. A cylinder is a surface that consists of all lines that are parallel to a given line and pass though a plane curve. Consider the following examples.

Example 2.35. Sketch the surface $z = x^2$.

If we fix $y = k$, the surface $z = x^2$ intersects $y = k$ in a curve which is a parabola with equation $z = x^2$. The surface $z = x^2$ can be obtained by taking this parabola and moving it in the y-axis direction. This surface is called a ***parabolic cylinder***.

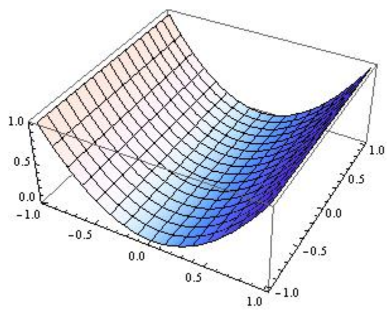


FIGURE 1. $z = x^2$

Example 2.36. Sketch the surface $x^2 + 2y^2 = 1$.

We fix $z = k$ and the surface intersects it in an ellipse $x^2 + 2y^2 = 1$. In this case, the surface can be obtained by moving the ellipse in the z-axis direction. This is an ***elliptic cylinder***.

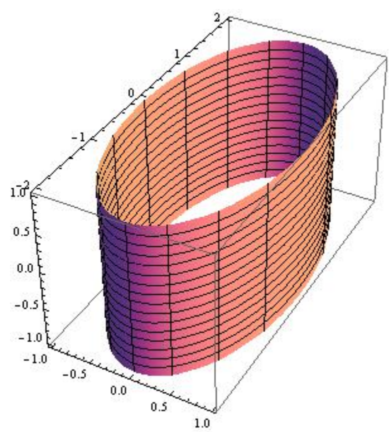


FIGURE 2. $x^2 + 2y^2 = 1$

Now, if A, B, C and I are non-zero, then we call the surface represented by the above equations a ***quadric surface***. Depending on the signs in front of the variables x, y, z, there are six different types of the surfaces. The standard equations are listed as the following.

The surfaces shown in these pictures are symmetric about the z-axis. If a quadric surface is symmetric about a different axis, its equation also changes accordingly.

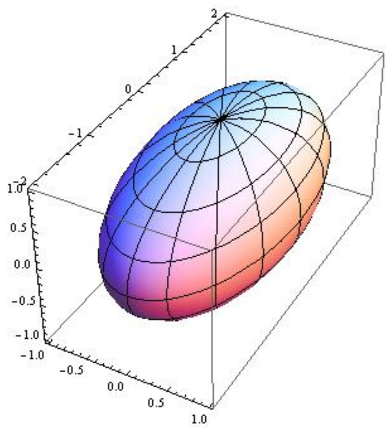


FIGURE 3. Ellipsoid $\frac{x^2}{a^2}+\frac{y^2}{b^2}+\frac{z^2}{c^2}=1$.

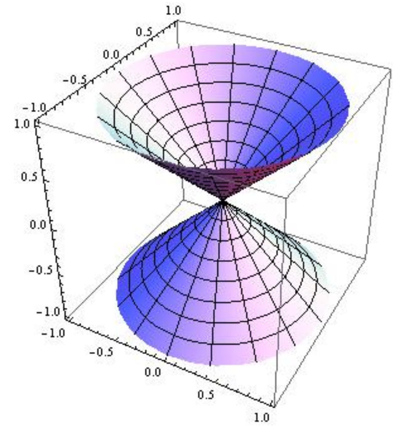


FIGURE 4. Cone $\frac{z^2}{c^2}=\frac{x^2}{a^2}+\frac{y^2}{b^2}$

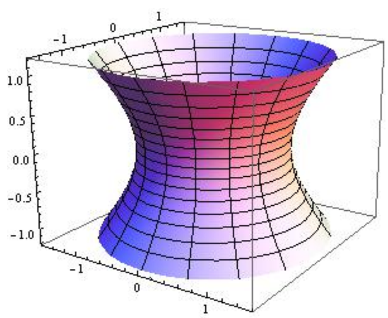


FIGURE 5. Hyperboloid of one sheet $\frac{x^2}{a^2}+\frac{y^2}{b^2}-\frac{z^2}{c^2}=1$

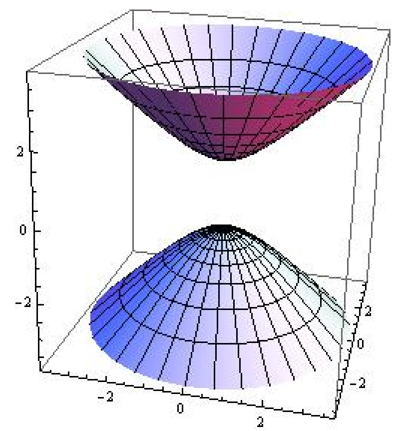


FIGURE 6. Hyperboloid of two sheets $-\frac{x^2}{a^2}-\frac{y^2}{b^2}+\frac{z^2}{c^2}=1$

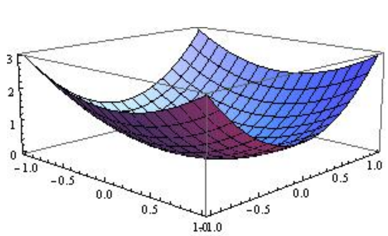


FIGURE 7. Elliptic Paraboloid $\frac{z}{c}=\frac{x^2}{a^2}+\frac{y^2}{b^2}$

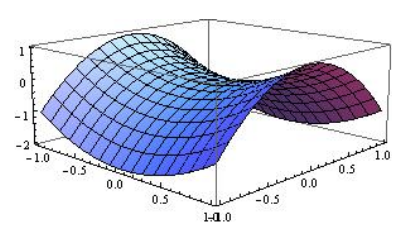


FIGURE 8. Hyperbolic Paraboloid $\frac{z}{c}=\frac{x^2}{a^2}-\frac{y^2}{b^2}$

Example 2.37. Identify the surface $x^2+2z^2-6x-y+8=0$.

Solution. By completing the square, we obtain that

$$y+1=(x-3)^2+2z^2$$

This is an elliptic paraboloid with vertex $(3, -1, 0)$.

3. Vector-valued functions

3.1. Definition. (Ch13.1)

Recall that a function is a relation that assigns to each element in a set, which is called ***domain***, an element in the ***range***. The function $y = f(x)$ we learned before is a function whose domain and range are both the set of real numbers $\mathbb{R}$.

A ***vector-valued function***, or sometimes called ***vector function***, is a function whose domain is the set of real numbers and the range is the set of vectors. This means for each real number t in the domain, there is a corresponding vector $\mathbf{r}(t)$. If this vector has components given by $x(t)$, $y(t)$, and $z(t)$, then we write

$$\mathbf{r}(t) = \langle x(t), y(t), z(t)\rangle.$$

Here, each $x(t), y(t)$ and $z(t)$ are some functions of t. Note that if $x(t), y(t)$ and $z(t)$ has different domains, then the domain of $\mathbf{r}(t)$ is the intersection of the domains of $x(t), y(t)$ and $z(t)$.

Example 3.1. Consider a vector-valued function $\mathbf{r}(t) = \langle 1/t, \log t, \cos t\rangle$. The domain of $1/t$ is $\{t \neq 0\}$, while the domain of $\log t$ is $\{t > 0\}$ and the domain of $\cos t$ is $\mathbb{R}$. Therefore, in this case, the domain of $\mathbf{r}(t)$ is the intersection $\{t > 0\}$.

The ***limit*** of a vector function is defined by taking the limits of its component functions.

Definition 3.2. If $\mathbf{r}(t) = \langle x(t), y(t), z(t)\rangle$, then we define

$$\lim_{t\to a} \mathbf{r}(t) = \langle \lim_{t\to a} x(t), \lim_{t\to a} y(t), \lim_{t\to a} z(t)\rangle.$$

Definition 3.3. We say that a vector function $\mathbf{r}(t)$ is continuous at $t = a$, if

$$\lim_{t\to a} \mathbf{r}(t) = \mathbf{r}(a).$$

Note that by definition, a vector function $\mathbf{r}(t)$ is continuous at $t = a$ if and only if all of its component functions $x(t), y(t)$ and $z(t)$ are continuous at $t = a$.

Example 3.4. Consider a vector-valued function $\mathbf{r}(t) = \langle \cos t, \sin t, t\rangle$. Clearly, at any $a \in \mathbb{R}$, the limit $\lim_{t\to a} \mathbf{r}(t) = \mathbf{r}(a)$. Therefore, this function is continuous at any $a \in \mathbb{R}$.

Vector-valued functions are closely related to space curves. In previous sections, we have seen identification between points in $\mathbb{R}^3$ with vectors whose initial point in the origin. Under this identification, a vector valued function can be used to represent a curve in $\mathbb{R}^3$. Suppose that $\mathbf{r}(t) = \langle x(t), y(t), z(t)\rangle$ is a vector valued function. Consider the curve traced out by the terminal points of the function. We call the components

$$x = x(t), y = y(t), z = z(t)$$

the ***parametric equation*** of the curve.

Example 3.5. The vector-valued function $\mathbf{r}(t) = \langle \cos t, \sin t, t\rangle$ in the previous example represents a curve in the space which is called a helix.

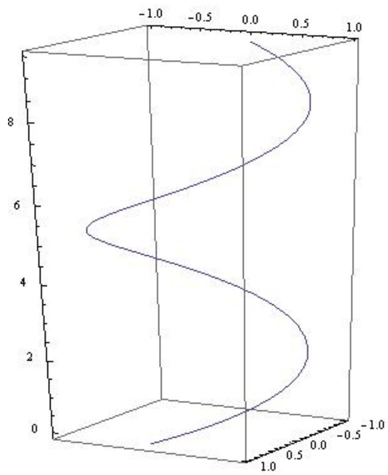


FIGURE 9. A helix.

Example 3.6. Find a vector function that represents the intersection of the cylinder $x^2 + y^2 = 1$ and the plane $y + z = 2$.

Solution. The points on the curve of intersection satisfies both equations of the cylinder and the plane. In other words, if

$$x = x(t), y = y(t), z = z(t)$$

is the parametric equation of the curve, then

$$x^2(t) + y^2(t) = 1, y(t) + z(t) = 1.$$

We only have to choose equations satisfies these relations. One possible choice would be the following

$$x(t) = \cos t, y(t) = \sin t, z(t) = 1 - \sin t, 0 \le t \le 2\pi.$$

Therefore, the curve of intersection can be represented by $\mathbf{r}(t) = \langle \cos t, \sin t, 1 - \sin t \rangle$.

Remark 3.7. In this example, one can see that the vector functions represent the same curve in the space is not unique. For example, we may simply pick $\langle \sin 2t, \cos 2t, 1 - \cos 2t \rangle$ and this still represents the curve in this example. The choice of a vector-valued function to represent a curve in the space is called a ***parametrization*** of the curve. Now what kind of parametrization of the curve is desired, let us study the following properties of the vector-valued function.

3.2. **Derivatives and integrals of vector functions.** (Ch13.1,13.2)

Since we have introduced the limit and continuity of vector-valued functions as component-wise definitions. We may also study the derivatives. Let $\mathbf{r}(t) = \langle x(t), y(t), z(t) \rangle$ be a vector function.

Definition 3.8. We define the derivative $\mathbf{r}'(t)$ of the function $\mathbf{r}(t)$ to be

$$\frac{d\mathbf{r}}{dt} = \mathbf{r}'(t) = \lim_{\Delta t \to 0} \frac{\mathbf{r}(t + \Delta t) - \mathbf{r}(t)}{\Delta t},$$

if the limit exists. In this case, we say $\mathbf{r}(t)$ is differentiable at t.

The geometric interpretation of this definition is the following. Let P, Q be the terminal points of $\mathbf{r}(t)$ and $\mathbf{r}(t+\Delta t)$. Then the vector $\overrightarrow{PQ}$ is $\mathbf{r}(t+\Delta t)-\mathbf{r}(t)$. The expression $\frac{1}{\Delta t}(\mathbf{r}(t+\Delta t)-\mathbf{r}(t))$ is a vector in the same direction of $\overrightarrow{PQ}$ but scaled by $\frac{1}{\Delta t}$ so that it does not vanish when we take the limit. When we take the limit $\Delta t \to 0$, the vector $\overrightarrow{PQ}$ becomes a tangent vector to the curve at P. Therefore we call $\mathbf{r}'(t)$ a ***tangent vector*** at t. The line passes through P in the direction of $\mathbf{r}'(t)$ is called a ***tangent line***.

Also note that because of the component-wise definition of the limit, the derivative, if exists, can be computed component-wise as well.

$$\mathbf{r}'(t) = \langle x'(t), y'(t), z'(t)\rangle.$$

Example 3.9. Consider the curve $\mathbf{r}(t) = \langle \cos t, \sin t, 1-\sin t\rangle$ in the previous example. Find the derivative of this function and write down the equation of the tangent line at the point $(1,0,1)$.

Solution. The derivative of the function can be computed component-wise,

$$\mathbf{r}'(t) = \langle -\sin t, \cos t, -\cos t\rangle.$$

Now, to find the equation of the tangent line at the point $(1,0,1)$, we must first figure out for which t does the vector represents the point $(1,0,1)$. Note that at $(1,0,1)$

$$\cos t = 1 \Rightarrow t = 0.$$

Therefore, a tangent vector at $t=0$ is

$$\mathbf{r}'(0) = \langle 0, 1, -1\rangle.$$

The tangent line, by definition, is a line that passes through $(1,0,1)$ in the direction of $\mathbf{r}'(0)$. Therefore, the parametric equation of the tangent line is

$$x = 1, y = t, z = 1-t.$$

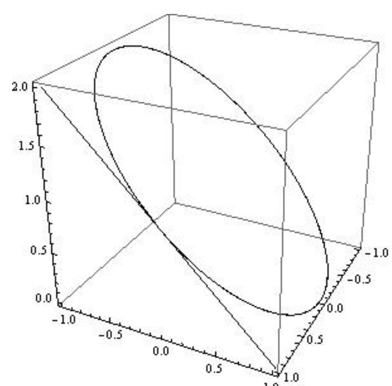


Figure 10. The curve and its tangent line.

The derivative of vector function satisfies the following differentiation rules.

Proposition 3.10. *Let* $\mathbf{u}$ *and* $\mathbf{v}$ *be two differentiable vector valued functions,* c *a scalar and* $f(t)$ *a function. Then*

- $\frac{d}{dt}(\mathbf{u}(t)+\mathbf{v}(t)) = \mathbf{u}'(t)+\mathbf{v}'(t)$,
  $\frac{d}{dt}(c\mathbf{u}(t)) = c\mathbf{u}'(t)$. *(Linearity)*
- $\frac{d}{dt}(f(t)\mathbf{u}(t)) = f'(t)\mathbf{u}(t) + f(t)\mathbf{u}'(t)$. *(Product rule)*
- $\frac{d}{dt}(\mathbf{u}(t)\cdot\mathbf{v}(t)) = \mathbf{u}'(t)\cdot\mathbf{v}(t) + \mathbf{u}(t)\cdot\mathbf{v}'(t)$. *(Product rule for dot product)*
- $\frac{d}{dt}(\mathbf{u}(t)\times\mathbf{v}(t)) = \mathbf{u}'(t)\times\mathbf{v}(t) + \mathbf{u}(t)\times\mathbf{v}'(t)$. *(Product rule for cross product)*
- $\frac{d}{dt}(\mathbf{u}(f(t))) = \mathbf{u}'(f)f'(t)$. *(Chain rule)*

Because of the component-wise definition, these properties can by proved by component-wise computation. Using these differentiation rules, we may immediately obtain the following result.

Example 3.11. Let $\mathbf{r}$ be a differentiable vector-valued function with constant magnitude. Then $\mathbf{r} \perp \mathbf{r}'$.

Solution. The magnitude square of the function $\mathbf{r}$ is

$$|\mathbf{r}|^2 = \mathbf{r}\cdot\mathbf{r},$$

which is also a constant. Therefore, by taking the derivatives of both sides, we obtain that

$$0 = \mathbf{r}'\cdot\mathbf{r} + \mathbf{r}\cdot\mathbf{r}'.$$

Now, by the symmetry of the dot product,

$$\mathbf{r}\cdot\mathbf{r}' = 0 \Rightarrow \mathbf{r}\perp\mathbf{r}'.$$

For vector-valued function, one may define ***higher derivatives*** recursively. For example, the ***second derivative*** $\mathbf{r}''(t)$ is the derivative of $\mathbf{r}'(t)$.

Example 3.12. Let $\mathbf{u}(t) = \mathbf{r}(t)\cdot(\mathbf{r}'(t)\times\mathbf{r}''(t))$. Find $\mathbf{u}'(t)$.

Solution. We compute the derivative using the differentiation rule.

$$\begin{aligned}
\mathbf{u}'(t) &= \frac{d}{dt}(\mathbf{r}(t)\cdot(\mathbf{r}'(t)\times\mathbf{r}''(t))) \\
&= \mathbf{r}'(t)\cdot(\mathbf{r}'(t)\times\mathbf{r}''(t)) + \mathbf{r}(t)\cdot\frac{d}{dt}(\mathbf{r}'(t)\times\mathbf{r}''(t)) \\
&= 0 + \mathbf{r}(t)\cdot\frac{d}{dt}(\mathbf{r}'(t)\times\mathbf{r}''(t)) \\
&= \mathbf{r}(t)\cdot(\mathbf{r}''(t)\times\mathbf{r}''(t)) + \mathbf{r}(t)\cdot(\mathbf{r}'(t)\times\mathbf{r}'''(t)) \\
&= \mathbf{r}(t)\cdot(\mathbf{r}'(t)\times\mathbf{r}'''(t)).
\end{aligned}$$

The ***definite and indefinite integral*** of vector-valued functions are defined in the same way as the derivatives. We define the definite and indefinite integral to be component-wise integrals.

Definition 3.13. Let $\mathbf{r}(t) = \langle x(t), y(t), z(t)\rangle$ be a continuous vector valued function, $a \le t \le b$. The definite integral

$$\int_a^b \mathbf{r}(t)dt = \langle \int_a^b x(t)dt, \int_a^b y(t)dt, \int_a^b z(t)dt\rangle.$$

And the indefinite integral

$$\int \mathbf{r}(t)dt = \langle \int x(t)dt, \int y(t)dt, \int z(t)dt\rangle.$$

Note the for the indefinite integral, suppose that $X(t), Y(t), Z(t)$ are anti-derivatives of $x(t), y(t), z(t)$ respectively. Then

$$\int x(t)dt = X(t) + c_1, \int y(t)dt = Y(t) + c_2, \int z(t)dt = Z(t) + c_3.$$

The c_1, c_2, c_3 in each integral may be different. If we let $\mathbf{c} = \langle c_1, c_2, c_3\rangle$, then the indefinite integral can be denoted by

$$\int \mathbf{r}(t)dt = \langle X(t), Y(t), Z(t)\rangle + \mathbf{c}.$$

Example 3.14. Find the indefinite integral of $\mathbf{r}(t) = \langle \cos t, t^2, e^t\rangle$.

Solution. We compute directly.

$$\int \mathbf{r}(t)dt = \langle \sin t, \frac{1}{3}t^3, e^t\rangle + \mathbf{c}.$$

In the previous section, we have discussed a curve can be represented by a vector-valued function $\mathbf{r}(t)$. Note that as a geometric object, the smoothness of a curve is different from the smoothness of a vector-valued function.

Let $\mathbf{r}(t)$ be a vector-valued function. We say that $\mathbf{r}(t)$ is smooth, if, as a function, it is infinitely differentiable. This coincides with the definition of the smoothness for scalar functions.

However, when we talk about the smoothness of the curves, we would like to avoid the possibility of have sharp angles in the curves. This in particular requires the tangent vector $\mathbf{r}'(t)$ to be non-vanishing. On the other hand, we have seen in the previous examples that the tangent vector really depends on the choice of the parametrization. Therefore, we introduce the following definition.

Definition 3.15. Given a curve in the space and a parametrization $\mathbf{r}(t)$ of the curve. We say that $\mathbf{r}(t)$ is a ***regular parametrization*** if its tangent vector $\mathbf{r}'(t) \neq 0$ for any t.

Definition 3.16. A curve is said to be a ***smooth curve***, if it admits a regular parametrization $\mathbf{r}(t)$ and this parametrization $\mathbf{r}(t)$ is smooth as a vector function.

Remark 3.17. Note that a smooth curve also admits a non-regular parametrization. Consider the following example.

Example 3.18. Let us consider a straight line which is parameterized by the equation $\mathbf{r}(t) = \langle t, t, t\rangle$. Clearly, this is a regular smooth parametrization since $\mathbf{r}'(t) = \langle 1, 1, 1\rangle$ will never be zero and its higher derivatives exist and all vanish. This suggests that the line is a smooth curve in the sense of the above definition.

On the other hand, consider a parametrization of the line using the function $\mathbf{u}(t) = \langle t^3, t^3, t^3\rangle$. This is still a parametrization of the line. However, it is not regular, since $\mathbf{u}'(t) = \langle 3t^2, 3t^2, 3t^2\rangle$ which vanishes at $t = 0$.

Food for thought. The reason why we require the tangent vector of a smooth curve is non-vanishing can be seen in the following example. Consider two functions

$$f(t) = \begin{cases} e^{-1/t^2} & t > 0 \\ 0 & t = 0 \\ -(e^{-1/t^2}) & t < 0 \end{cases}$$

$$g(t) = \begin{cases} e^{-1/t^2} & t \neq 0 \\ 0 & t = 0 \end{cases}$$

One may check that both these two functions are smooth at any value of t.

Now consider a curve which is represented by vector-valued function $\mathbf{r}(t) = \langle f(t), g(t), g(t) \rangle$. This is in fact a line which suddenly changes direction at $t = 0$ even though the vector-valued function $\mathbf{r}(t)$ is smooth. In fact, though these examples, we see that a parametrization is not only a collection of the vectors that trace out a curve in the space, but also it tells us how fast this tracing is. To precisely describe this phenomenon, let us study the arc length of a curve.

3.3. **Arc length and reparametrization.** (Ch13.3)

The length of a curve in the space can be computed as the limit of the lengths of its inscribed polygons. Suppose that the curve is represented by $\mathbf{r}(t)$, for $a \leq t \leq b$. If we subdivide the interval $[a, b]$ into small pieces $[t_i, t_{i+1}]$, then the length of a secant line is in fact the magnitude $|\mathbf{r}(t_{i+1}) - \mathbf{r}(t_i)|$. If we choose this subdivision even so that $\Delta t = t_{i+1} - t_i$ is a constant for every i, then the arc length is given by the limit of the following sum

$$\lim_{\Delta t \to 0} \sum_i |\mathbf{r}(t_{i+1}) - \mathbf{r}(t_i)| = \lim_{\Delta t \to 0} \sum_i \frac{|\mathbf{r}(t_{i+1}) - \mathbf{r}(t_i)|}{\Delta t} \cdot \Delta t$$

Now recall that by the definition of the Riemann sum of a definite integral, this is

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

Theorem 3.19. *Let $\mathbf{r}(t)$, $a \leq t \leq b$ be a regular parametrization of a curve. Then the* ***arc length*** *of the curve between $t = a$ and $t = b$ is the definite integral*

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

In particular, if $\mathbf{r}(t) = \langle x(t), y(t), z(t) \rangle$, then

$$\int_a^b |\mathbf{r}'(t)| dt = \int_a^b \sqrt{(x'(t))^2 + (y'(t))^2 + (z'(t))^2} dt.$$

Example 3.20. Consider a helix $\mathbf{r}(t) = \langle \cos t, \sin t, t \rangle$, $0 \leq t \leq 2\pi$. The arc length of this helix between $t = 0$ and 2π can be computed as the following.

$$\mathbf{r}'(t) = \langle -\sin t, \cos t, 1 \rangle$$

$$|\mathbf{r}'(t)| = \sqrt{\sin^2 t + \cos^2 t + 1} = \sqrt{2}.$$

$$\text{Arc length} = \int_0^{2\pi} \sqrt{2} dt = 2\sqrt{2}\pi.$$

Example 3.21. We have seen before that a curve can be represented by different vector functions. Let us consider two different parametrization of the same curve and compare the result of the arc length.

Let $\mathbf{r_1}(t) = \langle 2t, t^2, \frac{1}{3}t^3 \rangle$, $1 \leq t \leq 2$. Then

$$\mathbf{r_1}'(t) = \langle 2, 2t, t^2 \rangle$$

$$|\mathbf{r_1}'(t)| = \sqrt{4+4t^2+t^4} = t^2+2.$$

And

$$\text{Arc length} = \int_1^2 (t^2+2)dt = \frac{1}{3}t^3 + 2t\bigg|_{t=1}^{t=2} = \frac{7}{3}+2.$$

Let $\mathbf{r_2}(u) = \langle 2e^u, e^{2u}, \frac{1}{3}e^{3u}\rangle$, $0 \le u \le \ln 2$. Then

$$\mathbf{r_2}'(u) = \langle 2e^u, 2e^{2u}, e^{3u}\rangle$$

$$|\mathbf{r_1}'(t)| = \sqrt{4e^{2u}+4e^{4u}+e^{6u}} = 2e^u + e^{3u}.$$

And

$$\text{Arc length} = \int_0^{\ln 2} 2e^u + e^{3u}du = 2e^u + \frac{1}{3}e^{3u}\bigg|_{u=0}^{u=\ln 2} = \frac{7}{3}+2,$$

which gives us the same result.

Note that in these two different parametrization, the tangent vectors are very different. The x-component of the $\mathbf{r_1}'(t)$ is a constant 2 while for $\mathbf{r_2}'(u)$ it is always changing according to u. In fact, the variables in this example are related by $t = e^u$. If we perform a change of variable to the first integral of the arc length, then

$$dt = e^u du,$$

$$\int_1^2 (t^2+2)dt = \int_0^{\ln 2} (e^{2u}+2)e^u du,$$

which is precisely the second integral.

In fact the arc length is an invariant under the choice of the parametrization. Let $\mathbf{r}(t)$ be a parametrization of a curve in the space. We say that $\tilde{\mathbf{r}}(u)$ is a ***reparametrization***, if there is a smooth invertible function $u = \phi(t)$ with a smooth inverse $t = \phi^{-1}(u)$ such that

$$\tilde{\mathbf{r}}(u) = \mathbf{r}(\phi(u)).$$

Example 3.22. The change of variable $t = e^u, 0 \le u \le \ln 2$ is a smooth function with a smooth inverse $u = \ln t, 1 \le t \le 2$ in Example 3.21. Therefore $\mathbf{r_2}(u)$ is a reparametrization of $\mathbf{r_1}(t)$.

Example 3.23. In Example 3.18, there are two different parametrizations of the line $\mathbf{r_1}(t) = \langle t,t,t\rangle$, $\mathbf{r_2}(u) = \langle u^3, u^3, u^3\rangle$. Note that the function $t = u^3$ is smooth and invertible, but its inverse

$$u = \sqrt[3]{t}$$

is not smooth.

In fact, its derivative is not defined at 0. Therefore, $\mathbf{r_2}(u)$ is not a reparametrization of $\mathbf{r_1}(t)$. Note that $\mathbf{r_2}(u)$ is not regular.

These examples suggests that regularity and arc length are both preserved by reparametrization. Indeed, we have

Theorem 3.24. *Regularity is preserved under reparametrization.*

Proof. Let $\mathbf{r}(t)$ be a regular parametrization and $\tilde{\mathbf{r}}(u)$ a reparametrization. Suppose that $t = t(u)$. Then

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

By the chain rule,

$$\tilde{\mathbf{r}}'(u) = \mathbf{r}'(t)\frac{dt}{du}.$$

Since $\mathbf{r}(t)$ is regular, $\mathbf{r}'(t) \neq 0$. Now since $\tilde{\mathbf{r}}(u)$ is a reparametrization, $t = t(u)$ has a smooth inverse. Therefore, by the inverse function theorem,

$$\frac{dt}{du} \neq 0.$$

□

Theorem 3.25. *Arc length is preserved under reparametrization.*

Proof. If $\mathbf{r}(t)$, $a \leq t \leq b$ is a regular parametrization and $\tilde{\mathbf{r}}(u)$ is a reparametrization with $t = t(u)$. Since $\frac{dt}{du} \neq 0$, and it is continuous, we have either $\frac{dt}{du} > 0$ or $\frac{dt}{du} < 0$. If $\frac{dt}{du} > 0$, Consider the arc length under $\mathbf{r}(t)$ parametrization.

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

and a change of variable $t = t(u)$. Then $dt = \frac{dt}{du}du$.

$$\int_a^b |\mathbf{r}'(t)|dt = \int_{u(a)}^{u(b)} |\mathbf{r}'(u)|\frac{dt}{du}du$$

On the other hand, since

$$\tilde{\mathbf{r}}'(u) = \mathbf{r}'(t)\frac{dt}{du},$$

we obtain that

$$|\tilde{\mathbf{r}}'(u)| = |\mathbf{r}'(t)| \cdot |\frac{dt}{du}|,$$

and

$$\int_{u(a)}^{u(b)} |\mathbf{r}'(u)|\frac{dt}{du}du = \int_{u(a)}^{u(b)} |\tilde{\mathbf{r}}'(u)|du,$$

which is the arc length in $\tilde{\mathbf{r}}(u)$ parametrization.

If $\frac{dt}{du} < 0$, then

$$\int_a^b |\mathbf{r}'(t)|dt = \int_{u(b)}^{u(a)} |\mathbf{r}'(u)|(-\frac{dt}{du})du,$$

which will give the same result. □

The fact that the arc length is an invariant allows us to perhaps use the arc length itself as a parametrization of the curve. To define this parametrization, we can define a reparametrization using the ***arc length function***.

Let $\mathbf{r}(t)$, $a \leq t \leq b$ be a regular parametrization. Consider a function $s(t)$ defined with a changing variable t is the integral

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

Here, to avoid repeatedly using the variable t, we denote the variable in the integral by u. Note that u is no longer a variable after we integrated the function. This function is called the arc length function of a curve. And in fact, using the fundamental theorem of calculus,

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

If $\mathbf{r}(t)$ is a regular parametrization, $\frac{ds}{dt} = |\mathbf{r}'(t)| > 0$, which implies that $s(t)$ is smooth increasing and with a smooth inverse $t = t(s)$. Therefore, we may consider a reparametrization

$$\tilde{\mathbf{r}}(s) = \mathbf{r}(t(s)).$$

This is called a ***reparametrization by the arc length function***.

Proposition 3.26. *If $\tilde{\mathbf{r}}(s) = \mathbf{r}(t(s))$ is a reparametrization by the arc length function $s(t)$, then*

$$|\tilde{\mathbf{r}}'(s)| = 1.$$

Proof. By the chain rule,

$$\frac{d\tilde{\mathbf{r}}}{ds} = \frac{d\mathbf{r}}{dt}\frac{dt}{ds},$$

or

$$\frac{d\tilde{\mathbf{r}}}{ds}\frac{ds}{dt} = \frac{d\mathbf{r}}{dt}.$$

Note that

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

taking the magnitude of both sides we obtain that

$$|\frac{d\tilde{\mathbf{r}}}{ds}| \cdot |\frac{ds}{dt}| = |\frac{d\mathbf{r}}{dt}| \Rightarrow |\frac{d\tilde{\mathbf{r}}}{ds}| = 1.$$

□

Example 3.27. Recall that in Example 3.20, a helix $\mathbf{r}(t) = \langle \cos t, \sin t, t\rangle$, $0 \le t \le 2\pi$. We have $|\mathbf{r}'(t)| = \sqrt{2}$. Therefore the arc length function

$$s(t) = \int_0^t \sqrt{2}du = \sqrt{2}t.$$

The inverse of this function is given by

$$t = \frac{1}{\sqrt{2}}s.$$

Therefore, a reparametrization of the helix by the arc length function is given by

$$\tilde{\mathbf{r}}(s) = \langle \cos(\frac{1}{\sqrt{2}}s), \sin(\frac{1}{\sqrt{2}}s), \frac{1}{\sqrt{2}}s\rangle.$$

We may check that in this case,

$$\tilde{\mathbf{r}}'(s) = \langle -\frac{1}{\sqrt{2}}\sin(\frac{1}{\sqrt{2}}s), \frac{1}{\sqrt{2}}\cos(\frac{1}{\sqrt{2}}s), \frac{1}{\sqrt{2}}\rangle,$$

and

$$|\tilde{\mathbf{r}}'(s)| = \sqrt{\frac{1}{2}+\frac{1}{2}} = 1.$$

Reparametrization using the arc-length function allows us to simplify the computation in many cases because of Proposition 3.26 and Example 3.11. We shall see this in the following section about curvature.

3.4. **Curvature of a plane curve.** The curvature of a curve is a quantity which measures how far away a curve from being a straight line. Intuitively, the curvature should be a geometric quantity which does not depend on the parametrization of the curve. We will start with a plane curve.

Consider a curve $\mathbf{r}(t) = \langle x(t), y(t)\rangle$ in a plane. At $t = t_0$, for some fixed t_0, the "best" straight line which approximates the curve should be its tangent line $\mathbf{l}(t)$, whose directional vector is $\mathbf{r}'(t_0)$.

For a nearby point $\mathbf{r}(t_0 + \Delta t)$, the distance between $\mathbf{r}(t_0 + \Delta t)$ and $\mathbf{l}(t_0 + \Delta t)$ indicates how far away this curve is from being a straight line. Now if $\mathbf{n}$ is a unit normal vector which is perpendicular to the tangent vector $\mathbf{r}'(t_0)$, then this distance is

$$|\operatorname{proj}_{\mathbf{n}}(\mathbf{r}(t_0 + \Delta t) - \mathbf{r}(t_0))| = |(\mathbf{r}(t_0 + \Delta t) - \mathbf{r}(t_0)) \cdot \mathbf{n}|.$$

Using the Taylor expansion of the function $\mathbf{r}$ at t_0, we have that

$$\mathbf{r}(t_0 + \Delta t) = \mathbf{r}(t_0) + \mathbf{r}'(t_0)\Delta t + \frac{1}{2}\mathbf{r}''(t_0)(\Delta t)^2 + \varepsilon,$$

where $\varepsilon/(\Delta t)^2 \to 0$ when $\Delta t \to 0$.

In order to define a curvature which is independent of the parametrization, we may first choose the arc-length function as the parameter. In other words, suppose that the curve is parameterized by the arc-length function s and the above expression becomes

$$\mathbf{r}(s_0 + \Delta s) = \mathbf{r}(s_0) + \mathbf{r}'(s_0)\Delta s + \frac{1}{2}\mathbf{r}''(s_0)(\Delta s)^2 + \varepsilon,$$

for some s_0. Now by Proposition 3.26, $|\mathbf{r}'(s)| = 1$ is a constant. Therefore, by Example 3.11, $\mathbf{r}' \perp \mathbf{r}''$, and in particular, $\mathbf{r}''$ is parallel to $\mathbf{n}$.

Now if we rewrite the above equation by taking the dot product with $\mathbf{n}$, we obtain that

$$(\mathbf{r}(s_0 + \Delta s) - \mathbf{r}(s_0)) \cdot \mathbf{n} = \mathbf{r}'(s_0) \cdot \mathbf{n}\Delta s + \frac{1}{2}\mathbf{r}''(s_0) \cdot \mathbf{n}(\Delta s)^2 + \varepsilon$$

Note that the first term on the right-hand side of the equality is zero. And if $\mathbf{r}''(s_0) = \kappa\mathbf{n}$, for some constant $\kappa \geq 0$ by choosing the direction of $\mathbf{n}$,

$$|(\mathbf{r}(s_0 + \Delta s) - \mathbf{r}(s_0)) \cdot \mathbf{n}| = \frac{1}{2}\kappa(\Delta s)^2 + \varepsilon$$

If we normalize this dot product by $\frac{1}{2}(\Delta s)^2$ and let $\Delta s \to 0$, we obtain that this distance is characterized by the quantity

$$\kappa = |\mathbf{r}''(s_0)|,$$

which is called the curvature of the curve $\mathbf{r}(s)$ at s_0.

Definition 3.28. Let $\mathbf{r}(s)$ be a curve parameterized by the arc length function s. We define the curvature of the curve to be

$$\kappa = |\mathbf{r}''(s)|$$

the magnitude of the second derivative.

From the above discussion, it is clear that κ characterizes the tendency for a curve leaving its tangent line. However, the arc length parametrization is usually difficult to compute and it is not entirely clear that κ is independent of the parametrization. For this reason, let us also compute the expression of the curvature under different parametrizations.

Remark 3.29. In geometry, we say that a quantity is geometric, if it does not depend on the parametrization of the object. In fact, we have seen that from Theorem 3.25, the arc length of a curve does not depend on the parametrization and therefore, it is geometric. Integral is good way to extract geometric invariants from the parametrization. We will see below that curvature is also an example of a geometric quantity.

In Definition 3.28, we see that the curvature is the magnitude of the vector $\mathbf{r}''(s)$. On the other hand, $\mathbf{r}''(s) = \frac{d}{ds}\mathbf{r}'(s)$, where $\mathbf{r}'(s)$ is the tangent vector to the curve $\mathbf{r}(s)$ with magnitude 1. Therefore, if the curve has a reparametrization by some generic parameter t,

$$\kappa = \left| \frac{d}{ds}\mathbf{T} \right|,$$

where $\mathbf{T} = \frac{\mathbf{r}'(t)}{|\mathbf{r}'(t)|}$ is the unit tangent vector of the curve and s is the arc length parameter. Note that $\frac{ds}{dt} = |\mathbf{r}'(t)|$, by the chain rule,

$$\kappa = \left| \frac{d\mathbf{T}/dt}{ds/dt} \right| = \frac{|\mathbf{T}'(t)|}{|\mathbf{r}'(t)|}.$$

This formula gives us a way to compute the curvature in any parametrization t.

Example 3.30. Consider a circle on the plane of radius a. Suppose that the circle is parameterized by the function $\mathbf{r}(t) = \langle a\cos t, a\sin t\rangle$. Then

$$\mathbf{r}'(t) = \langle -a\sin t, a\cos t\rangle, |\mathbf{r}'(t)| = a.$$

And the unit tangent vector

$$\mathbf{T}(t) = \frac{\mathbf{r}'(t)}{|\mathbf{r}'(t)|} = \langle -\sin t, \cos t\rangle.$$

So

$$\mathbf{T}'(t) = \langle -\cos t, -\sin t\rangle \text{ and } |\mathbf{T}'| = 1.$$

Therefore,

$$\kappa = \frac{|\mathbf{T}'|}{|\mathbf{r}'|} = \frac{1}{a}.$$

The curvature of a circle of radius a is a constant $\frac{1}{a}$.

Remark 3.31. This example shows that the curvature of a circle of radius is a constant 1/radius. Note that for any curve, if the curvature of a curve at some point is k, then it must be tangent to a circle of radius $1/k$ at this point. This circle is called the circle of curvature.

Let us now try to obtain an equation of the curvature in terms of $\mathbf{r}(t)$. To do this, we still consider a plane curve, but the curve now is considered to be in the XY-plane in a three dimensional space (so that we can use the cross product).

Theorem 3.32. *The curvature of the curve given by* $\mathbf{r}(t)$ *is*

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

Proof. We first rewrite the equation $\mathbf{T} = \mathbf{r}'/|\mathbf{r}'|$ and using $|\mathbf{r}'| = \frac{ds}{dt}$,

$$\mathbf{r}' = |\mathbf{r}'|\mathbf{T} = \frac{ds}{dt}\mathbf{T}.$$

By taking the derivative and using the product rule,

$$\mathbf{r}'' = \frac{d^2s}{dt^2}\mathbf{T} + \frac{ds}{dt}\mathbf{r}.$$

Note that because $\mathbf{T} \times \mathbf{T} = \mathbf{0}$,

$$\mathbf{r}' \times \mathbf{r}'' = \left(\frac{ds}{dt}\right)^2 \mathbf{T} \times \mathbf{T}'.$$

Now because $|\mathbf{T}| = 1$, by Example 3.11, $\mathbf{T} \perp \mathbf{T}'$ and

$$|\mathbf{T} \times \mathbf{T}'| = |\mathbf{T}| \cdot |\mathbf{T}'| = |\mathbf{T}'|,$$

and therefore,

$$|\mathbf{r}' \times \mathbf{r}''| = \left(\frac{ds}{dt}\right)^2 |\mathbf{T}'| = |\mathbf{r}'|^2|\mathbf{T}'|.$$

$$\kappa = \frac{|\mathbf{T}'|}{|\mathbf{r}'|} = \frac{|\mathbf{r}'(t) \times \mathbf{r}''(t)|}{|\mathbf{r}'(t)|^3}.$$

□

From this expression, we may also obtain that

Theorem 3.33. κ *is geometric. In other words,* κ *is independent of the choice of the parametrization.*

Proof. Let $\mathbf{r}(t)$ be a parametrization of a curve and $\tilde{\mathbf{r}}(u)$ a reparametrization with $u = u(t)$. Then $\mathbf{r}(t) = \tilde{\mathbf{r}}(u(t))$. Now by the chain rule,

$$\mathbf{r}'(t) = \tilde{\mathbf{r}}'u'.$$

And

$$\mathbf{r}''(t) = \tilde{\mathbf{r}}''u'^2 + \mathbf{r}'u''.$$

Therefore, $\mathbf{r}' \times \mathbf{r}'' = u'^3\tilde{\mathbf{r}}' \times \tilde{\mathbf{r}}''$, and

$$\kappa = \frac{|\mathbf{r}'(t) \times \mathbf{r}''(t)|}{|\mathbf{r}'(t)|^3} = \frac{|\tilde{\mathbf{r}}'(u) \times \tilde{\mathbf{r}}''(u)|}{|\tilde{\mathbf{r}}'(u)|^3}.$$

□

Example 3.34. Suppose that a plane curve is given by the graph of a function $y = f(x)$. We may choose x as a parameter and the curve can be written by

$$\mathbf{r}(t) = \langle x, f(x), 0\rangle.$$

Now, using the expression in Theorem 3.32, we obtain that

$$\kappa = \frac{|f''(x)|}{(\sqrt{1+f'(x)^2})^3}.$$

Example 3.35. Find the curvature of the curve $y = x^2$.

Solution. Using the above equation, we obtain that

$$\kappa = \frac{2}{(1+4x^2)^{3/2}}.$$

3.5. **Curvature, torsion and TNB frame.** So far we have been discussing the curvature of a plane curve. A natural question is can we generalize this to the case of space curves? In fact, the expression of the curvature in Theorem 3.32 is already using the cross product in a 3-dimensional space. The main difference between a curve in the 3-dimensional space and a plane is that in the three dimensional space, the direction of the vector that is perpendicular to the tangent vector is not unique. Therefore, in order to discuss vectors in 3-dimensional space, let us first introduce the TNB-frame of a curve.

Consider a smooth curve $\mathbf{r}(t)$, at each fixed t, there are many vectors that are perpendicular to the unit tangent vector $\mathbf{T}$. Because $|\mathbf{T}| = 1$, by Example 3.11, $\mathbf{T} \perp \mathbf{T}'$. We define the unit vector in the direction of $\mathbf{T}'$ to be the unit normal vector of the curve and denote by

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

Now, given two unit and orthogonal vectors $\mathbf{T}$ and $\mathbf{N}$, we may form a third vector which is unit and orthogonal to both these two vectors using the cross product. We define

$$\mathbf{B} = \mathbf{T} \times \mathbf{N}$$

to be the unit binormal vector.

Definition 3.36. Given a smooth curve with parametrization $\mathbf{r}(t)$ in the space, for each t, the unit vectors $\mathbf{T}(t)$, $\mathbf{N}(t)$ and $\mathbf{B}(t)$ is called the TNB-frame (or Frenet frame) of the curve.

Example 3.37. Find out the TNB-frame of the helix $\mathbf{r}(t) = \langle \cos t, \sin t, t\rangle$.

Solution. We first compute the unit tangent vector.

$$\mathbf{r}' = \langle -\sin t, \cos t, 1\rangle, |\mathbf{r}'| = \sqrt{2}.$$

Therefore,

$$\mathbf{T}(t) = \frac{1}{\sqrt{2}}\langle -\sin t, \cos t, 1\rangle.$$

Now

$$\mathbf{T}'(t) = \frac{1}{\sqrt{2}}\langle -\cos t, -\sin t, 0\rangle, |\mathbf{T}'| = \frac{1}{\sqrt{2}},$$

and

$$\mathbf{N}(t) = \frac{\mathbf{T}'}{|\mathbf{T}'|} = \langle -\cos t, \sin t, 0\rangle.$$

And finally,

$$\mathbf{B}(t) = \mathbf{T}(t) \times \mathbf{N}(t) = \frac{1}{\sqrt{2}}\langle \sin t, -\cos t, 1\rangle.$$

Example 3.38. Consider a plane curve which is given by the graph of $y = f(x)$. If we use the parametrization $\langle x, f(x), 0\rangle$ of the curve, then the TNB-frame can be computed by

$$\mathbf{T}(x) = \frac{1}{\sqrt{1+f'^2}}\langle 1, f'(x), 0\rangle$$
$$\mathbf{N}(x) = \frac{1}{\sqrt{1+f'^2}}\langle -f'(x), 1, 0\rangle$$
$$\mathbf{B}(x) = \langle 0, 0, 1\rangle.$$

We have defined the TNB-frame of a curve. Note that by the definition of $\mathbf{N}$, we must have

$$\mathbf{T}' = \kappa\mathbf{N}$$

for some $\kappa = |\mathbf{T}'| > 0$. We may still define the curvature of a curve to be

$$\kappa = |\mathbf{T}'|,$$

here the derivative is with respect to the arc length parameter and therefore, the equation for the plane curve still follows in this case.

Theorem 3.39. *Given a smooth curve with a parametrization* $\mathbf{r}(t)$*, the curvature of the curve is*

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

which is independent of the choice of the parametrization.

One may ask that in this case, what is the relation between the vectors $\mathbf{B}$ and $\mathbf{T}$, $\mathbf{N}$. First note that the vectors $\mathbf{T}$, $\mathbf{N}$, and $\mathbf{B}$ forms a right-handed coordinate system, in other words,

$$\mathbf{T} = \mathbf{N} \times \mathbf{B}, \mathbf{N} = \mathbf{B} \times \mathbf{T}, \mathbf{B} = \mathbf{T} \times \mathbf{N}.$$

Since we used the cross product $\mathbf{T} \times \mathbf{N}$ to define the vector $\mathbf{B}$ and $\mathbf{T}' \parallel \mathbf{N}$, using the product rule for the cross product, the derivative

$$\mathbf{B}' = \mathbf{T}' \times \mathbf{N} + \mathbf{T} \times \mathbf{N}' = \mathbf{T} \times \mathbf{N}'.$$

This shows that $\mathbf{B}' \perp \mathbf{T}$. Now note that since $|\mathbf{B}| = 1$ is a constant, we have $\mathbf{B}' \perp \mathbf{B}$. And therefore $\mathbf{B}'$ must be parallel to $\mathbf{N}$ (since it is perpendicular to both $\mathbf{B}$ and $\mathbf{T}$.) Thus, there must be some scalar τ such that

$$\mathbf{B}' = -\tau\mathbf{N}.$$

This scalar function τ is called the torsion of the curve. Note that, here the torsion is defined using the arc length parameter s and the derivative of $\mathbf{B}$ with respect to s. Using the same computational techniques as in Theorem 3.32, we may obtain that

Theorem 3.40. *Given a smooth curve with a parametrization* $\mathbf{r}(t)$*, the torsion of the curve is*

$$\tau = \pm\frac{|\mathbf{B}'(t)|}{|\mathbf{r}'(t)|} = \frac{\mathbf{r}'(t) \times \mathbf{r}''(t) \cdot \mathbf{r}'''(t)}{|\mathbf{r}(t)' \times \mathbf{r}''(t)|^2},$$

where the $\pm$ *is determined by the sign of the triple product* $\mathbf{r}' \times \mathbf{r}'' \cdot \mathbf{r}'''$.

And

Theorem 3.41. *τ is geometic, in other words, it is independent of the choice of the parametrization.*

Example 3.42. From Example 3.38, since $\mathbf{B}$ is a constant vector, the torsion of the curve $\langle x, f(x), 0\rangle$ is 0. In fact, in general, a curve is contained in a plane, if and only if its torsion $\tau = 0$.

Example 3.43. Compute the curvature and torsion of the helix $\mathbf{r}(t) = \langle \cos t, \sin t, t\rangle$.

Solution. Let us compute this in two different ways. First, in Example 3.37, we have already computed the TNB-frame of this curve to be

$$\mathbf{T}(t) = \frac{1}{\sqrt{2}}\langle -\sin t, \cos t, 1\rangle.$$

$$\mathbf{N}(t) = \langle -\cos t, \sin t, 0\rangle.$$

$$\mathbf{B}(t) = \frac{1}{\sqrt{2}}\langle \sin t, -\cos t, 1\rangle.$$

From this and $|\mathbf{r}'| = \sqrt{2}$ we obtain that

$$\kappa = \frac{|\mathbf{T}'|}{|\mathbf{r}'|} = \frac{1}{2},$$

and

$$\tau = \pm\frac{|\mathbf{B}'|}{|\mathbf{r}'|} = \pm\frac{1}{2},$$

to determine the sign of τ, we still have to investigate the sign of $\mathbf{r}' \times \mathbf{r}'' \cdot \mathbf{r}'''$. One may check that in this case,

$$\tau = \frac{1}{2}.$$

On the other hand, we may use the derivatives of $\mathbf{r}$ to compute κ and τ directly, and because

$$\mathbf{r}' = \langle -\sin t, \cos t, 1\rangle, \mathbf{r}''(t) = \langle -\cos t, -\sin t, 0\rangle, \mathbf{r}'''(t) = \langle \sin t, -\cos t, 0\rangle,$$

By computing the cross products, we will still obtain

$$\kappa = \frac{1}{2}, \tau = \frac{1}{2}.$$

Remark 3.44. We have seen that the plane curve with constant curvature is the circle and now through this example a space curve with constant curvature and torsion is the helix. In fact, from Figure 9, one may see that the helix is a curve which is circling around a cylinder with the z-coordinate increasing at a constant rate. This provides a geometric intuition of the torsion τ. Intuitively, torsion τ describes the tendency for a plane curve to leave its plane. For the space curve, this plane is spanned by the unit vectors $\mathbf{T}$ and $\mathbf{N}$, which is called the osculating plane (with normal vector $\mathbf{B}$).

We have obtained that in the arc length parametrization, the TNB-frame satisfies the relation

$$\mathbf{T}' = \kappa\mathbf{N}, \mathbf{B}' = -\tau\mathbf{N}.$$

Using the dot and cross product, one may immediately obtain that

$$\mathbf{N}' = -\kappa\mathbf{T} + \tau\mathbf{B}.$$

These relations between the TNB-frame and their derivatives are called the Frenet-Serret formulas. In fact, given two functions τ and κ, the above equations becomes a set of differential equations for the curve $\mathbf{r}(t)$, and the Frenet-Serret theorem indicates that in this case, the curve $\mathbf{r}(t)$ is determined up to an isometry by its curvature and torsion.

Remark 3.45. This also implies that a curve with constant curvature and 0 torsion must be a circle and constant curvature and non-zero torsion must be a helix.

3.6. **Velocity and Acceleration.** And application of the normal vectors and curvature in physics is the notion of velocity and acceleration. Given a vector valued function $\mathbf{r}(t)$, the derivative

$$\mathbf{r}'(t) = \lim_{\Delta t \to 0} \frac{\mathbf{r}(t + \Delta t) - \mathbf{r}(t)}{\Delta t}$$

is the tangent vector to the curve. On the other hand, if we consider a particle moves along the curve $\mathbf{r}$, then the above derivative is the instantaneous rate of change of its position, which is called the velocity vector.

$$\mathbf{v}(t) = \mathbf{r}'(t).$$

The magnitude of the velocity vector $|\mathbf{v}|$ is called the speed of the particle. The velocity vector provides the information for both the change of direction and displacement. And in fact, from the equation

$$|\mathbf{v}| = |\mathbf{r}'(t)| = \frac{ds}{dt},$$

where s is the arc length function, we see that speed is the instantaneous rate of change of the displacement of a particle, which is measured by the arc length. And we define the acceleration to be the derivative of the velocity.

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

Example 3.46. Suppose that a particle is moving along the curve $\mathbf{r}(t) = \langle t^2, e^t, te^t \rangle$, find its velocity, speed and acceleration.

Solution. This is found by direct computation.

$$\mathbf{v}(t) = \langle 2t, e^t, (1+t)e^t \rangle$$
$$\mathbf{a}(t) = \langle 2, e^2, (2+t)e^t \rangle$$

and

$$|\mathbf{v}(t)| = \sqrt{4t^2 + e^{2t} + (1+t)^2 e^{2t}}.$$

Example 3.47. A particle with the initial position $\langle 1, 0, 0 \rangle$ and initial velocity $\langle 1, -1, 1 \rangle$ is moving with the acceleration $\mathbf{a}(t) = \langle 4t, 6t, 1 \rangle$. Find its position function at time t.

Solution. Because $\mathbf{a}(t) = \mathbf{v}'(t)$,

$$\mathbf{v}(t) = \int \mathbf{a}(t) dt = \langle 2t^2, 3t^2, t \rangle + \mathbf{C_1}.$$

Using $\mathbf{v}(0) = \langle 1, -1, 1 \rangle$, we obtain that $\mathbf{C_1} = \langle 1, -1, 1 \rangle$ and

$$\mathbf{v} = \langle 2t^2 + 1, 3t^2 + 1, t + 1 \rangle.$$

Now the position

$$\mathbf{r}(t) = \int \mathbf{v}(t) dt = \langle \frac{2}{3} t^3 + t, t^3 - t, \frac{1}{2} t^2 + t \rangle + \mathbf{C_2}.$$

And since $\mathbf{r}(0) = \langle 1, 0, 0 \rangle$, we obtain that

$$\mathbf{r}(t) = \langle \frac{2}{3}t^3 + t + 1, t^3 - t, \frac{1}{2}t^2 + t \rangle.$$

This example illustrates how to compute the position of a particle if the acceleration vector is given. In fact, the Newton's Second Law of Motion indicates that the acceleration is proportional to the force acting on the particle, which is given by

$$\mathbf{F}(t) = m\mathbf{a}(t),$$

where m is the mass of the particle.

Example 3.48. An object of mass m is moving in a circular trajectory with a constant angular speed w. In this case, the position vector is $\mathbf{r}(t) = \langle a \cos wt, a \sin wt \rangle$, where a is the radius of the circle.

In this case, the acceleration is

$$\mathbf{a}(t) = \mathbf{r}''(t) = \langle -aw^2 \cos wt, -aw^2 \sin wt \rangle,$$

and Newton's Second Law of Motion gives that the force is

$$\mathbf{F} = m\mathbf{a} = -mw^2\mathbf{r}(t),$$

which is in the direction of $-\mathbf{r}$ and proportional to w^2. This force is called a centripetal force.

Our last topic is the tangential and normal component of the acceleration. When we study the force or acceleration of a particle, sometimes, its is convenient to decompose it into components. In general, since $\mathbf{a}(t)$ is a vector in the space, and because we are always able to write $\mathbf{a}$ as a linear combination of the $\mathbf{T}$, $\mathbf{N}$, $\mathbf{B}$ vectors, since $\mathbf{T}$, $\mathbf{N}$, $\mathbf{B}$ forms an orthonormal basis of the vector space. On the other hand, since $\mathbf{a}(t) = \mathbf{v}'(t)$, let $v = |\mathbf{v}(t)|$ be the speed. Then we have

$$\mathbf{a}(t) = (v\mathbf{T})' = v'\mathbf{T} + v\mathbf{T}'.$$

Here, the derivative is with respect to the parameter t. Now because $\kappa = \frac{|\mathbf{T}'|}{v}$, and $\mathbf{N} = \mathbf{T}'/|\mathbf{T}'|$,

$$\mathbf{T}' = \kappa v \mathbf{N},$$

and

$$\mathbf{a} = a_T\mathbf{T} + a_N\mathbf{N} = v'\mathbf{T} + \kappa v^2\mathbf{N},$$

where

$$a_T = v', a_N = \kappa v^2$$

is called the tangential and normal component of the acceleration.

Remark 3.49. Note that this also indicates that in the TNB decomposition of the acceleration $\mathbf{a}$, there is no binormal component $\mathbf{B}$.

On the other hand, since

$$\mathbf{v} \cdot \mathbf{a} = v\mathbf{T} \cdot (v'\mathbf{T} + \kappa v^2\mathbf{N}) = vv',$$

$$v' = \frac{\mathbf{v} \cdot \mathbf{a}}{v}$$

and using the formula of κ we obtain that

Proposition 3.50. *Let* $\mathbf{a} = a_T\mathbf{T} + a_N\mathbf{N}$. *Then*

$$a_T = \frac{\mathbf{r}' \cdot \mathbf{r}''}{|\mathbf{r}'|}, a_N = \frac{|\mathbf{r}' \times \mathbf{r}''|}{|\mathbf{r}'|}.$$

4. Partial Derivatives

4.1. **Multi-variable Functions.** A function can be viewed as a relation between the set of domain and the set of range. This relation has to satisfy a basic condition, which is for each element in the domain, there is one unique element in the range that is related to this element.

In Calculus 1, We have learned the single variable function, whose domain and range are both subsets of $\mathbb{R}$, and the relation can be represented by $y = f(x)$.

Example 4.1. If we use the notation x to denote an element in the domain, and y an element in the range. Then the relation $x^2 + y^2 = 1$ is not a function, since for a fixed x, there are more than one y is "related" to x under this relation.

On the other hand, if we have a relation given by $y = \sqrt{1 - x^2}$. In this case, we obtain a function whose domain is $[-1, 1]$ and range $[0, 1]$.

Definition 4.2. A function of two variables is a relation between a domain, which is a subset of $\mathbb{R}^2$ and a range, which is a subset of $\mathbb{R}$. The function is denoted by $z = f(x, y)$.

If a function f is given by some formula and no domain is specified, then we usually define the domain to be all possible values of (x, y) so that $f(x, y)$ is well-defined.

Example 4.3. Find the domain of the following function.

(a) $f(x, y) = \frac{\sqrt{x+y+1}}{x-1}$

Solution. In order for $f(x, y)$ to be well-defined, the expression under the square root must be non-negative and the denominator must be non-zero.

Therefore, we have the domain $D = \{(x, y)|x + y + 1 \geq 0, x \neq 1\}$.

(b) $f(x, y) = x \ln(y^2 - x)$

Solution. For this function, $\ln(y^2 - x)$ is defined only when $y^2 - x > 0$. In other words, the domain is $D = \{(x, y)|x < y^2\}$.

Note that similar to the "open" and "closed" interval in the case of single variable functions, the domain of a function of variable can be open or closed or neither open nor closed regions in $\mathbb{R}^2$. We say a subset of $\mathbb{R}^2$ is closed, if the boundary of this region is contained in it. Usually, a closed region is described using an inequalities with $\geq$ or $\leq$ sign.

Similarly, we say a region is open, if its compliment is closed. In this case, the region does not include its boundary. Usually these regions are described by some strict inequalities with $>$ or $<$.

Once we are able to identify the domain of a function as a region in $\mathbb{R}^2$, one way to visualize a function is to consider its graph.

Definition 4.4. Let $z = f(x, y)$ be a function of two variables. If D is the domain of the function, we define the graph of the function to be the set $\{(x, y, z)|z = f(x, y), (x, y) \in D\}$.

Usually the graph of $z = f(x, y)$ is a surface such that if (x, y, z) is a point on the surface, then (x, y) is in the domain of the function and $z = f(x, y)$ is the value of the function at (x, y).

Sometimes we use another method, which is called the level curves, to visualize a function of two variables.

Definition 4.5. The level curve of a function of two variables are the curves with equations $f(x, y) = k$, where k is a constant. We may plot the curve on the xOy-plane.

Intuitively, the level curves of a function is dense, when the value of the function changes more drastically, and the level curves are more sparse, when the rate of change is small. We will see a precise characterization of this phenomenon during the gradient section.

In general, a function of multi-variable is a relation between a subset of $\mathbb{R}^n$ and a subset of $\mathbb{R}$. In this case, a function can be denoted by $f(x_1, x_2, \ldots, x_n)$. The level set of a function of three variable $f(x, y, z)$ is called a level surface. They are the surfaces given by the equations $f(x, y, z) = k$.

Example 4.6. Identify the level surfaces of the function $f(x, y, z) = x^2 + y^2 + z^2$.

Solution. The level surfaces have the equations $x^2 + y^2 + z^2 = k$, $k \geq 0$. These are the spheres centered at O with radius $\sqrt{k}$, when $k > 0$. When $k = 0$, it is a single point.

4.2. **Limit and Continuity.** One important property of a function is its continuity. Let us study the limit and continuity of functions of multi-variable.

We use the notation

$$\lim_{(x,y)\to(a,b)} f(x, y) = L$$

to indicate that the value of the function $f(x, y)$ approaches L when the point (x, y) approaches (a, b). The precise definition is the following.

Definition 4.7. Suppose that $f(x, y)$ is a function of two variables. We say the limit of $f(x, y)$ as (x,y) approaches (a, b) is L, if for every $\varepsilon > 0$, there is a $\delta(\varepsilon) > 0$, such that if the distance $\sqrt{(x - a)^2 + (x - b)^2} < \delta$, then $|f(x, y) - L| < \varepsilon$. And we write

$$\lim_{(x,y)\to(a,b)} f(x, y) = L.$$

Remark 4.8. The notation $\lim_{(x,y)\to(a,b)} f(x, y)$ is different from $\lim_{x\to a, y\to b} f(x, y)$. The latter one usually indicates we are taking the limit $y \to b$ first, and then $x \to a$.

Remark 4.9. The condition $\sqrt{(x - a)^2 + (x - b)^2} < \delta$ means when (x, y) approaching (a, b) is in the sense that the distance between (x, y) and (a, b) is decreasing. It does not specify a path or a direction such that (x, y) approaches (a, b). In fact, (x, y) may approach (a, b) along any path and in any direction. This gives us a way to show when the limit does not exist.

Proposition 4.10. *If along a path C_1, the limit $f(x, y) \to L_1$ and along C_2, the limit is $f(x, y) \to L_2$. If $L_1 \neq L_2$, then the limit $\lim_{(x,y)\to(a,b)} f(x, y)$ does not exist.*

Example 4.11. The limit $\lim_{(x,y)\to(0,0)} \frac{x^2 - y^2}{x^2 + y^2}$ does not exist.

Solution. Let us first consider the limit when (x, y) approaches $(0, 0)$ along the x-axis. In this case we have $y = 0$. $f(x, 0) = 1$ for all $x \neq 0$. Therefore, $\lim_{x\to 0} f(x, 0) = 1$.

On the other hand, if we consider the path y-axis. In this case, $x = 0$ and $f(0, y) = -1$, so $\lim_{y\to 0} f(0, y) = -1$.

Since $1 \neq -1$, the limit does not exist.

Example 4.12. The limit $\lim_{(x,y)\to(0,0)} \frac{xy}{x^2+y^2}$ does not exist.

Solution. For this limit, we see that either along $x = 0$ or $y = 0$ will give us the same result where $f(0, y) = f(x, 0) = 0$. Therefore, we obtain the identical limit along the axes.

However, if we consider the limit when we approach $(0, 0)$ along the line $y = x$, then for $x \neq 0$,

$$f(x, x) = \frac{x^2}{x^2 + x^2} = \frac{1}{2} \neq 0.$$

This implies that the limit does not exist.

Example 4.13. The limit $\lim_{(x,y)\to(0,0)} \frac{xy^2}{x^2+y^4}$ does not exist.

Solution. Follow the previous example, let us first consider the limit when we approach $(0, 0)$ along some line $y = kx$. Because

$$f(x, kx) = \frac{k^2x^3}{x^2 + k^4x^4},$$

we have $\lim_{x\to 0} f(x, kx) = 0$.

This shows that $f(x, y)$ has the same limit when we approach $(0, 0)$ along straight lines. But that does not show the limit is 0. In fact, if we consider the limit when we approach $(0, 0)$ along the parabola $x = y^2$, then for $x \neq 0$,

$$f(y^2, y) = \frac{y^4}{y^4 + y^4} = \frac{1}{2} \neq 0.$$

This again implies that the limit does not exist.

Let us now consider an example where the limit do exist.

Example 4.14. Compute the limit $\lim_{(x,y)\to(0,0)} \frac{3x^2y}{x^2+y^2}$.

Solution. Through a quick computation we see that the limit when we approach along $y = kx$ or parabola $y = x^2, x = y^2$ are all 0. In this case, we suspect that the limit exist and is 0.

The idea of determine whether the limit exists or not from Definition 4.7 is whether we are able to compare the value of the function with the limit (ε) as a function of the distance $\sqrt{(x-a)^2 + (y-b)^2}$ (δ).

In other words, if $\sqrt{x^2 + y^2} < \delta$, we would like to control $\left|\frac{3x^2y}{x^2+y^2} - 0\right|$.

Since when $x, y \neq 0$, $\frac{x^2}{x^2+y^2} \leq 1$,

$$\left|\frac{3x^2y}{x^2 + y^2} - 0\right| \leq 3|y| = 3\sqrt{y^2} \leq 3\sqrt{x^2 + y^2} < 3\delta,$$

if we pick $\varepsilon = 3\delta$, then

$$\lim_{(x,y)\to(0,0)} f(x, y) = 0.$$

Through the above examples we see that the limit of a multi-variable function when we approach a point which is not in the domain of the function can be quite complicated. On the other hand, similar to the case of single variable, when the point is in the domain, we can usually use the continuity property to compute the limit.

Definition 4.15. A function $f(x, y)$ is continuous at (a, b) if

$$\lim_{(x,y)\to(a,b)} f(x, y) = f(a, b).$$

We say f is continuous on a set D, if it is continuous at every point in D.

Let us introduce, without proof the continuity of the elementary functions.

Theorem 4.16. *If $f(x, y)$ is a composition of the elementary functions, then it is continuous on its domain.*

Example 4.17. Find $\lim_{(x,y)\to(1,0)} \arctan \frac{y}{x}$.

Solution. Note that $(1, 0)$ is in the domain of the function since $x = 1 \neq 0$. The function $\arctan \frac{y}{x}$ is continuous at $(1, 0)$ and thus,

$$\lim_{(x,y)\to(1,0)} \arctan \frac{y}{x} = \arctan \frac{0}{1} = 0.$$

If we have a function $f(x_1, x_2, \ldots, x_n)$, we say the limit of f when $\mathbf{x} = (x_1, \ldots, x_n)$ approaches $\mathbf{a} = (a_1, \ldots, a_n)$ is L, if for any $\varepsilon > 0$, we can find δ such that when $0 < |\mathbf{x} - \mathbf{a}| < \delta$, $|f - L| < \varepsilon$. Here the first $|, |$ is the magnitude. In this case, we write

$$\lim_{\mathbf{x}\to\mathbf{a}} f(\mathbf{x}) = L$$

And we say f is continuous at $\mathbf{a}$, if

$$\lim_{\mathbf{x}\to\mathbf{a}} f(\mathbf{x}) = f(\mathbf{a}).$$

4.3. **Partial Derivatives.** The relation between the value of a multi-variable function and the variables can be quite complicated. For example, let us consider the wind chill index $I(v, T)$ which is a function of the wind speed v and the temperature T. One way to study how I depends on v and T is that we fix one variable, say wind speed v as a constant, and we investigate how I changes according to the temperature T. This is the idea of the partial derivative.

Suppose that we have a function of two variables $f(x, y)$. Consider a point (a, b) in the domain of f. We may look at the function, when we fix $y = b$ to be a constant. In this case, we define the function $g(x) = f(x, b)$. Now g is a single variable function which only depends on x. We may investigate the rate of change of g by taking the derivative of it.

Definition 4.18. The partial derivative of f with respect to x at (a, b) is

$$\frac{\partial f}{\partial x}(a, b) = g'(a) = \lim_{\Delta x\to 0} \frac{g(a + \Delta x) - g(a)}{\Delta x} = \lim_{\Delta x\to 0} \frac{f(a + \Delta x, b) - f(a, b)}{\Delta x}.$$

Similarly, the partial derivative of f with respect to y at (a, b) is

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

This definition allows us to define the partial derivative of f at any points in its domain, provided that the limits exist. In this case, we call the functions f_x, f_y defined as follows, the partial derivatives of the function f.

Definition 4.19.

$$f_x(x,y) = \frac{\partial f}{\partial x}(x,y)$$

$$f_y(x,y) = \frac{\partial f}{\partial y}(x,y)$$

There are many alternative notations for partial derivatives. Let us list some common notations might be used to represent the partial derivative. Suppose that $z = f(x,y)$ is a function of two variables. Then

Definition 4.20 (Notation)**.**

$$z_x = f_x = \frac{\partial z}{\partial x} = \frac{\partial f}{\partial x} = \frac{\partial}{\partial x} f = D_x f = D_1 f = f_1.$$

$$z_y = f_y = \frac{\partial z}{\partial y} = \frac{\partial f}{\partial y} = \frac{\partial}{\partial y} f = D_y f = D_2 f = f_2.$$

To compute the partial derivatives, according to Definition 4.18, we simply just regard one variable as a constant and take the derivative with respect to the other variable.

Example 4.21.

- To find f_x, we regard y as a constant and differentiate with respect to x.
- To find f_y, we regard x as a constant and differentiate with respect to y.

Example 4.22. Compute the partial derivative f_x and f_y of $f = x^3 + x^2y^3 - 2y^2$. And find $f_x(2,1)$ and $f_y(2,1)$.

Solution. To find f_x, we fix y as constant and take derivative with respect to x. This gives us

$$f_x = 3x^2 + 2xy^3$$

and $f_x(2,1) = 16$. Now to find f_y, we fix x as constant so

$$f_y = 3x^2y^2 - 4y$$

and $f_y(2,1) = 8$.

Example 4.23. Let $f(x,y) = \sin\left(\frac{x}{1+y}\right)$. Find $\frac{\partial f}{\partial x}$ and $\frac{\partial f}{\partial y}$.

Solution. To compute the partial derivatives of this function, we have to apply the chain rule for the single variable functions, since when we view x or y as a constant, this is a composition of some functions of y or x.

$$\frac{\partial f}{\partial x} = \cos\left(\frac{x}{1+y}\right) \cdot \frac{1}{1+y},$$

$$\frac{\partial f}{\partial y} = \cos\left(\frac{x}{1+y}\right) \cdot \frac{-x}{(1+y)^2}.$$

Example 4.24. Find the partial derivative $\frac{\partial z}{\partial x}$ if z is defined implicity as $z = f(x,y)$ by the equation

$$x^3 + y^3 + z^3 + 6xyz = 1.$$

Solution. To find $\frac{\partial z}{\partial x}$, we view y as a constant and in this case the equation becomes an equation about x and z. So we use the implicit differentiation with respect to x.

$$3x^2 + 0 + 3z^2\frac{\partial z}{\partial x} + 6yz + 6xy\frac{\partial z}{\partial x} = 0$$

which gives

$$\frac{\partial z}{\partial x} = -\frac{y^2 + 2xz}{z^2 + 2xy}.$$

If f is a function of x, y, its partial derivative is also a function of x, y. Therefore, we may consider the partial derivatives of f_x and f_y, which are called the second partial derivatives. If $z = f(x, y)$, then we write

Definition 4.25.

$$(f_x)_x = f_{xx} = f_{11} = \frac{\partial}{\partial x}f_x = \frac{\partial}{\partial x}\left(\frac{\partial f}{\partial x}\right) = \frac{\partial^2 f}{\partial x^2} = \frac{\partial^2 z}{\partial x^2}.$$

$$(f_x)_y = f_{xy} = f_{12} = \frac{\partial}{\partial y}f_x = \frac{\partial}{\partial y}\left(\frac{\partial f}{\partial x}\right) = \frac{\partial^2 f}{\partial y\partial x} = \frac{\partial^2 z}{\partial y\partial x}.$$

$$(f_y)_x = f_{yx} = f_{21} = \frac{\partial}{\partial x}f_y = \frac{\partial}{\partial x}\left(\frac{\partial f}{\partial y}\right) = \frac{\partial^2 f}{\partial x\partial y} = \frac{\partial^2 z}{\partial x\partial y}.$$

$$(f_y)_y = f_{yy} = f_{22} = \frac{\partial}{\partial y}f_y = \frac{\partial}{\partial y}\left(\frac{\partial f}{\partial y}\right) = \frac{\partial^2 f}{\partial y^2} = \frac{\partial^2 z}{\partial y^2}.$$

Example 4.26. Find the second partial derivative of $f = x^3 + x^2y^3 - 2y^2$.
Solution. We have already found the first partial derivatives

$$f_x = 3x^2 + 2xy^3, f_y = 3x^2y^2 - 4y.$$

Now to compute the second partial derivatives, we take the derivatives of f_x and f_y with respect to x, y, which gives us

$$f_{xx} = 6x + 2y^3, f_{xy} = 6xy^2.$$

$$f_{yx} = 6xy^2, f_{yy} = 6x^2y - 4.$$

In this example, we obtained that $f_{xy} = f_{yx}$. In fact, this is not a coincidence.

Theorem 4.27 (Clairaut)**.** *Let $f(x, y)$ be a function of two variables. If the second partial derivatives f_{xy}, f_{yx} are continuous, then*

$$f_{xy} = f_{yx}.$$

In other words, the mixed order partial derivatives does not depend on the order.

Using the same idea we may define the higher derivatives $f_{xxx}, \dots$. Clairaut's Theorem holds for higher derivative of any order as long as those partial derivatives are continuous functions of x, y.

We may also generalize this to the case of function of more than two variables. Let $f = f(x_1, \dots, x_n)$. In this case, the partial derivative $\frac{\partial f}{\partial x_i}$ is the derivative with respect to x_i, while we view all other variables as constants. The computation is similar to the case of two variables.

4.4. **Tangent plane and linear approximation.** So far we have been discussing the definition and computation of the partial derivatives. We have not introduce what we mean by a function $f(x, y)$ is differentiable in this case. Note that due to the bizarre behaviour of the limit of multi-variable functions, the partial derivative of a function exist does not indicate that the function is differentiable. In fact, the partial derivatives are the derivatives along the $x-$axis or $y-$axis, and we would like to characterize the differentiability as a property with respect to the limit of a two-variable function.

To begin with, let us first consider a geometric interpretation of the partial derivative. Recall that $z = f(x, y)$ may be viewed as a surface in $\mathbb{R}^3$. To compute the partial derivative $\frac{\partial f}{\partial x}$ at some point (a, b), we fix y as the constant b and differentiate with respect to x and then evaluate at $x = a$.

Fixing $y = b$ in the 3-dimensional space can be viewed as looking at the intersection between the surface $z = f(x, y)$ and $y = b$. This intersection is a curve which has a natural parameterization by the parameter x. In fact, we can write it as $\mathbf{r}(x) = \langle x, b, f(x, b)\rangle$. Now the tangent vector to this curve is given by $\mathbf{r}' = \langle 1, 0, \frac{\partial f}{\partial x}\rangle$. If we view this curve on the xz-plane, then $\frac{\partial f}{\partial x}$ is precisely the slope of the tangent line to the curve.

Now similarly, if we look at the partial derivative $\frac{\partial f}{\partial y}$, where we fix $x = a$ to be a constant. Then we may obtain another curve which is the intersection between $z = f(x, y)$ and $x = a$ and can be parameterized by $\mathbf{r}(y) = \langle a, y, f(a, y)\rangle$. Therefore, we obatin another tangent vector which is $\langle 0, 1, \frac{\partial f}{\partial y}\rangle$ at the point (a, b).

In this way, the partial derivatives can be viewed as the slope of the tangent lines to the curve of intersection, when we place these curves onto the coordinate planes. Now at each point, there are two tangent lines in the direction of $\langle 1, 0, \frac{\partial f}{\partial x}\rangle$ and $\langle 0, 1, \frac{\partial f}{\partial y}\rangle$. The two tangent lines (they must be different since the (x, y) coordinate of the vectors are given by $(1, 0)$ and $(0, 1)$) determines a unique plane, which is called the tangent plane to the surface $z = f(x, y)$ at the point (a, b).

Let us now try to write down an equation of the tangent plane. We consider a surface $z = f(x, y)$ at the point (a, b). The tangent plane is a plane passes through $(a, b, f(a, b))$ and thus we may assume that its equation is given by

$$A(x - a) + B(y - b) + C(z - f(a, b)) = 0,$$

for some numbers A, B, C.

Now, if the plane is not vertical, in other words $C \neq 0$, we may rewrite the above equation as

$$z - f(a, b) = \alpha(x - a) + \beta(y - b),$$

for some constant α, β.

To see what is the numbers α and β in this case, note that when $x = a$, the line of the intersection

$$\begin{cases} z - f(a, b) & = \alpha(x - a) + \beta(y - b) \\ x & = a \end{cases}$$

must be a tangent line to the surface $z = f(x, y)$, and the directional vector of this line is $\langle 0, 1, \beta\rangle$. Compare this expression with the tangent vector $\langle 0, 1, \frac{\partial f}{\partial y}\rangle$, we immediately obtain that $\beta = \frac{\partial f}{\partial y}$ and similarly, $\alpha = \frac{\partial f}{\partial x}$, so that the equation of the tangent plane is given by

Theorem 4.28. *Suppose that $z = f(x, y)$ has continuous partial derivatives at the point (a, b), then the equation of the tangent plane to the surface at (a, b) is*

$$z - f(a, b) = \frac{\partial f}{\partial x}(a, b)(x - a) + \frac{\partial f}{\partial y}(a, b)(y - b).$$

Example 4.29. Find the equation of the tangent plane to $z = 2x^2 + y^2$ at the point $(1, 1, 3)$.

Solution. If we denote $f(x, y) = 2x^2 + y^2$, then $f_x = 4x, f_y = 2y$. At $x = 1, y = 1$,

$$f_x(1, 1) = 4, f_y(1, 1) = 2.$$

Therefore, the equation of the tangent plane is

$$z - 3 = 4(x - 1) + 2(y - 1).$$

In the case of single variable functions, the tangent line can be viewed as the linear approximation of the function. Similarly, in the case of function of two variables, we may view the tangent plane as the linear approximation of the surface $z = f(x, y)$.

To be more precisely, if $L(x, y)$ is the $z-$coordinate of a point on the tangent plane at (a, b), in other words,

$$L(x, y) = f(a, b) + \frac{\partial f}{\partial x}(a, b)(x - a) + \frac{\partial f}{\partial y}(a, b)(y - b),$$

then we call $L(x, y)$ the linear approximation of the function $f(x, y)$ at (a, b).

We may use the value of $L(x, y)$ to approximate $f(x, y)$, in other words, $f(x, y) \approx L(x, y)$.

Example 4.30. In the previous example we have found the tangent plane of $z = 2x^2 + y^2$ to be $z - 3 = 4(x - 1) + 2(y - 1)$. Use this equation to approximate z at $(1.1, 0.95)$.

Solution. $L(x, y) = 3 + 4(x - 1) + 2(y - 1) = 4x + 2y - 3$. Therefore, at $(1.1, 0.95)$,

$$f(1.1, 0.95) \approx L(1.1, 0.95) = 4(1.1) + 2(0.95) - 3 = 3.3$$

Example 4.31. Let $f(x, y) = xe^{xy}$. Find its linear approximation at $(1, 0)$ and use it to approximate $f(1.1, -0.1)$.

Solution. Let us first compute the partial derivatives,

$$f_x = e^{xy} + xye^{xy}, f_y = x^2e^{xy}$$
$$f_x(1, 0) = 1, f_y(1, 0) = 1.$$

Therefore the linear approximation is

$$L(x, y) = 1 + (x - 1) + y = x + y.$$

At $(1.1, -0.1)$,

$$f(1.1, -0.1) \approx L(1.1, -0.1) = 1.1 - 0.1 = 1.$$

Note that the actual value is $1.1e^{-0.1} \approx 0.985$.

The condition partial derivatives are continuous is important in the definition of the linear approximation. Let us consider the following example.

Example 4.32. Let $f(x, y) = \begin{cases} \frac{xy}{x^2+y^2} & (x, y) \neq (0, 0) \\ 0 & (x, y) = (0, 0) \end{cases}$ One may compute that the partial derivatives of f both exist at $(0, 0)$, (but not continuous). In fact, $f_x(0, 0) = f_y(0, 0) = 0$. So the linear approximation $L(x, y) = 0$. However, if we consider the path $y = x$, $f(x, y) = \frac{1}{2}$ for every point which is not $(0, 0)$. In this case, $f(x, y) \approx 0$, but $f(x, y) = \frac{1}{2}$.

This is an example of "not differentiable" function that the function near $(0,0)$ cannot by approximated by its linear approximation. We say a function is differentiable if locally it can be approximated by its linear approximation. Or more precisely,

Definition 4.33. Let $z = f(x,y)$ be a function of x,y. We say z is differentiable at (a,b), if for (x,y) near (a,b), $f(x,y) \approx L(x,y)$ and

$$\lim_{(x,y)\to(a,b)} f(x,y) - L(x,y) = 0.$$

Similar to the case of single variable, a different way to interpret linear approximation is by looking at the increment of the function and compare it with the linear increment.

If $z = f(x,y)$ with partial derivatives f_x, f_y, when (x,y) changes from (a,b) to $(a+\Delta x, b+\Delta y)$, we define the increment of z to be

$$\Delta z = f(a+\Delta x, b+\Delta y) - f(a,b).$$

And the linear increment

$$dz = L(a+\Delta x, b+\Delta y) - f(a,b) = f_x(a,b)\Delta x + f_y(a,b)\Delta y.$$

Now we see that if a function is differentiable, then the increment Δz can be approximated by the linear increment dz, and $\lim_{(\Delta x,\Delta y)\to(0,0)} \Delta z - dz = 0$. Indeed, we simply just subtract the value $f(a,b)$ from Definition 4.33. Here, we give a more precise definition of the differentiability.

Definition 4.34. If $z = f(x,y)$, then f is differentiable at point (a,b) if Δz can be expressed in the form

$$\Delta z = f_x(a,b)\Delta x + f_y(a,b)\Delta y + \varepsilon_1 \Delta x + \varepsilon_2 \Delta y,$$

where $\varepsilon_1, \varepsilon_2 \to 0$ when $(\Delta x, \Delta y) \to (0,0)$.

Remark 4.35. Sometimes we use the notation $dx = \Delta x$ and $dy = \Delta y$ to denote the increment of x, y. So that

$$dz = f_x dx + f_y dy = \frac{\partial z}{\partial x}dx + \frac{\partial z}{\partial y}dy.$$

In this case, we call dz the differential or the total differential of the function z. In this notation, the linear approximation can be also written as

$$f(x,y) \approx f(a,b) + dz$$

Example 4.36. Let $z = f(x,y) = x^2 + 3xy - y^2$. Find dz. If (x,y) changes from $(2,3)$ to $(2.05, 2.96)$, compare the values of Δz and dz.

Solution.

$$dz = \frac{\partial z}{\partial x}dx + \frac{\partial z}{\partial y}dy = (2x+3y)dx + (3x-2y)dy$$

Now if (x,y) changes from $(2,3)$ to $(2.05, 2.96)$, then

$$dx = \Delta x = 0.05, dy = \Delta y = -0.04.$$

So

$$dz = (2\cdot 2 + 3\cdot 3)\cdot 0.05 + (3\cdot 2 - 2\cdot 3)\cdot(-0.04) = 0.65$$

and

$$\Delta z = f(2.05, 2.96) - f(2,3) = 0.6449.$$

We see that in this case $\Delta z \approx dz$.

4.5. **A brief overview.** The differential $dz = \frac{\partial z}{\partial x}dx + \frac{\partial z}{\partial y}dy$ plays a central role in this section. We will give a brief overview of how it is related to the topics in this section. In fact, these are all the same object from different point of view.

- The equation of the tangent plane $z - f(a,b) = \frac{\partial z}{\partial x}(x-a) + \frac{\partial z}{\partial y}(y-b)$
- Differentiability $\Delta z \approx dz$
- Linear approximation $f(x,y) \approx f(a,b) + dz$
- Chain rule $\frac{dz}{dt} = \frac{\partial z}{\partial x}\frac{dx}{dt} + \frac{\partial z}{\partial y}\frac{dy}{dt}$
- Directional derivative $D_{\langle u,v\rangle} z = \frac{\partial z}{\partial x}u + \frac{\partial z}{\partial y}v$
- First order Taylor polynomial $f(x,y) = f(a,b) + \frac{\partial z}{\partial x}(x-a) + \frac{\partial z}{\partial y}(y-b)$.

4.6. **Chain rule.** Before we proceed further, let us introduce the chain rule, a computation rule for the partial derivatives, if we are taking the partial derivative of the composition of multi-variable functions.

Let $z = f(x,y)$ be a differentiable function. If $x = x(t)$, $y = y(t)$ are differentiable. Consider the function $z(t) = f(x(t), y(t))$. Let us compute the derivative $z'(t)$. By the differential formula we know that

$$dz = \frac{\partial z}{\partial x}dx + \frac{\partial z}{\partial y}dy,$$

and the fact that z is differentiable implies

$$\Delta z \approx dz = \frac{\partial z}{\partial x}\Delta x + \frac{\partial z}{\partial y}\Delta y.$$

Now if we divide Δt on both sides,

$$\frac{\Delta z}{\Delta t} \approx \frac{\partial z}{\partial x}\frac{\Delta x}{\Delta t} + \frac{\partial z}{\partial y}\frac{\Delta y}{\Delta t}.$$

And let $\Delta t \to 0$, we obtain that

$$\frac{dz}{dt} = \frac{\partial z}{\partial x}\frac{dx}{dt} + \frac{\partial z}{\partial y}\frac{dy}{dt}.$$

This gives us the derivative of $z(t)$.

Proposition 4.37 (Chain rule I)**.** *If $z = f(x,y)$ is differentiable and both $x = x(t)$, $y = y(t)$ are differentiable, then*

$$\frac{dz}{dt} = \frac{\partial z}{\partial x}\frac{dx}{dt} + \frac{\partial z}{\partial y}\frac{dy}{dt}.$$

Example 4.38. Let $z = x^2y + 3xy^4$, $x = \sin 2t$, $y = \cos t$. Find $\frac{dz}{dt}$ when $t = 0$.

Solution. By the Chain rule,

$$\frac{dz}{dt} = \frac{\partial z}{\partial x}\frac{dx}{dt} + \frac{\partial z}{\partial y}\frac{dy}{dt} = (2xy + 3y^4)(2\cos 2t) + (x^2 + 12xy^3)(-\sin t).$$

When $t = 0$, $x = \sin 0 = 0$, $y = \cos 0 = 1$. So

$$\frac{dz}{dt}(0) = (0+3)(2\cos 0) + (0+0)(-\sin 0) = 6.$$

The Chain rule 4.37 can be generalized to the case where $x = x(t,s)$, $y = y(t,s)$ differentiable, simply by the limit definition of the partial derivatives. In fact, we have

Proposition 4.39 (Chain rule II)**.** *If $z = f(x, y)$ is differentiable and both $x = x(t, s)$, $y = y(t, s)$ are differentiable, then*

$$\frac{\partial z}{\partial t} = \frac{\partial z}{\partial x}\frac{\partial x}{\partial t} + \frac{\partial z}{\partial y}\frac{\partial y}{\partial t}.$$

$$\frac{\partial z}{\partial s} = \frac{\partial z}{\partial x}\frac{\partial x}{\partial s} + \frac{\partial z}{\partial y}\frac{\partial y}{\partial s}.$$

Example 4.40. Let $z = e^x \sin y$ and $x = st^2$, $y = s^2 t$. Find $\frac{\partial z}{\partial t}$ and $\frac{\partial z}{\partial s}$.

Solution. Using the Chain rule,

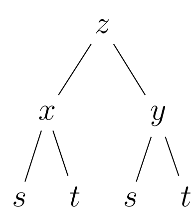
$$\frac{\partial z}{\partial t} = \frac{\partial z}{\partial x}\frac{\partial x}{\partial t} + \frac{\partial z}{\partial y}\frac{\partial y}{\partial t} = (e^x \sin y)(t^2) + (e^x \cos y)(2st)$$

$$\frac{\partial z}{\partial s} = \frac{\partial z}{\partial x}\frac{\partial x}{\partial s} + \frac{\partial z}{\partial y}\frac{\partial y}{\partial s} = (e^x \sin y)(2st) + (e^x \cos y)(s^2).$$

Proposition 4.41. *In general, if z is a differentiable function of $x_1, \dots, x_n$ and each x_i is a differentiable function of $t_1, \dots, t_m$. Then*

$$\frac{\partial z}{\partial t_j} = \frac{\partial z}{\partial x_1}\frac{\partial x_1}{\partial t_j} + \cdots + \frac{\partial z}{\partial x_n}\frac{\partial x_n}{\partial t_j}.$$

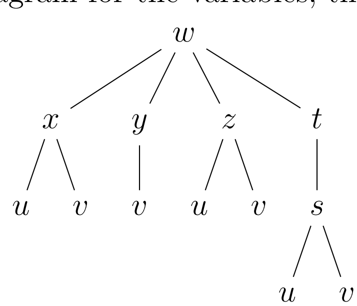
One way to remember the Chain rule is to draw the tree diagram. If a variable z is a function of some other variables x, y, then we draw a tree with a vertex z and two children x, y. If $x = x(s, t)$, $y = y(s, t)$, then we may draw a tree diagram as following



Each path represents a partial derivative. If we want to find a partial derivative, say $\frac{\partial z}{\partial s}$, we find the product of partial derivatives along each path form z to s and then we add these products together.

Example 4.42. Write down the chain rule for $\frac{\partial w}{\partial v}$ if $w = w(x, y, z, t)$, $x = x(u, v)$, $y = y(v)$, $z = z(u, v)$, $t = t(s)$, $s = s(u, v)$.

Solution. if we draw a tree diagram for the variables, then

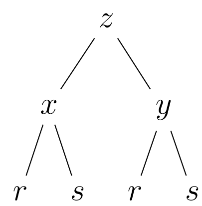


Therefore,

$$\frac{\partial w}{\partial v} = \frac{\partial w}{\partial x}\frac{\partial x}{\partial v} + \frac{\partial w}{\partial y}\frac{dy}{dv} + \frac{\partial w}{\partial z}\frac{\partial z}{\partial v} + \frac{\partial w}{\partial t}\frac{dt}{ds}\frac{\partial s}{\partial v}.$$

Example 4.43. Let $z = f(x, y)$ and $x = r^2 + s^2$, $y = 2rs$. Find $\frac{\partial z}{\partial r}$ and $\frac{\partial^2 z}{\partial r^2}$.

Solution. The relations between the variable can be expressed as the following

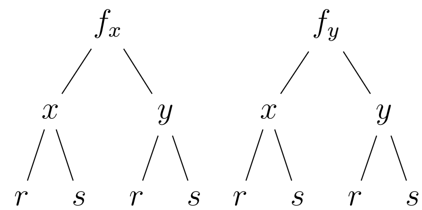


$$\frac{\partial z}{\partial r} = \frac{\partial z}{\partial x}\frac{\partial x}{\partial r} + \frac{\partial z}{\partial y}\frac{\partial y}{\partial r} = 2rf_x + 2sf_y.$$

To compute the second derivative, first note that

$$\frac{\partial^2 z}{\partial r^2} = \frac{\partial}{\partial r}\frac{\partial z}{\partial r} = \frac{\partial}{\partial r}\left(2rf_x + 2sf_y\right) = 2f_x + 2r\frac{\partial}{\partial r}f_x + 2s\frac{\partial}{\partial r}f_y.$$

Note that since z is a function of x, y, its partial derivatives f_x and f_y, in general should also be a function of x, y. In fact, the domain of f_x and f_y are subsets of the domain of f. Therefore,



Therefore,

$$\frac{\partial}{\partial r}f_x = \frac{\partial f_x}{\partial x}\frac{\partial x}{\partial r} + \frac{\partial f_x}{\partial y}\frac{\partial y}{\partial r} = 2r\frac{\partial^2 z}{\partial x^2} + 2s\frac{\partial^2 z}{\partial x \partial y}.$$

$$\frac{\partial}{\partial r}f_y = \frac{\partial f_y}{\partial x}\frac{\partial x}{\partial r} + \frac{\partial f_y}{\partial y}\frac{\partial y}{\partial r} = 2r\frac{\partial^2 z}{\partial x \partial y} + 2s\frac{\partial^2 z}{\partial y^2}.$$

Now if we combine these expressions together,

$$\frac{\partial^2 z}{\partial r^2} = 2f_x + 2r\frac{\partial}{\partial r}f_x + 2s\frac{\partial}{\partial r}f_y = 2f_x + 4r^2 f_{xx} + 8rsf_{xy} + 4s^2 f_{yy}.$$

4.7. **Directional derivative and gradient vector.** Let $z = f(x, y)$ be a differentiable function. Recall that two tangent vectors to the surface are given by $\langle 1, 0, \frac{\partial f}{\partial x}\rangle$ and $\langle 0, 1, \frac{\partial f}{\partial y}\rangle$. The partial derivatives $\frac{\partial f}{\partial x}$ or $\frac{\partial f}{\partial y}$ can be viewed as the rate of change of z in the direction of $\langle 1, 0\rangle$ or $\langle 0, 1\rangle$. Let us now try to find the rate of change of z in an arbitrary direction $\mathbf{u} = \langle a, b\rangle$, where $\mathbf{u}$ is a unit vector, i.e., $a^2 + b^2 = 1$. Suppose that this rate of change is $D_{\mathbf{u}}f$, then $\langle a, b, D_{\mathbf{u}}f\rangle$ is also a tangent vector. Now since we have three vectors that are in the same plane, we must have

$$\langle a, b, D_{\mathbf{u}}f\rangle = a\langle 1, 0, \frac{\partial f}{\partial x}\rangle + b\langle 0, 1, \frac{\partial f}{\partial y}\rangle.$$

This implies

$$D_{\mathbf{u}}f = a\frac{\partial f}{\partial x} + b\frac{\partial f}{\partial y}.$$

We call $D_{\mathbf{u}}f$ the directional derivative of f in the direction of $\mathbf{u}$. If the unit vector $\mathbf{u} = \langle \cos\theta, \sin\theta \rangle$, then the above equation becomes

$$D_{\mathbf{u}}f = \cos\theta \frac{\partial f}{\partial x} + \sin\theta \frac{\partial f}{\partial y}.$$

Example 4.44. Find the directional derivative of the function $f = x^3 - 3xy + 4y^2$ in the direction of $\mathbf{u}$, where $\mathbf{u} = \langle \cos\frac{\pi}{6}, \sin\frac{\pi}{6} \rangle$.

Solution. We compute this directly.

$$\begin{aligned} D_{\mathbf{u}}f &= \cos\frac{\pi}{6} f_x + \sin\frac{\pi}{6} f_y \\ &= \frac{\sqrt{3}}{2}(3x^2 - 3y) + \frac{1}{2}(-3x + 8y) \\ &= \frac{1}{2}(3\sqrt{3}x^2 - 3x + (8 - 3\sqrt{3})y). \end{aligned}$$

Note that the equation $D_{\mathbf{u}}f = a\frac{\partial f}{\partial x} + b\frac{\partial f}{\partial y}$ is the dot product between the vectors

$$D_{\mathbf{u}}f = \langle \frac{\partial f}{\partial x}, \frac{\partial f}{\partial y} \rangle \cdot \langle a, b \rangle = \langle \frac{\partial f}{\partial x}, \frac{\partial f}{\partial y} \rangle \cdot \mathbf{u}.$$

We give a special name to the vector $\langle \frac{\partial f}{\partial x}, \frac{\partial f}{\partial y} \rangle$.

Definition 4.45. Let $z = f(x, y)$ be a differentiable function. We define the gradient vector of f to be a vector function ∇f where

$$\nabla f = \langle \frac{\partial f}{\partial x}, \frac{\partial f}{\partial y} \rangle.$$

And in this case

$$D_{\mathbf{u}}f = \nabla f \cdot \mathbf{u}$$

Example 4.46. If $z = \sin x + e^{xy}$, then

$$\nabla z = \langle \cos x + ye^{xy}, xe^{xy} \rangle.$$

In particular, if we consider the point $(x, y) = (0, 1)$, then

$$\nabla z(0, 1) = \langle 2, 0 \rangle.$$

Example 4.47. Let $f(x, y) = x^2y^3 - 4y$. Find ∇f and use it to find the directional derivative of f at $(2, -1)$ in the direction of $\mathbf{v} = \langle 2, 5 \rangle$.

Solution. We first compute the gradient.

$$\nabla f = \langle 2xy^3, 3x^2y^2 - 4 \rangle.$$

At $(2, -1)$,

$$\nabla f(2, -1) = \langle -4, 8 \rangle.$$

Now the unit vector in the direction of $\langle 2, 5 \rangle$ is $\mathbf{u} = \langle \frac{2}{\sqrt{29}}, \frac{5}{\sqrt{29}} \rangle$. Therefore the directional derivative is

$$D_{\mathbf{u}}f(2, -1) = \nabla f(2, -1) \cdot \mathbf{u} = \frac{32}{\sqrt{29}}.$$

We may generalize the definition of the gradient vector and directional derivative to higher dimensions. In the case of function of three variables, if $f = f(x, y, z)$, we define

$$\nabla f = \langle \frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, \frac{\partial f}{\partial z} \rangle.$$

And if $\mathbf{u}$ is a unit vector, we define $D_{\mathbf{u}} f = \nabla f \cdot \mathbf{u}$. One significance of the gradient vector is that it tells us the direction of maximal rate of change. In fact, since

$$D_{\mathbf{u}} f = \nabla f \cdot \mathbf{u} = |\nabla f||\mathbf{u}| \cos\theta = |\nabla f| \cos\theta,$$

where θ is the angle between ∇f and $\mathbf{u}$ and $|\mathbf{u}| = 1$. The maximal value of $D_{\mathbf{u}} f$ occurs when $\theta = 0$, that is, when $\mathbf{u}$ is in the same direction as ∇f. And the maximal value is $|\nabla f|$.

Example 4.48. Let $f = xe^y$. Find the rate of change of f at $(2, 0)$ in the direction from $(2, 0)$ to $(\frac{1}{2}, 2)$. Find the maximal rate of change and the direction.

Solution. We first compute the gradient vector.

$$\nabla f = \langle e^y, xe^y \rangle.$$

At $(2, 0)$,

$$\nabla f(2, 0) = \langle 1, 2 \rangle.$$

The unit vector $\mathbf{u}$ in the direction from $(2, 0)$ to $(\frac{1}{2}, 2)$ is $\langle -\frac{3}{5}, \frac{4}{5} \rangle$, so the rate of change is

$$D_{\mathbf{u}} f = \langle 1, 2 \rangle \cdot \langle -\frac{3}{5}, \frac{4}{5} \rangle = 1.$$

On the other hand, according to the above discussion, the maximal rate of change is in the direction of $\nabla f(2, 0) = \langle 1, 2 \rangle$, or $\langle \frac{1}{\sqrt{5}}, \frac{2}{\sqrt{5}} \rangle$, if we represent it using a unit vector. And the value of the maximal rate of change is $|\nabla f| = \sqrt{5}$.

One property of the gradient vector is that geometrically, ∇f is always perpendicular to the level curves $f(x, y) = k$. In fact, if the level curve is parameterized by some vector valued function $\mathbf{r}(t) = \langle x(t), y(t) \rangle$, then $f(x(t), y(t)) = k$. Differentiate this with respect to t we obtain that

$$\frac{\partial f}{\partial x} x'(t) + \frac{\partial f}{\partial y} y'(t) = 0,$$

or

$$\nabla f \cdot \mathbf{r}'(t) = 0,$$

which shows that ∇f is perpendicular to $\mathbf{r}'(t)$. One may also compare this equation with the differential formula df. An application of this property is that in the case of three variables, if a surface if given by some implicit equation $F(x, y, z) = 0$, for example, an ellipsoid $x^2 + 2y^2 + z^2 - 1 = 0$, then ∇F being perpendicular to the surface means it is a normal vector to the tangent plane. This allows us to write down the equation of the tangent plane using ∇F. If the plane passes through the point (a, b, c), then the equation is given by

$$\frac{\partial F}{\partial x}(x - a) + \frac{\partial F}{\partial y}(y - b) + \frac{\partial F}{\partial z}(z - c) = 0.$$

In the special case where the surface is given by $z = f(x, y)$, we may rewrite it as $f(x, y) - z = 0$ and let $F = f(x, y) - z$, then $\nabla F = \langle f_x, f_y, -1 \rangle$ and the equation of the tangent plane becomes

$$\frac{\partial f}{\partial x}(x - a) + \frac{\partial f}{\partial y}(y - b) - (z - c) = 0,$$

which is the same as before.

Example 4.49. Consider an ellipsoid $\frac{x^2}{4} + y^2 + \frac{z^2}{9} = 3$. Find the equation of the tangent plane to the ellipsoid at $(-2, 1, -3)$.

Solution. We may view the ellipsoid as a level surface of the function $F = \frac{x^2}{4} + y^2 + \frac{z^2}{9}$ (at the level $k = 3$), and

$$\nabla F = \langle \frac{x}{2}, 2y, \frac{2z}{9} \rangle.$$

At $(-2, 1, -3)$, the gradient is

$$\nabla F(-2, 1, -3) = \langle -1, 2, -\frac{2}{3} \rangle.$$

Therefore, the equation of the tangent line is

$$-1(x + 2) + 2(y - 1) - \frac{2}{3}(z + 3) = 0,$$

or $3x - 6y + 2z + 18 = 0$.

4.8. **Maximum and minimum values, Taylor polynomial.** Our last topic in this chapter is the maximum and minimum values of multi-variable functions. In this section, we shall see how partial derivative allows us to find these values.

Definition 4.50. Let $f(x, y)$ be a function of two variables. We say $f(x, y)$ has a local maximum at (a, b) if $f(x, y) \leq f(a, b)$ at any (x, y) near (a, b). Similarly, We say $f(x, y)$ has a local minimum at (a, b) if $f(x, y) \geq f(a, b)$ at any (x, y) near (a, b).

Definition 4.51. If in the above definition, the inequalities hold for all (x, y) in the domain of the function f, then we say f has an absolute maximum or absolute minimum at (a, b).

If f is differentiable and has a local maximum or local minimum at some point (a, b), then by Fermat's theorem, its partial derivatives must be 0 at (a, b). That is,

Theorem 4.52. *If f has a local maximum or local minimum at (a, b) and its first-order partial derivative exists, then $f_x(a, b) = f_y(a, b) = 0$.*

Geometrically, f has a horizontal tangent plane at these local extreme value points.

Therefore, similar to the case of single variable, to locate a potential local maximum or minimum point of f, we may solve the equations $f_x = 0$ and $f_y = 0$.

Definition 4.53. We say a point (a, b) is a critical point, if $f_x(a, b) = 0$ and $f_y(a, b) = 0$, or if one of these partial derivative does not exist.

Note that a critical point might be a local maximum or minimum point, or it can be neither a local maximum nor a local minimum point.

Example 4.54. Let $f(x, y) = x^2 + y^2 - 2x - 6y + 14$. Then f is differentiable everywhere in $\mathbb{R}^2$. So the critical points are when $f_x = f_y = 0$. Therefore,

$$f_x = 2x - 2 = 0$$
$$f_y = 2y - 6 = 0,$$

implies there is only one critical point of f which is $(1, 3)$. On the other hand, $f = (x - 1)^2 + (y - 3)^2 + 4$, we see that f has a local minimum, and in fact, an absolute minimum, at the point $(1, 3)$.

Example 4.55. Consider $f(x, y) = y^2 - x^2$. The function f is differentiable in the entire $\mathbb{R}^2$ and therefore we solve $f_x = 0$, $f_y = 0$ for the critical points.

$$f_x = -2x = 0, f_y = 2y = 0.$$

The only critical point is $(0, 0)$. However, in this case, $(0, 0)$ is neither a local maximum nor a local minimum of f, since $f(0, 0) = 0$ and if we check the points on the x-axis, then $f(x, 0) = -x^2 < 0$ and along y-axis, $f(0, y) = y^2 > 0$.

The critical point in the above example is called a saddle point. If the function $f(x, y)$ has continuous second order partial derivatives, then we may determine whether a critical point is a local minimum or maximum using the second derivative test.

The idea of the second derivative test is similar to the case of single variable. In the case of single variable, the sign of the second derivative $f''(x)$ indicates whether the graph of $f(x)$ is concave up or down. In the case of function of two variable, it is not sufficient to just look at the sign of the second derivatives f_{xx} or f_{yy} since we have learned that the partial derivatives are just the rate of change in two directions. In order for a critical point to be a local minimum or a local maximum, the curve which is the intersection of the surface $z = f(x, y)$ and a vertical plane must be concave down or up for all possible directions.

To be more precisely, if $\mathbf{u} = \langle u_1, u_2 \rangle$ is a unit vector. Consider a curve which is the intersection between the surface $z = f(x, y)$ and a vertical plane P that passes through $(a, b, f(a, b))$ and contains the vector $\mathbf{u}$. In this case, we may parameterize the curve by

$$\mathbf{r}(t) = \langle a + u_1 t, b + u_2 t, f(a + u_1 t, b + u_2 t) \rangle.$$

On the vertical plane P, the concavity of the curve $\mathbf{r}(t)$ is determined by the sign of the second derivative of $f(a + u_1 t, b + u_2 t)$. If we let $F(t) = f(a + u_1 t, b + u_2 t)$, then

$$F'(t) = u_1 \frac{\partial f}{\partial x} + u_2 \frac{\partial f}{\partial y},$$

and taking the derivative again we obtain that,

$$F''(t) = u_1^2 \frac{\partial^2 f}{\partial x^2} + 2u_1 u_2 \frac{\partial^2 f}{\partial x \partial y} + u_2^2 \frac{\partial^2 f}{\partial y^2}.$$

If (a, b) is a local maximum, then $F''(0) < 0$ for all directions $\langle u_1, u_2 \rangle$. And similarly, if (a, b) is a local minimum, then $F''(0) > 0$ for all directions. Now in order to obtain the second derivative test, it is convenient to denote the expression of $F''(t)$ using the following notation.

Definition 4.56. Let $f(x, y)$ be a function which has continuous second order partial derivatives. We call the matrix

$$H = \begin{pmatrix} \frac{\partial^2 f}{\partial x^2} & \frac{\partial^2 f}{\partial y \partial x} \\ \frac{\partial^2 f}{\partial x \partial y} & \frac{\partial^2 f}{\partial y^2} \end{pmatrix}$$

the Hessian matrix of f.

Using matrix multiplication,

$$F'' = \mathbf{u} \cdot H \cdot \mathbf{u}^T$$

Now using the theory of linear algebra, $F'' > 0$ or $F'' < 0$ for all $\mathbf{u}$ if and only if H has two positive or two negative eigenvalues. In the case of 2×2 matrices, this is equivalent to the following theorem.

Theorem 4.57 (Second Derivative Test)**.** *Let $f(x, y)$ be a function with continuous second order partial derivatives, and $f_x(a, b) = f_y(a, b) = 0$. Let*

$$D = \det H = f_{xx} f_{yy} - f_{xy}^2.$$

Then

- *f has a local maximum at (a, b), if H has two negative eigenvalues, in other words,*
$$f_{xx} < 0, D > 0.$$
- *f has a local minimum at (a, b), if H has two positive eigenvalues, in other words,*
$$f_{xx} > 0, D > 0.$$
- *f has a saddle point at (a, b), if H has two eigenvalues of opposite signs, in other words,*
$$D < 0.$$
- *If $D = 0$, the test gives no information.*

Example 4.58. Find the local maximum and local minimum and saddle points of $f(x, y) = x^4 + y^4 - 4xy + 1$.

Solution. We first compute the critical points.

$$f_x = 4x^3 - 4y, f_y = 4y^3 - 4x.$$

Let $f_x = 0, f_y = 0$. We obtain that

$$x^3 - y = 0, y^3 - x = 0.$$

So

$$x^9 - x = 0 \Rightarrow x = 0, -1, 1,$$

and

$$y = x^3 = 0, -1, 1, \text{respectively}.$$

Therefore, there are three critical points

$$(0, 0), (-1, -1), (1, 1).$$

Now we apply the second derivative test. Let us first compute D.

$$f_{xx} = 12x^2, f_{xy} = -4, f_{yy} = 12y^2.$$

$$D = f_{xx} f_{yy} - f_{xy}^2 = 144x^2 y^2 - 16.$$

Now at $(0, 0)$, $D = -16 < 0$ and therefore it is a saddle point.
At $(-1, -1)$, $f_{xx}(-1, -1) = 12 > 0, D = 128 > 0$, so f has a local minimum at $(-1, -1)$.
And at $(1, 1)$, $f_{xx}(1, 1) = 12 > 0, D = 128 > 0$, so f also has a local minimum at $(1, 1)$.

The computation of the derivatives of $F(t)$ in fact allows us to obtain the Taylor series for function of two variables. Recall that $F(t) = f(a + u_1 t, b + u_2 t)$. Let us now consider the Taylor series of $F(t)$ at $t = 0$, that is

$$F(t) = F(0) + F'(0)t + \frac{1}{2!} F''(0)t^2 + \dots$$

If we view u_1, u_2 as changing variables, and let us denote by Δx, Δy, evaluating the above equation at $t = 1$, we obtain that

$$f(a+\Delta x, b+\Delta y) = f(a,b)+\frac{\partial f}{\partial x}\Delta x+\frac{\partial f}{\partial y}\Delta y+\frac{1}{2}\left(\frac{\partial^2 f}{\partial x^2}\Delta x^2 + 2\frac{\partial^2 f}{\partial x \partial y}\Delta x \Delta y + \frac{\partial^2 f}{\partial y^2}\Delta y^2\right)+\dots$$

This is called the Taylor series of $f(x, y)$ at (a, b). The polynomials

$$T_1 = f(a,b) + \frac{\partial f}{\partial x}(x-a) + \frac{\partial f}{\partial y}(y-b)$$

$$T_2 = f(a,b)+\frac{\partial f}{\partial x}(x-a)+\frac{\partial f}{\partial y}(y-b)+\frac{1}{2}\left(\frac{\partial^2 f}{\partial x^2}(x-a)^2 + 2\frac{\partial^2 f}{\partial x \partial y}(x-a)(y-b) + \frac{\partial^2 f}{\partial y^2}(y-b)^2\right),$$

$$\dots$$

are called the Taylor polynomials of degree $1, 2, \dots,$ of f. In fact, T_1 is the linear approximation of f and T_2, from this point of view, can be regarded as an approximation of f using quadratic polynomials.

4.9. **Absolute maximum and minimum, Lagrange Multipliers.** Let us now consider the absolute maximum and minimum of a function f. For a function of one variable, if f is continuous on a closed interval, then the Extreme value theorem says f has an absolute minimum value and an absolute maximum value on this interval. This can be generalized to the case of two variables. Recall that a closed set in $\mathbb{R}^2$ is a set that contains all its boundary points. Usually these sets can be described by $\geq$ or $\leq$. We say a set is bounded if it is finite in extent. In other words it can be contained within some disk.

Theorem 4.59 (Extreme Value Theorem)**.** *If $f(x, y)$ is continuous on a closed bounded set D in $\mathbb{R}^2$, then f has an absolute maximum and an absolute minimum in D.*

Note that since an absolute maximum or minimum must also be a local maximum or minimum, if f is a differentiable function, to find the extreme values, we may use the following method.

Proposition 4.60. *Let f be a differentiable function on a closed bounded set D. Then to find the absolute maximum or minimum:*

- *Find the values of f at all critical points in the interior of D.*
- *Find the extreme values of f on the boundary of D.*
- *The largest value from the above steps is the absolute maximum. The smallest value is the absolute minimum.*

Example 4.61. Find the absolute maximum and minimum values of the function $f = x^2 - 2xy + 2y$ on $D = \{(x, y)|0 \leq x \leq 3, 0 \leq y \leq 2\}$.

Solution. We first find the critical point of f inside D.

$$f_x = 2x - 2y = 0, f_y = -2x + 2 = 0 \Rightarrow x = 1, y = 1.$$

There is one critical point $(1, 1)$ inside D. And $f(1, 1) = 1$. Now to obtain the absolute extreme value, we have to find the extreme value of f on the boundary of D. The boundary consists of 4 segments.

- $y = 0, 0 \leq x \leq 3$.
  We have $f(x, 0) = x^2$, and its maximum is $f(3, 0) = 9$ and minimum $f(0, 0) = 0$.

- $x = 3, 0 \leq y \leq 2$.
  We have $f(3, y) = 9-4y$, and its maximum is $f(3, 0) = 9$ and minimum $f(3, 2) = 1$.
- $y = 2, 0 \leq x \leq 3$.
  We have $f(x, 2) = x^2 - 4x + 4$, and and its maximum is $f(0, 2) = 4$ and minimum $f(2, 2) = 0$.
- $x = 0, 0 \leq y \leq 2$.
  We have $f(0, y) = 2y$, and its maximum is $f(0, 2) = 4$ and minimum $f(0, 0) = 0$.

Now if we compare all these values, we conclude that f has an absolute maximum at $(3, 0)$ which is 9 and f has absolute minimums at $(0, 0)$ and $(2, 2)$ which is 0.

One may also check that the critical point $(1, 1)$ is a saddle point.

To find extreme values on the boundary of a region can be complicated in some cases. Especially when the boundary of D is not just segments but is described by some equation of the curves. This problem in general can be described by finding the extreme value of the function $f(x, y)$, where x, y satisfies a constraint, that is, an equation $g(x, y) = k$, where k is some constant. In this case, we may use the Lagrange Multiplier method to obtain the extreme values.

The idea is the following. Let us consider the level curves $f = c$. To maximize or minimize f along $g = k$ is to maximize or minimize the value of c such that $f = c$ intersects $g = k$. This happens when the curve $g = k$ "touches" the curve $f = c$. In other words, $g = k$ and $f = c$ share the same tangent line at the tangent point. This implies that the normal vectors to the tangent line must be parallel. Recall that the gradient vector is perpendicular to the tangent line. Therefore,

Theorem 4.62 (Method of Lagrange Multipliers)**.** *To find the maximum and minimum values of $f(x, y)$ under the constraint $g(x, y) = k$, assuming the extreme value exists and $\nabla g \neq \mathbf{0}$,*

- *First find all values of (x, y) such that*

$$\nabla f = \lambda \nabla g$$
$$g(x, y) = k.$$

- *Evaluate f at all points find in the first step. The largest value is the maximum of f and the smallest is the minimum value.*

Example 4.63. Find the extreme values of the function $f(x, y) = x^2 + 2y^2$ on the circle $x^2 + y^2 = 1$.

Solution. In this case we may pick $g = x^2 + y^2$ and the constraint is $g = 1$. Using Lagrange multipliers, $\nabla f = \lambda \nabla g$ and $g = 1$, so that

$$f_x = \lambda g_x, f_y = \lambda g_y, x^2 + y^2 = 1 \Rightarrow$$
$$2x = 2x\lambda, 4y = 2y\lambda, x^2 + y^2 = 1.$$

From the first equation we have $x = 0$ or $\lambda = 1$. If $x = 0$, then $y = \pm 1$ and λ can be any number (which we don't care).

If $\lambda = 1$, then $y = 0$ and hence $x = \pm 1$. Therefore we obtained four possible points

$$(0, 1), (0, -1), (1, 0), (-1, 0).$$

Evaluating f at these points we obtain that the maximum value is $f(0, \pm 1) = 2$ and the minimum value is $f(\pm 1, 0) = 1$.

Example 4.64. Find the extreme value of $f = x^2 + 2y^2$ on the disk $D = \{x^2 + y^2 \leq 1\}$.

Solution. D is a closed region. To find the extreme value, we compute the critical points and compare the extreme value with the ones on the boundary.

Since $f_x = 2x, f_y = 4y$, the critical point inside D is $(0,0)$ and $f(0,0) = 0$. We have computed the extreme value of f on the boundary of D in the previous example. Now by comparing these values, we conclude that f has an absolute minimum 0 at $(0,0)$ and absolute maximum 2 at $(0, \pm 1)$.

Remark 4.65. Note that in this example, we do not have to apply the second derivative test to see $(0,0)$ is a local minimum. However, one may still check that using the second derivative test, $(0,0)$ is a local minimum. But note that a local minimum does not apply its is an absolute minimum. We still have to use the extreme value theorem methods to conclude this.

Theorem 4.62 can be generalized to function of three variables or with multiple constraints.

Theorem 4.66. *To find the maximum and minimum values of $f(x,y,z)$ under the constraint $g(x,y,z) = k$, assuming the extreme value exists and $\nabla g \neq \mathbf{0}$,*

- *First find all values of (x,y,z) such that*
$$\nabla f = \lambda \nabla g$$
$$g(x,y,z) = k.$$
- *Evaluate f at all points find in the first step. The largest value is the maximum of f and the smallest is the minimum value.*

Theorem 4.67. *To find the maximum and minimum values of $f(x,y,z)$ under the constraint $g(x,y,z) = k, h(x,y,z) = c$, assuming the extreme value exists and the gradients are non zero.*

- *First find all values of (x,y,z) such that*
$$\nabla f = \lambda \nabla g + \mu \nabla h$$
$$g(x,y,z) = k$$
$$h(x,y,z) = c.$$
- *Evaluate f at all points find in the first step. The largest value is the maximum of f and the smallest is the minimum value.*

Let us now investigate a few examples.

Example 4.68. A rectangular box without a lid is to be made from $12m^2$ of cardboard. Find the maximum volume of the box.

Solution. Assume that the length, width and height are x, y, z respectively. Then the volume $V = xyz$. We wish to maximize V under the constraint the total area of the five faces of the box (without lid) is 12. In other words,

$$g(x,y,z) = 2xz + 2yz + xy = 12.$$

Applying lagrange multipliers,

$$\nabla V = \lambda \nabla g$$
$$2xz + 2yz + xy = 12$$

which become

$$yz = \lambda(2z + y), xz = \lambda(2z + x), xy = \lambda(2x + 2y), 2xz + 2yz + xy = 12.$$

There is no general methods of solving the equation of Lagrange multipliers. In this example, we may multiply the first three equations by x, y, z respectively, and we obtain that

$$xyz = \lambda(2xz + xy), xyz = \lambda(2yz + xy), xyz = \lambda(2xz + 2yz).$$

Now if $\lambda = 0$, we would have $yz = zy = xy = 0$, which contradicts the last equation. So $\lambda \neq 0$. Therefore,

$$2xz + xy = 2yz + xy \Rightarrow xz = yz.$$

Note that $z \neq 0$, or otherwise $V = 0$. So $x = y$. Then

$$2yz + xy = 2xz + 2yz \Rightarrow y = 2z$$

Now finally, if we put $x = y = 2z$ into $g = 12$, we obtain that

$$4z^2 + 4z^2 + 4z^2 = 12,$$

and $z = 1$, $x = y = 2$, since x, y, z must be positive in this question. The maximal volume is 4.

Example 4.69. Find the points on the sphere $x^2 + y^2 + z^2 = 4$ that are closet to the point $(3, 1, -1)$.

Solution. The distance between a point (x, y, z) and $(3, 1, -1)$ is

$$\text{dist} = \sqrt{(x-3)^2 + (y-1)^2 + (z+1)^2}$$

It is simpler if we minimize d^2 instead. Let $f = d^2 = (x-3)^2 + (y-1)^2 + (z+1)^2$. In this problem we would like to minimize f while (x, y, z) is on the sphere. In other words,

$$x^2 + y^2 + z^2 = 4.$$

Let $g = x^2 + y^2 + z^2$. Then the constraint is $g = 4$. According to the method of Lagrange Multipliers, we solve the equations $\nabla f = \lambda \nabla g$ and $g = 4$.

$$2(x-3) = 2x\lambda$$
$$2(y-1) = 2y\lambda$$
$$2(z+1) = 2z\lambda$$
$$x^2 + y^2 + z^2 = 4$$

From the first equation we have

$$x(1-\lambda) = 3 \Rightarrow x = \frac{3}{1-\lambda}.$$

Note that here $\lambda \neq 1$, otherwise we would have $0 = 3$, which is impossible. Similarly, we obtain from the other two equations that

$$y = \frac{1}{1-\lambda}, z = -\frac{1}{1-\lambda}.$$

Now the last equation tells us

$$\frac{9}{(1-\lambda)^2} + \frac{1}{(1-\lambda)^2} + \frac{1}{(1-\lambda)^2} = 4$$

$$(1-\lambda)^2 = \frac{11}{4}$$
$$\lambda = 1 \pm \frac{\sqrt{11}}{2}.$$

Therefore, we got two points

$$\left(\frac{6}{\sqrt{11}}, \frac{2}{\sqrt{11}}, -\frac{2}{\sqrt{11}}\right), \left(-\frac{6}{\sqrt{11}}, -\frac{2}{\sqrt{11}}, \frac{2}{\sqrt{11}}\right)$$

Now one may check that f has smaller value on $\left(\frac{6}{\sqrt{11}}, \frac{2}{\sqrt{11}}, -\frac{2}{\sqrt{11}}\right)$. Therefore, it is the closest point.

Example 4.70. Find the maximum value of $f = x + 2y + 3z$ on the curve of intersection of $x - y + z = 1$ and $x^2 + y^2 = 1$.

Solution. Let $g = x - y + 1$ and $h = x^2 + y^2$. We maximize f under the constraint $g = 1$ and $h = 1$. According to the method of Lagrange multiplier,

$$\nabla f = \lambda \nabla g + \mu \nabla h, g = 1, h = 1.$$

That is,

$$1 = \lambda + 2x\mu, 2 = -\lambda + 2y\mu, 3 = \lambda$$
$$x - y + z = 1, x^2 + y^2 = 1$$

Thus, we have

$$x = -\frac{1}{\mu}, y = \frac{5}{2\mu}$$
$$\frac{1}{\mu^2} + \frac{25}{4\mu^2} = 1$$

So $\mu = \pm\frac{\sqrt{29}}{2}$. Therefore, we have $x = \mp\frac{2}{\sqrt{29}}$ and $y = \pm\frac{5}{\sqrt{29}}$. So $z = 1 - x + y = 1 \pm \frac{7}{\sqrt{29}}$. The corresponding values of f are $3 \pm \sqrt{29}$ and therefore, the maximum is $3 + \sqrt{29}$.

Our last example of this chapter is the Taylor polynomials.

Example 4.71. Let $f(x, y) = xe^y$. Find the first and second order Taylor polynomial for f at $(1, 0)$. Compare the value of f and the Taylor polynomials at $(0.9, 0.1)$.

The first order Taylor polynomial is in fact the linear approximation.

$$f_x = e^y, f_y = xe^y.$$

At $(1, 0)$, the values are

$$f(1, 0) = 1, f_x(1, 0) = 1, f_y(1, 0) = 1$$

Therefore,

$$T_1 = f(1, 0) + f_x(1, 0)(x - 1) + f_y(1, 0)y = 1 + (x - 1) + y = x + y.$$

To obtain T_2, we compute

$$f_{xx} = 0, f_{xy} = e^y, f_{yy} = xe^y.$$

At $(1, 0)$,

$$f_{xx}(1, 0) = 0, f_{xy}(1, 0) = 1, f_{yy}(1, 0) = 1.$$

Therefore,

$$T_2 = T_1 + \frac{1}{2}f_{xx}(1,0)(x-1)^2 + f_{xy}(1,0)(x-1)y + \frac{1}{2}f_{yy}(1,0)y^2 = x + y + (x-1)y + \frac{1}{2}y^2.$$

Now at $(0.9, 0.1)$, $x - 1 = -0.1, y = 0.1$. So

$$f \approx 0.99465, T_1 = 1, T_2 = 0.995$$

5. Integrals

5.1. **Double integral over rectangles.** Double integral is used to find the volume of a solid. To see how double integral is defined, let us first recall the definition of the definite integral for single variable functions.

If $f(x)$ is a single variable function defined over the interval $[a, b]$, then we start by dividing $[a, b]$ into n small subintervals $[x_i, x_{i+1}]$ of length $\Delta x = (b - a)/n$. And in each interval, we choose a sample point x_i^*. Then the Riemann sum is defined to be the sum of the areas of the rectangle with height $f(x_i^*)$. That is,

$$\sum_{i=1}^{n} f(x_i^*)\Delta x.$$

And the limit when $n \to \infty$ defines the definite integral,

$$\int_a^b f(x)dx = \lim_{n\to\infty} \sum_{i=1}^{n} f(x_i^*)\Delta x$$

Geometrically, if the function is non-negative, then it represents the area bounded by the curve $y = f(x)$, the vertical lines $x = a$, $x = b$ and the x-axis. One may generalize the idea to the case of multi-variables.

Note that one difference between the case of single-variable and multi-variables is that the domain of a function $f(x, y)$ in general can be very complicated. Therefore, let us first start on the simplest region rectangles in $\mathbb{R}^2$. Consider a function $f(x, y)$ defined on a closed rectangle $R = [a, b] \times [c, d]$. Let us assume that $f \geq 0$. Let us denote by S the solid region bounded between $f(x, y)$ and the xOy-plane. Our goal is to define an integral which computes the volume of S.

The first step is to subdivide R into subrectangles. In fact, since R is a product of the intervals $[a, b]$ and $[c, d]$, we may subdivide R by subdividing each of these intervals. Suppose that $[a, b]$ is divided into m intervals $[x_i, x_{i+1}]$ of length $\Delta x = (b - a)/m$ and $[c, d]$ is divided into n intervals $[y_i, y_{i+1}]$ of length $\Delta y = (c - d)/n$. We call the product $R_{ij} = [x_i, x_{i+1}] \times [y_i, y_{i+1}]$, which is a subrectangle in R. The area of this subrectangle is

$$\Delta A = \Delta x \Delta y.$$

Now similar to the case of single variable functions, in each subrectangle, we choose a sample point (x_{ij}^*, y_{ij}^*). The height of a small column which approximates a portion of S is given by $f(x_{ij}^*, y_{ij}^*)$. And its volume is given by

$$f(x_{ij}^*, y_{ij}^*)\Delta A.$$

In this way, the volume of the entire solid S can be approximated by the sum

$$V \approx \sum_{i=1}^{m} \sum_{j=1}^{n} f(x_{ij}^*, y_{ij}^*) \Delta A.$$

And in the limit case,

$$V = \lim_{m,n \to \infty} \sum_{i=1}^{m} \sum_{j=1}^{n} f(x_{ij}^*, y_{ij}^*) \Delta A.$$

Definition 5.1. We define the double integral of f over the rectangular region R to be

$$\iint_R f(x,y)dA = \lim_{m,n \to \infty} \sum_{i=1}^{m} \sum_{j=1}^{n} f(x_{ij}^*, y_{ij}^*) \Delta A.$$

We say f is integrable if the above limit exists.

Example 5.2. Estimate the volume of the solid that lies above the square $R = [0,2] \times [0,2]$ and below $z = 16 - x^2 - 2y^2$. Divide R into four equal squares and choose the sample point to be the upper right corner of each small squares.

Solution. The area of each small square is $\Delta A = 1$. The sample points are $(1,1), (1,2), (2,1), (2,2)$. So

$$V \approx f(1,1)\Delta A + f(1,2)\Delta A + f(2,1)\Delta A + f(2,2)\Delta A = 13 + 7 + 10 + 4 = 34.$$

Example 5.3. If $R = [-1,1] \times [-2,2]$, compute the integral $\iint_R \sqrt{1-x^2}dA$.

Solution. This integral would be very difficult to compute directly. However, since $\sqrt{1-x^2}$ is non-negative, the integral gives us the volume of the half-cylinder bounded by R and $z = \sqrt{1-x^2}$. The radius of the base of the cylinder is 1 and the height is 4. Therefore,

$$\iint_R \sqrt{1-x^2}dA = \frac{1}{2}\pi \cdot 1^2 \cdot 4 = 2\pi.$$

The evaluation of a double integral can be done by calculating two "iterated" single variable integrals. Suppose that $f(x,y)$ is a function defined on a rectangle $R = [a,b] \times [c,d]$. We define the partial integration with respect to y to be

$$A(x) = \int_c^d f(x,y)dy.$$

In this case, we regard x as constant in this computation. The resulting integral is a expression $A(x)$ which depends on x. Now we preform the second integral

$$\int_a^b A(x)dx = \int_a^b \left(\int_c^d f(x,y)dy \right) dx.$$

This expression is called an iterated integral, and we usually write

$$\int_a^b \int_c^d f(x,y)dydx = \int_a^b \left(\int_c^d f(x,y)dy \right) dx.$$

Similarly, one can consider the partial integral with respect to x and define

$$\int_c^d \int_a^b f(x,y)dxdy = \int_c^d \left(\int_a^b f(x,y)dx \right) dy.$$

Example 5.4. Compute $\int_0^3 \int_1^2 x^2 y dy dx$ and $\int_1^2 \int_0^3 x^2 y dx dy$.

Solution. In the first integral, we first integrate with respect to y,

$$\int_1^2 x^2 y dy = x^2 \frac{1}{2} y^2 \Big|_{y=1}^{y=2} = \frac{3}{2} x^2.$$

We then integrate with respect to x,

$$\int_0^3 \frac{3}{2} x^2 dx = \frac{1}{2} x^3 \Big|_{x=0}^{x=3} = \frac{27}{2}.$$

In other words,

$$\int_0^3 \int_1^2 x^2 y dy dx = \frac{27}{2}.$$

Now to compute the second integral, we first integrate with respect to x and then y.

$$\int_1^2 \int_0^3 x^2 y dx dy = \int_1^2 \frac{x^3}{3} y \Big|_{x=0}^{x=3} dy = \int_1^2 9y dy = \frac{27}{2}.$$

In fact, it is not a coincidence that the integrals in the above example gives us the same result.

Theorem 5.5 (Fubini)**.** *If f is continuous over the rectangle $R = [a, b] \times [c, d]$, then*

$$\iint_R f(x, y) dA = \int_a^b \int_c^d f(x, y) dy dx = \int_c^d \int_a^b f(x, y) dx dy.$$

This in fact holds for the case that f is bounded and is discontinuous only on a finite number of smooth curves, and the double integral exists.

Fubini's theorem allows us to explicitly compute a double integral, if we know how to compute the single variable integrals.

Example 5.6. Evaluate the integral $\iint_R y \sin(xy) dA$, where $R = [1, 2] \times [0, \pi]$.

Solution. Perhaps we can start with the partial integral with respect to y

$$\iint_R y \sin(xy) dA = \int_1^2 \int_0^\pi y \sin(xy) dy dx$$

To evaluate this integral, we have to use integrate by part. However, before we proceed, we should check the other iterated integral.

$$\iint_R y \sin(xy) dA = \int_0^\pi \int_1^2 y \sin(xy) dx dy.$$

This looks much easier,

$$\int_0^\pi \int_1^2 y \sin(xy) dx dy = \int_0^\pi (-\cos(xy))|_{x=1}^{x=2} dy = \int_0^\pi (-\cos 2y + \cos y) dy$$

$$= \frac{1}{2} \sin 2y + \sin y \Big|_{y=0}^{y=\pi} = 0.$$

Example 5.7. Find the volume of the solid that lies above the square $R = [0,2] \times [0,2]$ and below $z = 16 - x^2 - 2y^2$.

Solution. The volume is given by

$$V = \iint_R 16 - x^2 - 2y^2 dA = \int_0^2 \int_0^2 (16 - x^2 - 2y^2) dx dy$$
$$= \int_0^2 (16x - \frac{1}{3}x^3 - 2y^2 x)|_{x=0}^{x=2} dy$$
$$= \int_0^2 \frac{88}{3} - 4y^2 dy = 48.$$

5.2. **Double integral over general region.** Suppose that we would like to compute the volume of a solid over a general bounded closed region D instead of a rectangle. One way to define this double integral is the as follows. Suppose that $f(x,y)$ is an integrable function over a bounded finite region D. Let us consider an extension of the function. Pick R to be a rectangle that covers D. We define

$$F(x,y) = \begin{cases} f(x,y), & (x,y) \in D \\ 0, & (x,y) \in R - D. \end{cases}$$

Then the double integral of F over R should give us the same "volume" as f over D.

Definition 5.8. We define the double integral

$$\iint_D f(x,y) dA = \iint_R F(x,y) dA.$$

Now since the double integral is defined by integral over rectangles, we may use the iterated integral to evaluate it. In fact, by the generalized Fubini's theorem, the set of points of discontinuity of F must be contained in the points of discontinuity of f plus the boundary of D. In this case, depending on how the region D looks like, we may apply the iterated integrals as follows. Note that the integral of 0 is always 0.

Type I Region

We say a region D is of Type I if

$$D = \{(x,y) | a \leq x \leq b, g_1(x) \leq y \leq g_2(x)\}.$$

In this case,

$$\iint_D f(x,y) dA = \iint_R F(x,y) dA = \int_a^b \int_c^d F(x,y) dy dx$$

Because for each fixed x, $F(x,y) = 0$ if $y > g_2(x)$ or $y < g_1(x)$, so

$$\int_c^d F(x,y) dy = \int_{g_1(x)}^{g_2(x)} F(x,y) dy = \int_{g_1(x)}^{g_2(x)} f(x,y) dy.$$

As a result,

$$\iint_D f(x,y) dA = \int_a^b \int_{g_1(x)}^{g_2(x)} f(x,y) dy dx.$$

Type II Region

Simiplar to the case of Type I region, we say a region D is of Type II if

$$D = \{(x, y)|h_1(y) \leq x \leq h_2(y), c \leq y \leq d\}.$$

In this case,

$$\iint_D f(x, y)dA = \iint_R F(x, y)dA = \int_c^d \int_a^b F(x, y)dxdy$$

Now for each fixed y, $F(x, y) = 0$ if $x > h_2(y)$ or $x < h_1(y)$, so

$$\int_a^b F(x, y)dy = \int_{h_1(y)}^{h_2(y)} F(x, y)dx = \int_{h_1(y)}^{h_2(y)} f(x, y)dx.$$

As a result,

$$\iint_D f(x, y)dA = \int_c^d \int_{h_1(y)}^{h_2(y)} f(x, y)dxdy.$$

Remark 5.9. A Type I region can also be a Type II region at the same time. And there are regions which are neither Type I nor Type II.

Example 5.10. Evaluate $\iint_D (x + 2y)dA$, where D is the region bounded between $y = 2x^2$ and $y = 1 + x^2$.

Solution. To compute the integral, the first step is to find out whether this is a Type I or Type II region. In this case, we see that the two parabolas intersect when

$$2x^2 = 1 + x^2 \Rightarrow x = \pm 1$$

This is in fact a Type I region and not a Type II region. We can write

$$D = \{(x, y)| -1 \leq x \leq 1, 2x^2 \leq y \leq 1 + x^2\}.$$

Therefore,

$$\begin{aligned}\iint_D (x + 2y)dA &= \int_{-1}^1 \int_{2x^2}^{1+x^2} (x + 2y)dydx \\ &= \int_{-1}^1 (xy + y^2)|_{y=2x}^{y=1+x^2} dx \\ &= \int_{-1}^1 (-3x^4 - x^3 + 2x^2 + x + 1)dx = \frac{32}{15}.\end{aligned}$$

Sometimes a region can be both Type I and Type II.

Example 5.11. Evaluate $\iint_D x^2 + y^2 dA$ where D is the region bounded by $y = 2x$ and $y = x^2$.

We will solve this problem in two different ways. *Solution.* Let us first find out the intersection of these two curves. Let $2x = x^2$, we obtain that $x = 0, 2$. So the intersection points are $(0, 0)$ and $(2, 4)$.

Now if we view D as a Type I region, then it can be described by

$$\{(x, y)|0 \leq x \leq 2, x^2 \leq y \leq 2x\}.$$

So that in this case,

$$\iint_D x^2 + y^2 dA = \int_0^2 \int_{x^2}^{2x} x^2 + y^2 dydx$$

$$= \int_0^2 (x^2y + y^3/3)_{y=x^2}^{y=2x} dx$$

$$= \int_0^2 \left(-\frac{x^6}{3} - x^4 + \frac{14x^3}{3}\right) dx = \frac{216}{35}.$$

On the other hand, D may be also expresses as a Type II region,

$$D = \{(x, y)|0 \leq y \leq 4, \frac{y}{2} \leq x \leq \sqrt{y}\}$$

And

$$\iint_D x^2 + y^2 dA = \int_0^4 \int_{y/2}^{\sqrt{y}} x^2 + y^2 dx dy$$

$$= \int_0^4 \left(\frac{y^{3/2}}{3} + y^{5/2} - \frac{y^3}{24} - \frac{y^3}{2}\right) dy = \frac{216}{35}.$$

In some cases, it might be difficult to integrate a function over a certain type of the region. But we may rewrite the region in the other type to compute the integral.

Example 5.12. Evaluate $\int_0^1 \int_x^1 \sin(y^2) dy dx$.

Solution. It is very difficult to integrate $\sin(y^2)$ with respect to y. However, it is easy to integrate with respect to x since it is just a constant with respect to x. In this case, we have to first write down the region and then rewrite it as the other type so that we can switch the order of the iterated integral.

$$D = \{(x, y)|0 \leq x \leq 1, x \leq y \leq 1\}.$$

Now rewriting D as a Type II region, it is

$$D = \{(x, y)|0 \leq y \leq 1, 0 \leq x \leq y\}.$$

Now using Fubini's theorem,

$$\int_0^1 \int_x^1 \sin(y^2) dy dx = \int_0^1 \int_0^y \sin(y^2) dx dy = \int_0^1 y \sin(y^2) dy = \frac{1}{2}(1 - \cos 1).$$

Finally, to evaluate a double integral on a general region which is neither Type I nor Type II, let us study some properties of the integral. Because the integral is defined as a limit, if the integrals exists, then

$$\iint_D (f(x, y) + g(x, y)) dA = \iint_D f(x, y) dA + \iint g(x, y) dA.$$

$$\iint_D cf(x, y) dA = c \iint_D f(x, y) dA.$$

The next property actually shows us a way to evaluate integral over general region.

Proposition 5.13. *If $D = D_1 \cup D_2$, where D_1 and D_2 don't intersect except for their boundaries, then*

$$\iint_D f(x, y) dA = \iint_{D_1} f(x, y) dA + \iint_{D_2} f(x, y) dA.$$

Therefore, if a region D is neither Type I nor Type II, we subdivide it into small regions until each region is either Type I or Type II and then we evaluate the integral.

The next two properties are useful when we would like to estimate the integrals.

Proposition 5.14. *The integral of the constant function $f = 1$ over a region D is the area of the region.*

$$\text{area}(D) = \iint_D 1 dA.$$

Proposition 5.15. *If $f(x, y) \geq g(x, y)$ on a region D, then $\iint_D f(x, y)dA \geq \iint_D g(x, y)dA$. And if $m \leq f(x, y) \leq M$, then*

$$m \cdot \text{area}(D) \leq \iint_D f(x, y)dA \leq M \cdot \text{area}(D).$$

5.3. **Double integral in polar coordinates.** Sometimes it is easier to use the polar coordinate to describe a region. Recall that the polar coordinates (r, θ) of a point are related to the Cartesian coordinates (x, y) by the equations

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

$$r^2 = x^2 + y^2, \tan\theta = \frac{y}{x}.$$

We consider a polar rectangle in the plane. That is, a region R with

$$R = \{(r, \theta)|a \leq r \leq b, \alpha \leq \theta \leq \beta\}.$$

Let us compute the double integral over the region R. Similar to the case of double integral over rectangular region, in this case, we consider a partition of the region in to subrectangles. We divide $[a, b]$ into m intervals $[r_{i-1}, r_i]$ and $[c, d]$ into n intervals $[\theta_{i-1}, \theta_i]$. Suppose that $\Delta r = (b - a)/m$ and $\Delta\theta = (\beta - \alpha)/n$. Then a subrectangle is $R_{ij} = \{(r, \theta)|r_{i-1} \leq r \leq r_i, \theta_{j-1} \leq \theta \leq \theta_j\}$.

The area of the subractangle can be computed in the following way. In fact, since R_{ij} is in between two sectors of the circles of radius r_{i-1} and r_i, and the angle of these sectors are both $\Delta\theta = \theta_j - \theta_{j-1}$. Therefore, the area is

$$\Delta A_i = \frac{1}{2}r_i^2\Delta\theta - \frac{1}{2}r_{i-1}^2\Delta\theta.$$

$$= \frac{1}{2}(r_i + r_{i-1})(r_i - r_{i-1})\Delta\theta = \frac{1}{2}(r_i + r_{i-1})\Delta r\Delta\theta.$$

If we denote by $r_i^* = \frac{1}{2}(r_i + r_{i-1})$, then $\Delta A_i = r_i^*\Delta r\Delta\theta$. Note that r_i^* is the middle point of the interval $[(r_{i-1}, r_i)]$. Therefore, we can pick it to be the sample point. Now the double integral $\iint_R f(x, y)dA$ can be computed using the Riemann sum

$$\sum_{i=1}^{m}\sum_{j=1}^{n} f(r_i^*\cos\theta_j^*, r_i^*\sin\theta_j^*)\Delta A_i = \sum_{i=1}^{m}\sum_{j=1}^{n} f(r_i^*\cos\theta_j^*, r_i^*\sin\theta_j^*)r_i^*\Delta r\Delta\theta.$$

If we take the limit $n, m \to \infty$, this becomes

$$\iint_R f(x, y)dA = \iint_R f(r\cos\theta, r\sin\theta)rdrd\theta.$$

Theorem 5.16. *If f is a continuous function on a polar rectangle R with $a \le r \le b$ and $\alpha \le \theta \le \beta$, and $0 \le \beta - \alpha \le 2\pi$, then*

$$\iint_R f(x, y)dA = \int_\alpha^\beta \int_a^b f(r\cos\theta, r\sin\theta)rdrd\theta.$$

Remark 5.17. One can view $dA = rdrd\theta$.

Example 5.18. Compute $\iint_R (3x + 4y^2)dA$, where R is the region in the upper half plane between the circles $x^2 + y^2 = 1$ and $x^2 + y^2 = 4$.

Solution. Using the polar coordinate, the region R can be written as

$$R = \{(r, \theta)|1 \le r \le 2, 0 \le \theta \le \pi\}.$$

Therefore, using the above formula, the integral can be computed by

$$\begin{aligned}\iint_R f(x, y)dA &= \int_0^\pi \int_1^2 (3r\cos\theta + 4(r\sin\theta)^2)rdrd\theta \\ = \int_0^\pi r^3\cos\theta + r^4\sin^2\theta|_{r=1}^{r=2} &= \int_0^\pi (7\cos\theta + 15\sin^2\theta)d\theta \\ = 7\sin\theta + \frac{15}{2}\theta - \frac{15}{4}\sin 2\theta|_{\theta=0}^{\theta=\pi} &= \frac{15\pi}{2}.\end{aligned}$$

Example 5.19. Find the volume of the solid between the plane $z = 0$ and the paraboloid $z = 1 - x^2 - y^2$.

Solution. If we put $z = 0$ in the equation of the paraboloid we get $x^2 + y^2 = 1$. This means the intersection between the paraboloid and the plane is the unit circle. So that the solid is bounded above the unit disk $R = \{x^2 + y^2 \le 1\}$ and below the paraboloid $z = 1 - x^2 - y^2$. Therefore, the volume is

$$\iint_R (1 - x^2 - y^2)dA.$$

Now using the polar coordinate

$$R = \{(r, \theta)|0 \le r \le 1, 0 \le \theta \le 2\pi\}.$$

So the integral becomes

$$\begin{aligned}&\int_0^{2\pi} \int_0^1 (1 - r^2\cos^2\theta - r^2\sin^2\theta)rdrd\theta \\ &= \int_0^{2\pi} \int_0^1 (1 - r^2)rdrd\theta \\ &= \int_0^{2\pi} \frac{1}{4}d\theta = \frac{\pi}{2}.\end{aligned}$$

Similar to the case of iterated integral with respect to x, y, the iterated integral with respect to r, θ can be computed in different order, if necessary. And as an extension to a more general type of region, if

$$D = \{(r, \theta)|\alpha \le \theta \le \beta, h_1(\theta) \le r \le h_2(\theta)\},$$

then

$$\iint_D f(x, y)dA = \int_\alpha^\beta \int_{h_1(\theta)}^{h_2(\theta)} f(r\cos\theta, \rho\sin\theta)rdrd\theta.$$

Example 5.20. In the special case where $f = 1$, $h_1 = 0$ and $h_2 = h(\theta)$, this gives us the formula of the area of a polar region

$$\text{area}(D) = \int_\alpha^\beta \int_0^{h(\theta)} r dr d\theta = \int_\alpha^\beta \frac{1}{2} h(\theta)^2 d\theta.$$

Example 5.21. Find the area enclosed by one loop of the four leaved rose curve $r = \cos 2\theta$.

Solution. We may write the region enclosed by one loop with

$$D = \{(r,\theta) | -\frac{\pi}{4} \le \theta \le \frac{\pi}{4}, 0 \le r \le \cos 2\theta\}.$$

Therefore, the area is

$$\begin{aligned} A = \iint_D dA &= \int_{-\frac{\pi}{4}}^{\frac{\pi}{4}} \int_0^{\cos 2\theta} r dr d\theta = \frac{1}{2} \int_{-\frac{\pi}{4}}^{\frac{\pi}{4}} \cos^2 2\theta d\theta \\ &= \frac{1}{4} \int_{-\frac{\pi}{4}}^{\frac{\pi}{4}} (1 + \cos 4\theta) d\theta = \frac{\pi}{8}. \end{aligned}$$

Example 5.22. Find the volume of the solid above the xOy-plane and below $z = x^2 + y^2$ inside the cylinder $x^2 + y^2 = 2x$.

Solution. The solid lies above the disk $D = \{(x-1)^2 + y^2 \le 1\}$ on the xOy-plane. In polar coordinates, this is

$$D = \{(r,\theta) | -\frac{\pi}{2} \le \theta \le \frac{\pi}{2}, 0 \le r \le 2\cos\theta\}.$$

Therefore, the volume

$$\begin{aligned} V = \iint_D (x^2 + y^2) dA &= \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \int_0^{2\cos\theta} r^2 \cdot r dr d\theta \\ &= \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} 4\cos^4\theta d\theta = 4 \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \left(\frac{1 + \cos 2\theta}{2}\right)^2 d\theta \\ &= 2 \int_0^{\frac{\pi}{2}} (1 + 2\cos 2\theta + \frac{1}{2}(1 + \cos 4\theta)) d\theta = \frac{3\pi}{2}. \end{aligned}$$

5.4. **Applications of Double Integrals.** In this section we shall see some application of the double integral. The main idea behind these applications is that if we have a double Riemann sum, then in the limit case, we obtain a double integral.

Density and Mass Let us consider the mass of a thin plate which is a region D. Suppose that the plate has a density function $\rho(x,y)$, that is,

$$\rho(x,y) = \lim \frac{\Delta m}{\Delta A},$$

where Δm and ΔA are the mass and area of a small region which contains the point (x,y). The limit is taken when the region approaches the point (x,y). In other words, if we consider a small rectangle contains the point (x,y), then its mass can be approximated by

$$m \approx \rho(x,y)\Delta A$$

Now to compute the total mass of the plate with a given density function $\rho(x,y)$, we first subdivide the plate into subrectangles R_{ij} with area ΔA. The mass of each R_{ij} can be computed using $\rho(x_{ij}^*, y_{ij}^*)\Delta A$, where (x_{ij}^*, y_{ij}^*) is a point in the subrectangle. Then the total mass can be approximated by the Riemann sum

$$m \approx \sum_{i=1}^{m} \sum_{j=1}^{n} \rho(x_{ij}^*, y_{ij}^*)\Delta A.$$

In the limit case

$$m = \iint_D \rho(x, y)dA$$

Example 5.23. Suppose that the mass density of a triangular plate with vertices $(1, 0), (1, 1)$ and $(0, 1)$ is given by $\rho = xy$. Find the total mass of the plate.

Solution. Let D be the region of the triangular plate. Then we may write D as a Type I region.

$$D = \{(x, y)|0 \leq x \leq 1, 1 - x \leq y \leq 1\}.$$

Therefore, the total mass is

$$m = \iint_D \rho(x, y)dA = \int_0^1 \int_{1-x}^1 xydydx$$

$$= \int_0^1 \frac{1}{2}(2x^2 - x^3)dx = \frac{5}{24}.$$

In physics, one may consider other type of density which is not necessarily positive. For example, consider the electric charge distributed over a region D. If the surface charge density is $\sigma(x, y)$, then the total charge is

$$Q = \iint_D \sigma(x, y)dA$$

Moments and center of mass Recall that the moment of a particle about an axis is the product of its mass and its distance from the axis. Given a thin plate or a lamina, let us consider the moment of the plate with respect to the coordinate axis. Again, we first subdivide the plate into small rectangles and then the mass of each subrectangle R_{ij} is given by $\rho(x_{ij}^*, y_{ij}^*)\Delta A$. The moment of R_{ij} with respect to the x-axis can be approximated by $\rho(x_{ij}^*, y_{ij}^*)\Delta A \cdot y_{ij}^*$. Now if we add the moment of each R_{ij} together and take the limit of the Riemann sum, we obtain that the moment of the plate with respect to the x-axis is

$$M_x = \lim_{m,n\to\infty} \sum_{i=1}^{m} \sum_{j=1}^{n} \rho(x_{ij}^*, y_{ij}^*)y_{ij}^*\Delta A = \iint_D y\rho(x, y)dA.$$

Similarly, the moment about the y-axis is

$$M_y = \lim_{m,n\to\infty} \sum_{i=1}^{m} \sum_{j=1}^{n} \rho(x_{ij}^*, y_{ij}^*)x_{ij}^*\Delta A = \iint_D x\rho(x, y)dA$$

We define the center of the mass $(\bar{x}, \bar{y})$ to be the point such that $m = \bar{x}M_y = \bar{y}M_x$. In other words,

$$\bar{x} = \frac{M_y}{m} = \frac{\iint_D x\rho(x, y)dA}{\iint_D \rho(x, y)dA}$$

$$\bar{y} = \frac{M_x}{m} = \frac{\iint_D y\rho(x,y)dA}{\iint_D \rho(x,y)dA}$$

Example 5.24. Suppose that the density of a semicircular plate which is places as the upper half disk $R = \{x^2 + y^2 \leq a^2, y \geq 0\}$ is $\rho = K\sqrt{x^2 + y^2}$, where $K > 0$ is some positive constant. Find the center of mass.

Solution. The region described using the polar coordinate is

$$\{(r,\theta)|0 \leq r \leq a, 0 \leq \theta \leq \pi\}.$$

The mass of the plate is given by

$$m = \iint_R K\sqrt{x^2 + y^2}dA = \int_0^{\pi}\int_0^a Kr \cdot rdrd\theta$$

$$= K\int_0^{\pi}\frac{a^3}{3}d\theta = \frac{K\pi a^3}{3}.$$

Let us now compute the moment. The moment with respect to the y-axis M_y in fact is 0 since both the region and the density function is symmetric about the y-axis.

The moment about the x-axis is

$$M_x = \iint_D y\rho(x,y)dA = \int_0^{\pi}\int_0^a r\sin\theta \cdot Kr \cdot rdrd\theta$$

$$= K\int_0^{\pi}\sin\theta \cdot \frac{r^4}{4}|_{r=0}^{r=a}d\theta = \frac{Ka^4}{2}$$

Thus,

$$\bar{y} = \frac{M_x}{m} = \frac{Ka^4}{2}\frac{3}{K\pi a^3} = \frac{3a}{2\pi}.$$

Therefore, the center of mass is $(0, \frac{3a}{2\pi})$.

Moment of Inertia Recall that the moment of inertia or second moment of a particle of mass m about an axis is md^2, where d is the distance between the particle and the axis. We may also generalize this to the case of a plate or lamina. Suppose that a lamina occupies the region D has a density function $\rho(x,y)$. Then the moment of inertia about the x-axis, after passing to the limit of a subdivision, is

$$I_x = \lim_{m,n\to\infty}\sum_{i=1}^{m}\sum_{j=1}^{n}\rho(x_{ij}^*, y_{ij}^*)y_{ij}^{*2}\Delta A = \iint_D y^2\rho(x,y)dA.$$

Similarly, the moment of inertia about the y-axis is

$$I_y = \lim_{m,n\to\infty}\sum_{i=1}^{m}\sum_{j=1}^{n}\rho(x_{ij}^*, y_{ij}^*)x_{ij}^{*2}\Delta A = \iint_D x^2\rho(x,y)dA$$

It is also interest to consider the moment of inertia about the origin, which is also called the polar moment of inertia, which is

$$I_0 = \lim_{m,n\to\infty}\sum_{i=1}^{m}\sum_{j=1}^{n}\rho(x_{ij}^*, y_{ij}^*)(x_{ij}^{*2} + y_{ij}^{*2})\Delta A = \iint_D (x^2 + y^2)\rho(x,y)dA = I_x + I_y.$$

Example 5.25. Find out the I_0 of a disk D centered at the origin and radius a with constant density ρ.

Solution. In polar coordinates, the disk is

$$D = \{(r, \theta) | 0 \le r \le a, 0 \le \theta \le 2\pi\}$$

So

$$I_0 = \iint_D (x^2 + y^2)\rho dA = \rho \int_0^{2\pi} \int_0^a r^2 r dr d\theta$$
$$= \rho \cdot 2\pi \cdot \frac{a^4}{4} = \frac{\rho\pi a^4}{2}.$$

Note that in this case, by symmetry, $I_x = I_y$. And we may found $I_x = I_y = \frac{I_0}{2} = \frac{\rho\pi a^4}{4}$. And in fact, the mass of the disk is $m = \rho\pi a^2$, so

$$I_0 = \frac{1}{2} m a^2$$

If we increase the mass, then we increase the moment of inertia.

Probability The probability density function of a random variable X is a function $f(x) \ge 0$ such that $\int_{\mathbb{R}} f(x)dx = 1$. In this case the probability of X between a, b is $\int_a^b f(x)dx$. This can be generalized to the case of two random variables X, Y. The joint density function $f(x, y)$ is a function such that the probability of (X, Y) in a region D is

$$P((X, Y) \in D) = \iint_D f(x, y)dA,$$

and

$$f(x, y) \ge 0, \iint_{\mathbb{R}^2} f(x, y)dA = 1$$

Example 5.26. Suppose that the joint density function of X, Y is given by

$$f(x, y) = \begin{cases} C \cdot (x + 2y), & 0 \le x, y \le 10 \\ 0, & \text{otherwise} \end{cases}$$

Find the value of C and $P(X \le 7, Y \ge 2)$.

Solution. The value of C and be found by integrating $f(x, y)$ over the entire plane. Because $f = 0$ outside the rectangle $[0, 10] \times [0, 10]$,

$$\iint_{\mathbb{R}^2} f dA = \int_0^{10} \int_0^{10} C(x + 2y) dy dx = 1500C.$$

Therefore, $1500C = 1 \Rightarrow C = \frac{1}{1500}$. Now

$$P(X \le 7, y \ge 2) = \int_0^7 \int_2^{10} \frac{1}{1500}(x + 2y) dy dx$$
$$= \frac{1}{1500} \int_0^7 (8x + 96) dx = \frac{868}{1500} \approx 0.579$$

If X, Y are two random variables with joint density function f, then the expected values of X and Y is defined to be

$$\mu_X = \iint_{\mathbb{R}^2} x f(x, y) dA, \mu_Y = \iint_{\mathbb{R}^2} y f(x, y) dA.$$

Example 5.27. In Example 5.26, the expected value for X and Y are given by

$$\mu_X = \int_0^{10}\int_0^{10} \frac{1}{1500}x(x+2y)dydx = \frac{50}{9} \approx 5.56.$$

$$\mu_Y = \int_0^{10}\int_0^{10} \frac{1}{1500}y(x+2y)dydx = \frac{55}{9} \approx 6.11.$$

5.5. **Triple Integrals.** Let $f = f(x,y,z)$ be a function of three variables. To define the triple integral of the function, let us first consider the case where f is defined on a rectangular box

$$B = \{(x,y,z)|a \le x \le b, c \le y \le d, r \le z \le s\} = [a,b]\times[c,d]\times[r,s].$$

Intuitively, the function f can be viewed as a distribution or density on the box B and the triple integral is the total distribution or the quantity on the box. To define the triple integral, we first take a partition of the box B and define the triple Riemann sum.

We divide the intervals $[a,b],[c,d],[r,s]$ into m,n,l subintervals of length $\Delta x, \Delta y, \Delta z$ respectively. A sub-box B_{ijk} is $[x_{i-1},x_i]\times[y_{j-1},y_j]\times[z_{k-1},z_k]$. The volume of each sub-box is $\Delta V = \Delta x\Delta y\Delta z$. In each box, we pick a sample point $(x^*_{ijk}, y^*_{ijk}, z^*_{ijk})$. Now we can form the triple Riemann sum.

$$\sum_{i=1}^{m}\sum_{j=1}^{n}\sum_{k=1}^{l} f(x^*_{ijk}, y^*_{ijk}, z^*_{ijk})\Delta V.$$

And the triple integral is defined to be the limit of the Riemann sum when $n,m,l\to\infty$.

Definition 5.28. The triple integral of f over B is

$$\iiint_B f(x,y,z)dV = \lim_{l,m,n\to\infty}\sum_{i=1}^{m}\sum_{j=1}^{n}\sum_{k=1}^{l} f(x^*_{ijk}, y^*_{ijk}, z^*_{ijk})\Delta V.$$

Just as double integrals, the way we compute the triple integral is usually through iterated integrals. In this case,

Theorem 5.29 (Fubini)**.** *If f is continuous on the box B, or the point of discontinuity is contained in finitely many surfaces, then*

$$\iiint_B f(x,y,z)dV = \int_a^b\int_c^d\int_r^s f(x,y,z)dzdydx.$$

The integral can also be computed in 5 different orders of dx, dy, dz.

Example 5.30. Evaluate $\iiint_B xyz^2dV$, where $B = [0,1]\times[-1,2]\times[0,3]$.

Solution. We can choose any of the six orders to compute this integral. If we choose "$dxdydz$", then

$$\begin{aligned}\iiint_B xyz^2dV &= \int_0^3\int_{-1}^2\int_0^1 xyz^2dxdydz\\ &= \int_0^3\int_{-1}^2 yz^2\cdot\frac{1}{2}dydz\\ &= \int_0^3 \frac{3z^2}{4}dz = \frac{27}{4}.\end{aligned}$$

Now we may define the triple integral over general regions. Similar to the case of double integral, if E is a bounded closed region in $\mathbb{R}^3$, then we may always find a box B contains the region E. In this case, if f is a function defined on E, we extend it to a function F on B so that $F = 0$ on $B - E$. And we define

$$\iiint_E f dV = \iiint_B F dV.$$

To compute the integral over general regions using the iterated integral, similar to the case of double integral, let us consider the following types of the region.

We say that a bounded closed solid region E is of Type 1, if it lies in between the graph of two surfaces of x, y. In other words,

$$E = \{(x, y, z)|(x, y) \in D, u_1(x, y) \leq z \leq u_2(x, y)\},$$

where D is a region on the xOy-plane, which is the projection of the solid onto the xOy-plane. In this case, the triple integral can be computed as

$$\iiint_E f(x, y, z)dV = \iint_D \int_{u_1(x,y)}^{u_2(x,y)} f(x, y, z)dzdA$$

Note that in the integral $\int_{u_1(x,y)}^{u_1(x,y)} f(x, y, z)dz$, x, y should be viewed as constants and after we integrated the function with respect to z, we put the upper and lower bound u_2 and u_1 in the expression so the resulting function is a function of x, y. And then we perform the double integral with respect to x, y.

If the region D is a Type I region, then

$$E = \{(x, y, z)|a \leq x \leq b, g_1(x) \leq y \leq g_2(x), u_1(x, y) \leq z \leq u_2(x, y)\},$$

and

$$\iiint_E f(x, y, z)dV = \int_a^b \int_{g_1(x)}^{g_2(x)} \int_{u_1(x,y)}^{u_2(x,y)} f(x, y, z)dzdydx$$

And if the region D is a Type II region, then

$$\iiint_E f(x, y, z)dV = \int_c^d \int_{h_1(y)}^{h_2(y)} \int_{u_1(x,y)}^{u_2(x,y)} f(x, y, z)dzdxdy$$

Similarly, we say the region E is of Type 2 if

$$E = \{(x, y, z)|(y, z) \in D, u_1(y, z) \leq x \leq u_2(y, z)\}.$$

And E is called a Type 3 region, if

$$E = \{(x, y, z)|(x, z) \in D, u_1(x, z) \leq y \leq u_2(x, z)\}.$$

In each case one may compute the integral using the iterated integral.

Example 5.31. Evaluate $\iiint_E z dV$, where the region E is the solid tetrahedron bounded by $x = 0, y = 0, z = 0$ and $x + y + z = 1$.

Solution. In fact, the region E can be viewed as both Type 1,2 or 3 region. As a Type 1 region, it is bounded between $z = 0$ and $z = 1 - x - y$. And the x, y are in the projection of the region onto the coordinate plane, which is a triangle between $x = 0, y = 0$ and $y = 1 - x$. Therefore,

$$E = \{(x, y, z)|0 \leq x \leq 1, 0 \leq y \leq 1 - x, 0 \leq z \leq 1 - x - y\}.$$

Now the triple integral can be computed as follows

$$\iiint_E z dV = \int_0^1 \int_0^{1-x} \int_0^{1-x-y} z dz dy dx.$$

$$= \int_0^1 \int_0^{1-x} \frac{1}{2}(1-x-y)^2 dy dx$$

$$= \int_0^1 \frac{1}{6}(1-x)^3 dx = -\frac{1}{6} \cdot \frac{(1-x)^4}{4}\bigg|_0^1 = \frac{1}{24}.$$

Sometimes, the double integral after we finished the first iterated integral can be computed using the polar coordinates.

Example 5.32. Evaluate $\iiint_E \sqrt{x^2+z^2} dV$, where E is the solid region bounded by the parabola $y = x^2 + z^2$ and $y = 4$.

Solution. If we write the region E as Type 1 region, then E is bounded between $z = \pm\sqrt{y-x^2}$. And the projection of the region onto the xOy-plane is a region in between $y = 4$ and the parabola $y = x^2$, so

$$\iiint_E \sqrt{x^2+z^2} dV = \int_{-2}^2 \int_{x^2}^4 \int_{-\sqrt{y-x^2}}^{\sqrt{y-x^2}} \sqrt{x^2+z^2} dz dy dx.$$

Although this integral is correct, it is very difficult to compute. We should try to write E as other types of the region. The observation is that both the integrand and the region has some expressions of x^2+z^2. Therefore, we can write E as a Type 3 region. In this case, $x^2+z^2 \le y \le 4$. And the projection of E onto the xOz-plane is the disk $D = \{x^2+z^2 \le 4\}$.

Now the integral can be computed as

$$\iiint_E \sqrt{x^2+z^2} dV = \iint_D \int_{x^2+z^2}^4 \sqrt{x^2+z^2} dy dA = \iint_D (4-x^2-z^2)\sqrt{x^2+z^2} dA.$$

To compute the double integral, we may use the polar coordinates. In this case, let $x = r\cos\theta, z = r\sin\theta$, we have $dA = r dr d\theta$. And

$$\iint_D (4-x^2-z^2)\sqrt{x^2+z^2} dA = \int_0^{2\pi} \int_0^2 (4-r^2) r \cdot r dr d\theta$$

$$= 2\pi \cdot \left(\frac{4r^3}{3} - \frac{r^5}{5}\right)_0^2 = \frac{128\pi}{15}.$$

In the next example, let us practise writing the triple integral in different orders.

Example 5.33. Rewrite the integral $\int_0^1 \int_0^{x^2} \int_0^y f(x,y,z) dz dy dx$ in the order of $dx dz dy$.

Solution. If E is the region of the integral, then

$$E = \{0 \le x \le 1, 0 \le y \le x^2, 0 \le z \le y\}$$

In other words, the region E is between $z = 0$ and $z = y$, and the projection of the region onto the xOy-plane is $D_1 = \{0 \le x \le 1, 0 \le y \le x^2\}$.

Now, to rewrite the region as a Type 2 region, the x-coordinate of point in E is between $x = 1$ and $y = x^2$ or $x = \sqrt{y}$. The projection of E onto the yOz-plane is a triangle D_2 bounded by $z = 0$, $y = 1$ and $z = y$. Therefore, $D_2 = \{0 \leq y \leq 1, 0 \leq z \leq y\}$. And

$$\int_0^1 \int_0^{x^2} \int_0^y f(x,y,z)dzdydx = \int_0^1 \int_0^y \int_{\sqrt{y}}^1 f(x,y,z)dxdzdy.$$

Some of the application we studied in the previous section may be generalized to the case of triple integral.

Volume vs Area Recall that in the case of double integral, the value of integral of the constant function $f = 1$ equals to the area of the region. That is,

$$\text{area}(D) = \iint_D dA.$$

If E is a bounded solid region, then

$$\text{vol}(E) = \iiint_E dV.$$

Example 5.34. Find out the volume of the tetrahedron T bounded by the planes $x = 0, z = 0, x + 2y + z = 2$ and $x = 2y$.

Solution. The tetrahedron can be viewed as a Type 1 region between $z = 0$ and $z = 2-x-2y$. The projection of T onto the xOy-plane is a triangle bounded by $x = 2y, x = 2-2y$ and $x = 0$.

$$T = \{(x,y,z)|0 \leq x \leq 1, x/2 \leq y \leq 1 - x/2, 0 \leq z \leq 2 - x - 2y\}.$$

The volume of T is

$$\text{vol}(T) = \iiint_T dV = \int_0^1 \int_{x/2}^{1-x/2} \int_0^{2-x-2y} dzdydx$$

$$= \int_0^1 \int_{x/2}^{1-x/2} (2 - x - 2y)dydx = \int_0^1 (x^2 - 2x + 1)dx = \frac{1}{3}.$$

Mass We may also generalize the computation of the mass to the case of triple integral. Suppose that the density function of a solid E is $\rho(x,y,z)$. Then the mass is

$$m = \iiint_E \rho(x,y,z)dV.$$

And the moments about the coordinate planes are

$$M_{yz} = \iiint_E x\rho(x,y,z)dV, M_{xz} = \iiint_E y\rho(x,y,z)dV, M_{xy} = \iiint_E z\rho(x,y,z)dV.$$

Then the center of mass is

$$(\bar{x}, \bar{y}, \bar{z}) = \left(\frac{M_{yz}}{m}, \frac{M_{xz}}{m}, \frac{M_{xy}}{m}\right).$$

Example 5.35. Find out the center of mass of a solid of constant density k that is bounded by the parabolic cylinder $x = y^2$ and the planes $x = z, z = 0, x = 1$.

Solution. We may view the solid E as a Type 1 region where $0 \le z \le x$. The projection onto the xOy-plane is a Type II region between the curves $x = 1$ and $x = y^2$.

$$E = \{(x, y, z)| -1 \le y \le 1, y^2 \le x \le 1, 0 \le z \le x\}.$$

Since the density $\rho = k$ is a constant, the mass is

$$m = \iiint_E kdV = \int_{-1}^{1}\int_{y^2}^{1}\int_{0}^{x} kdzdxdy$$

$$= k\int_{-1}^{1}\int_{y^2}^{1} xdxdy = \frac{k}{2}\int_{-1}^{1}(1-y^4)dy = \frac{4k}{5}.$$

By symmetry, $M_{xz} = 0$.

$$M_{yz} = \iiint_E x \cdot kdV = \int_{-1}^{1}\int_{y^2}^{1}\int_{0}^{x} kxdzdxdy$$

$$= k\int_{-1}^{1}\int_{y^2}^{1} x^2dxdy = \frac{2k}{3}\int_{-1}^{1}(1-y^6)dy = \frac{4k}{7}.$$

$$M_{xy} = \iiint_E z \cdot kdV = \int_{-1}^{1}\int_{y^2}^{1}\int_{0}^{x} kzdzdxdy$$

$$= \frac{k}{2}\int_{-1}^{1}\int_{y^2}^{1} x^2dxdy = \frac{k}{3}\int_{-1}^{1}(1-y^6)dy = \frac{2k}{7}.$$

Therefore the center of mass is

$$(\bar{x}, \bar{y}, \bar{z}) = \left(\frac{M_{yz}}{m}, \frac{M_{xz}}{m}, \frac{M_{xy}}{m}\right) = \left(\frac{5}{7}, 0, \frac{5}{14}\right).$$

Moments of inertia The moments of inertia with respect to the three coordinate axes are

$$I_x = \iiint_E (y^2+z^2)\rho(x, y, z,)dV$$

$$I_y = \iiint_E (x^2+z^2)\rho(x, y, z,)dV$$

$$I_z = \iiint_E (x^2+y^2)\rho(x, y, z,)dV$$

We may also consider the moment of inertia about the origin, in this case, it is

$$I_0 = \iint_E (x^2+y^2+z^2)\rho(x, y, z)dV.$$

Probability Finally, if we have three random variables X, Y, Z. We consider their joint density function $f(x, y, z)$, that is, a function such that

$$P((X, Y, Z) \in E) = \iiint_E f(x, y, z)dV, f(x, y, z) \ge 0, \iiint_{\mathbb{R}^3} f(x, y, z)dV = 1$$

5.6. **Triple integrals in cylindrical and spherical coordinate.** Recall that the cylindrical coordinates of a point in $\mathbb{R}^3$ is (r, θ, z) such that

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

where (x, y, z) is the Cartesian coordinate of the point. The coordinates (r, θ) are the same as the polar coordinates. Therefore, if E is a Type 1 region such that the projection of E onto the xOy-plane is a general polar region D. That is,

$$E = \{(x, y, z)|(x, y) \in D, u_1(x, y) \leq z \leq u_2(x, y)\}.$$

$$D = \{(r, \theta)|\alpha \leq \theta \leq \beta, h_1(\theta) \leq r \leq h_2(\theta)\}$$

Then the triple integral

$$\iiint_E f(x, y, z)dV = \iint_D \int_{u_1(x,y)}^{u_2(x,y)} f(x, y, z)dzdA$$

$$= \int_\alpha^\beta \int_{h_1(\theta)}^{h_2(\theta)} \int_{u_1(r\cos\theta, r\sin\theta)}^{u_2(r\cos\theta, r\sin\theta)} f(x, y, z)rdzdrd\theta$$

The above formula is the triple integral in cylindrical coordinates. If we covert a triple integral from Cartesian coordinate to cylindrical coordinate using $x = r\cos\theta, y = r\sin\theta, z = z$, then $dV = rdzdrd\theta$.

Example 5.36. A solid E lies within the cylinder $x^2 + y^2 = 1$ between the surfaces $z = 4$ and $z = 1 - x^2 - y^2$. Evaluate the triple integral $\iiint_E \sqrt{x^2 + y^2}dV$

Solution. The surfaces in cylindrical coordinates are $z = 4$ and $z = 1 - r^2$. The projection of E onto the xOy-plane is the disk $x^2 + y^2 \leq 1$, which is $r \leq 1$ so we may write

$$E = \{(r, \theta, z)|0 \leq \theta \leq 2\pi, 0 \leq r \leq 1, 1 - r^2 \leq z \leq 4\}.$$

And

$$\iiint_E \sqrt{x^2 + y^2}dV = \int_0^{2\pi} \int_0^1 \int_{1-r^2}^4 r^2 dzdrd\theta$$

$$= \int_0^{2\pi} \int_0^1 (3r^2 + r^4)drd\theta = \frac{12\pi}{5}.$$

Example 5.37. Evaluate $\int_{-2}^2 \int_{-\sqrt{4-x^2}}^{\sqrt{4-x^2}} \int_{\sqrt{x^2+y^2}}^2 (x^2 + y^2)dzdydx$

Solution. The region of the integral is

$$E = \{(x, y, z)| -2 \leq x \leq 2, -\sqrt{4 - x^2} \leq y \leq \sqrt{4 - x^2}, \sqrt{x^2 + y^2} \leq z \leq 2\}.$$

The projection of E onto the xOy-plane is the disk $\{-2 \leq x \leq 2, -\sqrt{4 - x^2} \leq y \leq \sqrt{4 - x^2}\}$. Thus, in cylindrical coordinates,

$$E = \{(r, \theta, z)|0 \leq \theta \leq 2\pi, 0 \leq r \leq 2, r \leq z \leq 2\}.$$

And

$$\iiint_E (x^2 + y^2)dzdydx = \int_0^{2\pi} \int_0^2 \int_r^2 r^2 \cdot rdzdrd\theta$$

$$= \int_0^{2\pi} \int_0^2 r^3(2 - r)drd\theta = \frac{16\pi}{5}.$$

The spherical coordinates of a point (ρ,θ,ϕ) is such that

$$x=\rho\sin\phi\cos\theta, y=\rho\sin\phi\sin\theta, z=\rho\cos\phi$$

Let us consider a triple integral in the spherical coordinate system. We start from a "spherical rectangular region" or a spherical wedge

$$E=\{(\rho,\theta,\phi)|a\le\rho\le b,\alpha\le\theta\le\beta,c\le\phi\le d\}$$

If we subdivide the region E into small spherical wedges E_{ijk} by dividing the intervals into subintervals of length $\Delta\rho,\Delta\theta,\Delta\phi$, then each E_{ijk} can be approximated by a rectangular box with dimensions $\Delta\rho$, $\rho_i\Delta\phi$, and $\rho_i\sin\phi_k\Delta\theta$. We approximate the volume of E_{ijk} with

$$\Delta V=\Delta\rho\cdot\phi_i\Delta\phi\cdot\rho_i\sin\phi_k\Delta\theta=\rho_i^2\sin\phi_k\Delta\rho\Delta\theta\Delta\phi.$$

Now if we take the limit of the Riemann sum which defines the triple integral, and let

$$F(\rho,\theta,\phi)=f(\rho\sin\phi\cos\theta,\rho\sin\phi\sin\theta,\rho\cos\phi).$$

Then

$$\iiint_E f(x,y,z)dV=\int_c^d\int_\alpha^\beta\int_a^b F(\rho,\theta,\phi)\rho^2\sin\phi d\rho d\theta d\phi.$$

This formula is the triple integral in the spherical coordinates. Note that in this case $dV=\rho^2\sin\phi d\rho d\theta d\phi$. One may also extend this formula to the regions such as

$$E=\{(\rho,\theta,\phi)|c\le\phi\le d,\alpha\le\theta\le\beta,g_1(\theta,\phi)\le\rho\le g_2(\theta,\phi)\}.$$

Example 5.38. Find the integral $\iiint_B e^{(x^2+y^2+z^2)^{3/2}}dV$, where B is the unit ball $\{x^2+y^2+z^2\le 1\}$.

Solution. We use spherical coordinates to compute this integral. In spherical coordinates, the region is

$$B=\{(\rho,\theta,\phi)|0\le\rho\le 1,0\le\theta\le 2\pi,0\le\phi\le\pi\}.$$

And $x^2+y^2+z^2=\rho^2$. So the integral is

$$\begin{aligned}\iiint_B e^{(x^2+y^2+z^2)^{3/2}}dV&=\int_0^\pi\int_0^{2\pi}\int_0^1 e^{\rho^3}\rho^2\sin\phi d\rho d\theta d\phi\\&=\int_0^\pi\sin\phi d\phi\cdot\int_0^{2\pi}d\theta\cdot\int_0^1 e^{\rho^3}\rho^2 d\rho\\&=(-\cos\phi|_0^\pi)\cdot 2\pi\cdot(\frac{1}{3}e^{\rho^3})|_0^1=\frac{4}{3}\pi(e-1).\end{aligned}$$

Example 5.39. Find the volume of the solid that lies above the cone $z=\sqrt{x^2+y^2}$ and below $x^2+y^2+z^2=z$.

Solution. In spherical coordinates, the cone is $\rho\cos\phi=\sqrt{\rho^2\sin^2\phi}=\rho\sin\phi$. So $\cos\phi=\sin\phi\Rightarrow\phi=\frac{\pi}{4}$. And the sphere is $\rho^2=\rho\cos\phi$ or $\rho=\cos\phi$. Therefore, the solid is

$$E=\{(\rho,\theta,\phi)|0\le\theta\le 2\pi,0\le\phi\le\pi/4,0\le\rho\le\cos\phi\}.$$

The volume of the solid is

$$\iiint_E dV=\int_0^{2\pi}\int_0^{\pi/4}\int_0^{\cos\phi}\rho^2\sin\phi d\rho d\phi d\theta$$

$$=2\pi\cdot\int_0^{\pi/4}\frac{1}{3}\sin\phi\cos^3\phi d\phi=\frac{2\pi}{3}(-\frac{1}{4}\cos^4\phi)|_0^{\pi/4}=\frac{\pi}{8}.$$

5.7. **Change of variables.** In this section we will study general change of variables. Recall that we have studied several different type of change of variables before. In the case of single variable, the substitution rule is

$$\int_a^b f(x)dx = \int_c^d f(x(u))\frac{dx}{du}du.$$

The general multi-variable case is slightly different. We have seen some examples,

$$\textbf{Polar coordinates: } \iint_R f(x,y)dA = \iint_S f(r\cos\theta, r\sin\theta)rdrd\theta$$

$$\textbf{Cylindrical coordinates: } \iiint_E f(x,y,z)dV = \iiint_B f(r\cos\theta, r\sin\theta, z)rdzdrd\theta$$

$$\textbf{Spherical coordinates: } \iiint_E f(x,y,z)dV = \iiint_B f(\rho\sin\phi\cos\theta, \rho\sin\phi\sin\theta, \rho\cos\phi)\rho^2\sin\phi d\rho d\phi d\theta$$

We shall see how the expressions r or $\rho^2\sin\phi$ are related to the partial derivatives of the change of variables. Let us begin with the definition of the change of variables in the case of two variables. A change of variable is a correspondence between two sets of the coordinates. This correspondence can be viewed as a transformation

$$(x,y) = T(u,v) = (x(u,v), y(u,v)).$$

Here, one can view the transformation T as a vector-valued multi-variable function. The relation between (x,y) and (u,v) is

$$\begin{cases} x &= x(u,v) \\ y &= y(u,v). \end{cases}$$

Example 5.40. For example, in the polar coordinates, $x = r\cos\theta$, $y = r\sin\theta$. This is a correspondence between (x,y) and (r,θ). We can write $(x,y) = T(r,\theta) = (r\cos\theta, r\sin\theta)$.

We say that a transformation T is a C^1 transformation, if the functions $x(u,v), y(u,v)$ has continuous first-order partial derivatives. The point (x,y) is called the image of (u,v) under the transformation T. The collection of all points $\{(x,y)|(x,y) = T(u,v)\}$ is called the image of T. T is a one-to-one transformation if no two points have the same image. If T is a one-to-one transformation, then it has an inverse transformation T^{-1}. Usually, we write the inverse transformation as $(u,v) = T^{-1}(x,y)$, or componentwise,

$$\begin{cases} u &= u(x,y) \\ v &= v(x,y). \end{cases}$$

Example 5.41. In the case of the polar coordinates, T is one-to-one when, for example, $\theta \in (0,\pi/2)$ and $r \in (0,\infty)$. In this case the inverse transform is $r = \sqrt{x^2+y^2}, \theta = \arctan\frac{y}{x}$.

Similar to the case of the projection, for a closed region, one may determine its image under a C^1 transformation by looking at the image of the boundary.

Example 5.42. Consider a transformation $T : x = u^2 - v^2, y = 2uv$. Find the image of $S = \{(u,v)|0 \le u \le 1, 0 \le v \le 1\}$.

Solution. Since S is a closed region, the image of S is determined by the image of its boundary. The boundary of S consists of four segments. Let us label them counterclockwise using S_1, S_2, S_3, S_4 and let S_1 be the segment on the u-axis. That is, $S_1 : v = 0, 0 \le u \le 1$.

In this case, the image of S_1 is given by $x = u^2, y = 0$. Since $0 \leq 1 \leq 1$, we obtain that $0 \leq x \leq 1$. Therefore, it is the segment on the x-axis with $0 \leq x \leq 1$.

Now let us consider $S_2 : u = 1, 0 \leq v \leq 1$. In this case, we get $x = 1 - v^2, y = 2v$. By eliminating v, we see that (x, y) is on the curve $x = 1 - \frac{y^2}{4}$, $0 \leq x \leq 1$.

Similarly, we may obtain that the image of S_3 is given by the graph of the curve $x = \frac{y^2}{4} - 1$, $-1 \leq x \leq 0$. And the image of S_4 is the segment $y = 0$, $-1 \leq x \leq 1$. Finally, the image of S is the region R enclosed by the image of S_1, S_2, S_3, S_4. If we describe it as a Type II region, then it is

$$R = \{(x, y)|0 \leq y \leq 2, \frac{y^2}{4} - 1 \leq x \leq 1 - \frac{y^2}{4}\}.$$

Now let us consider change of variables in double integrals. Our goal is to compute the area of the image of a small rectangle under the transformation. Let S be a rectangle in uv-coordinates with dimensions Δu and Δv. Assume that one of the vertices on the lower left corner of S is (u_0, v_0). Let $(x_0, y_0) = T(u_0, v_0)$.

Consider the vector $\mathbf{r}(u, v) = \langle x(u, v), y(u, v)\rangle$. Then the image of the lower side of the rectangle S is given by the vector valued function $\mathbf{r}(u, v_0)$. The tangent vector at (x_0, y_0) to the curve $\mathbf{r}(u, v_0)$ is $\mathbf{r}_u = \langle \frac{\partial x}{\partial u}, \frac{\partial y}{\partial u}\rangle$. Similarly, the image of the left side of S is $\mathbf{r}(u_0, v)$, and the tangent vector at the vertex is $\mathbf{r}_v = \langle \frac{\partial x}{\partial v}, \frac{\partial y}{\partial v}\rangle$

Now the image of S can be approximated by a parallelogram determined by the secant vectors

$$\mathbf{r}(u_0 + \Delta u, v_0) - \mathbf{r}(u_0, v_0), \mathbf{r}(u_0, v_0 + \Delta v) - \mathbf{r}(u_0, v_0)$$

and

$$\mathbf{r}_u \approx \frac{\mathbf{r}(u_0 + \Delta u, v_0) - \mathbf{r}(u_0, v_0)}{\Delta u}$$
$$\mathbf{r}_v \approx \frac{\mathbf{r}(u_0, v_0 + \Delta v) - \mathbf{r}(u_0, v_0)}{\Delta v}$$

which means we may approximate the sides of the image of the rectangle by the vectors $\mathbf{r}_u \Delta u$ and $\mathbf{r}_v \Delta v$. The area, in this case can be approximated by

$$|\mathbf{r}_u \Delta u \times \mathbf{r}_v \Delta v| = |\mathbf{r}_u \times \mathbf{r}_v| \Delta u \Delta v.$$

Note that the cross product

$$\mathbf{r}_u \times \mathbf{r}_v = \det \begin{pmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ \frac{\partial x}{\partial u} & \frac{\partial y}{\partial u} & 0 \\ \frac{\partial x}{\partial v} & \frac{\partial y}{\partial v} & 0 \end{pmatrix} = \det \begin{pmatrix} \frac{\partial x}{\partial u} & \frac{\partial y}{\partial u} \\ \frac{\partial x}{\partial v} & \frac{\partial y}{\partial v} \end{pmatrix} \mathbf{k}$$

Let us give a special name to this determinant.

Definition 5.43. We define the Jacobian of the transformation $T : x = x(u, v), y = y(u, v)$ to be

$$\frac{\partial(x, y)}{\partial(u, v)} = \det \begin{pmatrix} \frac{\partial x}{\partial u} & \frac{\partial y}{\partial u} \\ \frac{\partial x}{\partial v} & \frac{\partial y}{\partial v} \end{pmatrix} = \frac{\partial x}{\partial u}\frac{\partial y}{\partial v} - \frac{\partial x}{\partial v}\frac{\partial y}{\partial u}$$

Using the notion of the Jacobian, we may write the area as

$$\Delta A \approx \left|\frac{\partial(x, y)}{\partial(u, v)}\right| \Delta u \Delta v,$$

where $|,|$ is the absolute value, and the Jacobian is evaluated at (u_0, v_0).

Now let us consider the double integral of $f(u,v)$ over a region R.

$$\iint_R f(x,y)dA = \lim \sum_i \sum_j f(x_i, y_j)\Delta A$$

$$\approx \lim \sum_i \sum_j f(x(u_i,v_j), y(u_i,v_j)) \left| \frac{\partial(x,y)}{\partial(u,v)} \right| \Delta u \Delta v$$

$$= \iint_{T^{-1}(R)} f(x(u,v), y(u,v)) \left| \frac{\partial(x,y)}{\partial(u,v)} \right| dudv$$

and we obtained that

Theorem 5.44 (Change of Variables)**.** *Suppose that $T: x = x(u,v), y = y(u,v)$ is a C^1 transformation with non-zero Jacobian. Suppose that f is integrable on R. Suppose also that T is one-to-one, except perhaps on the boundary of R. Then*

$$\iint_R f(x,y)dA = \iint_{T^{-1}(R)} f(x(u,v), y(u,v)) \left| \frac{\partial(x,y)}{\partial(u,v)} \right| dudv$$

Example 5.45. In polar coordinates, $T: x = r\cos\theta, y = r\sin\theta$, so

$$\frac{\partial(x,y)}{\partial(r,\theta)} = \det \begin{pmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{pmatrix} = r\cos^2\theta + r\sin^2\theta = r > 0.$$

If the image of the region R is S, then the above formula becomes

$$\iint_R f(x,y)dxdy = \iint_S f(r\cos\theta, r\sin\theta) r dr d\theta.$$

Example 5.46. Evaluate the integral $\iint_R e^{(x+y)/(x-y)}dA$, where R is the trapezoidal region with vertices $(1,0), (2,0), (0,-2)$ and $(0,-1)$.

Solution. Let us consider a change of variables

$$u = x + y, v = x - y$$

Note that this corresponds to the inverse of the transformation T. The transformation can be computed by solving x, y from the equation.

$$x = \frac{1}{2}(u+v), y = \frac{1}{2}(u-v).$$

Now the Jacobian is

$$\frac{\partial(x,y)}{\partial(u,v)} = det \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} \end{pmatrix} = -\frac{1}{2}.$$

We then find the image of this transformation. The sides of the region R are given by

$$y = 0, x - y = 2, x = 0, x - y = 1.$$

The images of these linea are

$$u - v = 0, v = 2, u + v = 0, v = 1.$$

Thus, the image is a trapezoidal region

$$S = \{(u,v) | 1 \le v \le 2, -v \le u \le v\}.$$

Therefore,

$$\iint_R e^{(x+y)/(x-y)}dA = \iint_S e^{u/v} \cdot \left|-\frac{1}{2}\right| dudv$$

$$= \int_1^2 \int_{-v}^{v} \frac{1}{2} e^{u/v} dudv = \frac{1}{2}\int_1^2 (ve^{u/v})|_{-v}^{v} dv = \frac{1}{2}\int_1^2 (e-e^{-1})vdv = \frac{3}{4}(e-e^{-1}).$$

Let us now consider the triple integrals. The idea is similar to the case of double integrals. Let T be a transformation from uvw-coordinates to xyz-coordinates, that is,

$$x = x(u,v,w), y = y(u,v,w), z = z(u,v,w).$$

Then the volume of the image of the rectangular box can be approximated by the following

$$\Delta V = \Delta x \Delta y \Delta z = \left|\frac{\partial(x,y,z)}{\partial(u,v,w)}\right| \Delta u \Delta v \Delta w,$$

where $\frac{\partial(x,y,z)}{\partial(u,v,w)}$ is the Jacobian of the transformation T.

$$\frac{\partial(x,y,z)}{\partial(u,v,w)} = \det \begin{pmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} & \frac{\partial x}{\partial w} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} & \frac{\partial y}{\partial w} \\ \frac{\partial z}{\partial u} & \frac{\partial z}{\partial v} & \frac{\partial z}{\partial w} \end{pmatrix}.$$

Theorem 5.47. *Suppose that T is a one-to-one, except perhaps on the boundary of the region R and f is integrable. Then*

$$\iiint_R f(x,y,z)dV = \iiint_{T^{-1}(R)} f(x(u,v,w), y(u,v,w), z(u,v,w)) \left|\frac{\partial(x,y,z)}{\partial(u,v,w)}\right| dudvdw.$$

Example 5.48. Compute the Jacobian of the change of variables of the spherical coordinates.

Solution. The change of variables or the transformation is given by

$$x = \rho \sin\phi \cos\theta, y = \rho\sin\phi\sin\theta, z = \rho\cos\phi.$$

The Jacobian is

$$\frac{\partial(x,y,z)}{\partial(\rho,\theta,\phi)} = \det \begin{pmatrix} \sin\phi\cos\theta & -\rho\sin\phi\sin\theta & \rho\cos\phi\cos\theta \\ \sin\phi\sin\theta & \rho\sin\phi\cos\theta & \rho\cos\phi\sin\theta \\ \cos\phi & 0 & -\rho\sin\phi \end{pmatrix}$$

$$= \cos\phi(-\rho^2\sin\phi\cos\phi\sin^2\theta - \rho^2\sin\phi\cos\phi\cos^2\theta)$$
$$-\rho\sin\phi(\rho\sin^2\phi\cos^2\theta + \rho\sin^2\phi\sin^2\theta)$$
$$= -\rho^2\sin\phi\cos^2\phi - \rho^2\sin\phi\sin^2\phi = -\rho^2\sin\phi.$$

Thus,

$$\left|\frac{\partial(x,y,z)}{\partial(\rho,\theta,\phi)}\right| = \rho^2\sin\phi.$$

As a result, we obtain that

$$\iiint_R f(x,y,z)dV = \iiint_{T^{-1}(R)} f(\rho\sin\phi\cos\theta, \rho\sin\phi\sin\theta, \rho\cos\phi)\rho^2\sin\phi d\rho d\theta d\phi.$$

Additional Examples

Example 5.49. While evaluating a double integral over a rectangle $R = [a, b] \times [c, d]$, if $f(x, y) = g(x)h(y)$, then

$$\iint_R f(x,y)dA = \int_a^b \int_c^d g(x)h(y)dydx = \int_a^b g(x)dx \cdot \int_c^d h(y)dy.$$

Example 5.50. Evaluate the double integral $\iint_R x\sec^2 y dA$, where $R = [0, 2] \times [0, \pi/4]$.

Solution. Note that $x \sec^2 y$ is separable $x \cdot \sec^2 y$.

$$\iint_R x\sec^2 y dA = \int_0^2 x dx \cdot \int_0^{\pi/4} sec^2 y dy$$
$$= 2 \cdot \tan(\pi/4) = 2.$$

Example 5.51. In probability, we sometimes would like to compute the improper integral $\iint_{\mathbb{R}^2} f(x, y)dA$. We define the improper integral

$$I = \iint_{\mathbb{R}^2} f(x,y)dA$$
$$= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x,y)dA$$
$$= \lim_{a\to\infty} \iint_{D_a} f(x,y)dA,$$

where D_a is a disk of radius a centered at the origin.

In the next example, we compute $\int_{-\infty}^{\infty} e^{-x^2}dx$ using the improper double integral.

Example 5.52. Find $\int_{-\infty}^{\infty} e^{-x^2}dx$ by first computing $\iint_{\mathbb{R}^2} e^{-(x^2+y^2)}dA$.

Solution. Let us first compute $\iint_{D_a} e^{-(x^2+y^2)}dA$. We compute it in polar coordinates.

$$D_a = \{0 \le r \le a, 0 \le \theta \le 2\pi\}$$

so that

$$\iint_{D_a} e^{-(x^2+y^2)}dA = \int_0^{2\pi} \int_0^a e^{-r^2} r dr d\theta$$
$$= 2\pi \cdot -\frac{e^{-r^2}}{2}|_0^a = \pi \cdot (1 - e^{-a^2}).$$

Now

$$\iint_{\mathbb{R}^2} e^{-(x^2+y^2)}dA = \lim_{a\to\infty} \iint_{D_a} e^{-(x^2+y^2)}dA$$
$$= \lim_{a\to\infty} \pi \cdot (1 - e^{-a^2}) = \pi.$$

Another way to evaluate this improper integral is to compute it as a limit when $a \to \infty$ over a square S_a with vertices $(\pm a, \pm a)$. In this case,

$$\iint_{\mathbb{R}^2} e^{-(x^2+y^2)}dA = \lim_{a\to\infty} \iint_{S_a} e^{-(x^2+y^2)}dA$$
$$= \lim_{a\to\infty} \int_{-a}^{a} \int_{-a}^{a} e^{-(x^2+y^2)}dxdy = \lim_{a\to\infty} \int_{-a}^{a} \int_{-a}^{a} e^{-x^2} \cdot e^{-y^2}dxdy$$

$$= \lim_{a\to\infty} \int_{-a}^{a} e^{-x^2} dx \cdot \int_{-a}^{a} e^{-y^2} dy = \int_{-\infty}^{\infty} e^{-x^2} dx \cdot \int_{-\infty}^{\infty} e^{-y^2} dy$$

$$= \left(\int_{-\infty}^{\infty} e^{-x^2} dx \right)^2 = \pi$$

$$\Rightarrow \int_{-\infty}^{\infty} e^{-x^2} dx = \sqrt{\pi}.$$

By doing a change of variable, we obtain the famous Gauss distribution

$$\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx = 1.$$

6. Vector Calculus

6.1. **Vector Fields.** A vector field is a function whose domain is a subset of $\mathbb{R}^2$ or $\mathbb{R}^3$ and whose range is a set of vectors. At each point in the domain, there is a vector associated to the point.

Definition 6.1. Let D be a set in $\mathbb{R}^2$ (or $\mathbb{R}^3$). A vector field on D is a function that assigns to each point (x, y) (or (x, y, z)) in D a vector $\mathbf{F}(x, y)$ (or $\mathbf{F}(x, y, z)$).

One way to represent a vector field is to draw arrow representing the vector $\mathbf{F}(x, y)$ with the initial point at (x, y). Since $\mathbf{F}(x, y)$ is a vector, we may write down its component form.

$$\mathbf{F}(x, y) = \langle P(x, y), Q(x, y) \rangle,$$

where the functions P, Q are called the component function of $\mathbf{F}$. Sometimes, we also write

$$\mathbf{F}(x, y) = P(x, y)\mathbf{i} + Q(x, y)\mathbf{j} \text{ or } \mathbf{F} = P\mathbf{i} + Q\mathbf{j}.$$

In the case of $\mathbb{R}^3$, a vector field can be represented by

$$\mathbf{F}(x, y, z) = \langle P(x, y, z), Q(x, y, z), R(x, y, z) \rangle = P\mathbf{i} + Q\mathbf{j} + R\mathbf{k}.$$

Example 6.2. Consider the vector field $\mathbf{F} = -y\mathbf{i} + x\mathbf{j}$. Some of the vectors of $\mathbf{F}(x, y)$ is showed in the following table.

(x, y)	$\mathbf{F}(x, y)$
$(1, 0)$	$\langle 0, 1 \rangle$
$(0, 1)$	$\langle -1, 0 \rangle$
$(-1, 0)$	$\langle 0, -1 \rangle$
$(0, -1)$	$\langle 1, 0 \rangle$

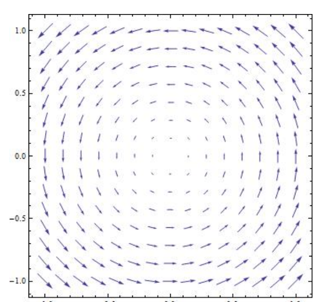


Figure 11. $\mathbf{F} = -y\mathbf{i} + x\mathbf{j}$

Example 6.3. Here are some sketch of vector fields in $\mathbb{R}^2$ and $\mathbb{R}^3$.

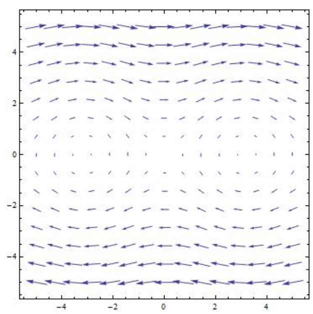


FIGURE 12. $\mathbf{F} = y\mathbf{i} + \sin x\mathbf{j}$

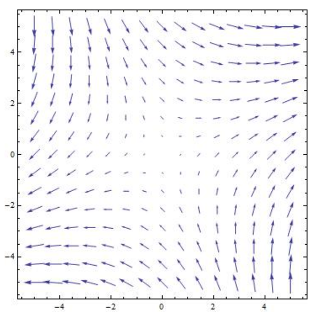


FIGURE 13. $\mathbf{F} = (x+y)\mathbf{i} + (x-y)\mathbf{j}$

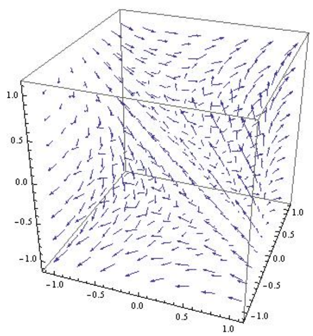


FIGURE 14. $\mathbf{F} = y\mathbf{i} + z\mathbf{j} + x\mathbf{k}$

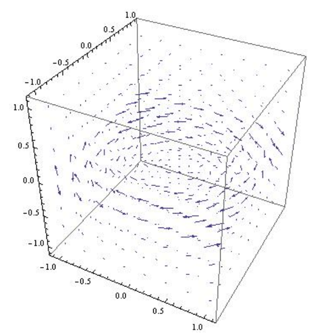


FIGURE 15. $\mathbf{F} = \frac{y}{z}\mathbf{i} - \frac{x}{z}\mathbf{j} + \frac{z}{4}\mathbf{k}$

Example 6.4. Newton's law of Gravitation states that the magnitude of the gravitational force between two objects with mass m and M is given by

$$|\mathbf{F}| = \frac{mMG}{r^2}.$$

Suppose that the object with mass M is placed at the origin, and the object with mass m is located at (x, y, z). Then the gravitation force on the second object is in the direction of $-\langle x, y, z\rangle$. If we write $\mathbf{x} = \langle x, y, z\rangle$, then

$$\begin{aligned} \mathbf{F} &= -\frac{mMG}{|\mathbf{x}|^3}\mathbf{x} \\ &= \frac{-mMGx}{(x^2+y^2+z^2)^{3/2}}\mathbf{i} + \frac{-mMGy}{(x^2+y^2+z^2)^{3/2}}\mathbf{j} + \frac{-mMGz}{(x^2+y^2+z^2)^{3/2}}\mathbf{k}. \end{aligned}$$

If f is a scalar function of two or three variables, the gradient ∇f is a vector field defined by

$$\nabla f(x, y) = f_x\mathbf{i} + f_y\mathbf{j}$$

or

$$\nabla f(x, y, z) = f_x\mathbf{i} + f_y\mathbf{j} + f_z\mathbf{k}.$$

We say that a vector field $\mathbf{F}$ is conservative, if it is the gradient of some scalar function f, that is, $\mathbf{F} = \nabla f$ for some function f. In this case, f is called a potential function of $\mathbf{F}$.

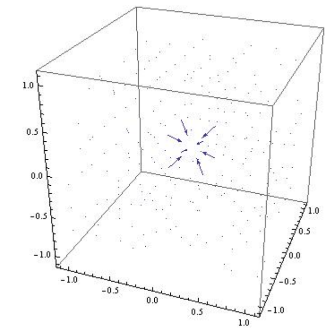


FIGURE 16. Gravitational force field

Example 6.5. In the case of the gravitational force field, one may check that

$$f = \frac{mMG}{\sqrt{x^2 + y^2 + z^2}}$$

is a potential function.

$$\nabla f = \frac{-mMGx}{(x^2 + y^2 + z^2)^{3/2}}\mathbf{i} + \frac{-mMGy}{(x^2 + y^2 + z^2)^{3/2}}\mathbf{j} + \frac{-mMGz}{(x^2 + y^2 + z^2)^{3/2}}\mathbf{k}.$$

There are two important operations that can be performed on vector fields. We may view the gradient ∇ as a vector-valued operator so that we can perform operations on vector fields.

$$\nabla = \langle \frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial z} \rangle$$

Definition 6.6. If $\mathbf{F} = P\mathbf{i} + Q\mathbf{j} + R\mathbf{k}$, such that P, Q, R has first order partial derivatives. We define the divergence of the vector field $\mathbf{F}$ to be

$$\operatorname{div} \mathbf{F} = \nabla \cdot \mathbf{F} = \frac{\partial P}{\partial x} + \frac{\partial Q}{\partial y} + \frac{\partial R}{\partial z}.$$

Example 6.7. If $\mathbf{F} = xz\mathbf{i} + xyz\mathbf{j} - y^2\mathbf{k}$, find $\operatorname{div} \mathbf{F}$.

Solution. We compute

$$\operatorname{div} \mathbf{F} = \nabla \cdot \mathbf{F} = z + xz.$$

Definition 6.8. If $\mathbf{F} = P\mathbf{i} + Q\mathbf{j} + R\mathbf{k}$, such that P, Q, R has first order partial derivatives. We define the curl of the vector field $\mathbf{F}$ to be

$$\operatorname{curl} \mathbf{F} = \nabla \times \mathbf{F} = det \begin{pmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ \frac{\partial}{\partial x} & \frac{\partial}{\partial y} & \frac{\partial}{\partial z} \\ P & Q & R \end{pmatrix}$$

$$= \left(\frac{\partial R}{\partial y} - \frac{\partial Q}{\partial z}\right)\mathbf{i} + \left(\frac{\partial P}{\partial z} - \frac{\partial R}{\partial x}\right)\mathbf{j} + \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y}\right)\mathbf{k}$$

Example 6.9. Let $\mathbf{F} = xz\mathbf{i} + xyz\mathbf{j} - y^2\mathbf{k}$. Find $\operatorname{curl} \mathbf{F}$.

Solution.

$$\operatorname{curl} \mathbf{F} = \nabla \times \mathbf{F} = det \begin{pmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ \frac{\partial}{\partial x} & \frac{\partial}{\partial y} & \frac{\partial}{\partial z} \\ xz & xyz & -y^2 \end{pmatrix}$$

$$= (-2y - xy)\mathbf{i} - (0 - x)\mathbf{j} + (yz - 0)\mathbf{k}$$
$$= (-2y - xy)\mathbf{i} + x\mathbf{j} + yz\mathbf{k}$$

Recall that in cross product, if two vectors are parallel, then the product is zero. In this case, we have a similar result for curl.

Theorem 6.10. *If f is a function with continuous second partial derivatives, then*
$$\mathrm{curl}(\nabla f) = \mathbf{0}.$$

Proof. Indeed,
$$\mathrm{curl}\,\nabla f = \nabla \times \nabla f = det \begin{pmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ \frac{\partial}{\partial x} & \frac{\partial}{\partial y} & \frac{\partial}{\partial z} \\ \frac{\partial f}{\partial x} & \frac{\partial f}{\partial y} & \frac{\partial f}{\partial z} \end{pmatrix}$$
$$= \left(\frac{\partial^2 f}{\partial y \partial z} - \frac{\partial^2 f}{\partial z \partial y}\right)\mathbf{i} + \left(\frac{\partial^2 f}{\partial z \partial x} - \frac{\partial^2 f}{\partial x \partial z}\right)\mathbf{j} + \left(\frac{\partial^2 f}{\partial x \partial y} - \frac{\partial^2 f}{\partial y \partial x}\right)\mathbf{k} = \mathbf{0}.$$
□

In other words, if $\mathbf{F}$ is conservative, then $\mathrm{curl}\,\mathbf{F} = \mathbf{0}$. In fact, the reverse is also true. The proof of this theorem requires Stokes' theorem which will be introduced in later sections.

Theorem 6.11. *If $\mathbf{F}$ is a vector field whose components have continuous partial derivatives, then $\mathbf{F}$ is conservative if and only if $\mathrm{curl}\,\mathbf{F} = \mathbf{0}$.*

Another property of $\mathrm{curl}\,\mathbf{F}$ is that the divergence of this vector field must be 0. In this case, one may check whether a vector field can be the curl of some other vector field or not.

Theorem 6.12. *If $\mathbf{F}$ is a vector field whose components have continuous second order partial derivatives, then*
$$\mathrm{div}\,\mathrm{curl}\,\mathbf{F} = 0.$$

Proof. The proof is straight forward computation.
$$\mathrm{div}\,\mathrm{curl}\,\mathbf{F} = \nabla \cdot (\nabla \times \mathbf{F})$$
$$= \frac{\partial}{\partial x}\left(\frac{\partial R}{\partial y} - \frac{\partial Q}{\partial z}\right) + \frac{\partial}{\partial y}\left(\frac{\partial P}{\partial z} - \frac{\partial R}{\partial x}\right) + \frac{\partial}{\partial z}\left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y}\right) = 0$$
□

Example 6.13. If $\mathbf{F} = xz\mathbf{i} + xyz\mathbf{j} - y^2\mathbf{k}$, then $\mathrm{div}\,\mathbf{F} = z + xz \neq 0$. So $\mathbf{F}$ can not be the curl of some other vector field $\mathbf{F} \neq \mathrm{curl}\,\mathbf{G}$.

There is one last operator which is defined using the divergence.

Definition 6.14. If f is a scalar function with continuous second order partial derivatives, we define the Laplace of f to be $\triangle f$, where
$$\triangle f = \mathrm{div}\,\nabla f = \nabla \cdot \nabla f = \nabla^2 f = \frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2} + \frac{\partial^2 f}{\partial z^2}.$$

6.2. **Line Integrals.** In this section we will study a new type of integral which is similar to a single variable integral except that instead of integrating over an interval, we integrate over a curve. Such integrals are called line integrals. We have seen that the double and triple integral viewed as the limit of a Riemann sum, represents some kind of total quantity of a distribution over a region. The line integral is a way to find this total quantity along a curve.

Recall that a curve C in the plane can be parameterized by a vector valued function

$$\mathbf{r}(t) = \langle x(t), y(t) \rangle,$$

or in component forms,

$$x = x(t), y = y(t), a \leq t \leq b.$$

Let us assume that the curve is regular, in other words, $\mathbf{r}'$ is continuous and $\mathbf{r}' \neq \mathbf{0}$. Assume that $f(x, y)$ is a function defined on the curve C. Let us compute the Riemann sum along the curve C. We subdivide the parameter interval $[a, b]$ into n subintervals $[t_{i-1}, t_i]$ of length Δt. This allows us to subdivide the curve C into n arcs $P_{i-1}P_i$. Assume that the length of each arc is given by Δs_i, and we pick a sample point (x_i, y_i) from each arc. Then the Riemann sum can be written as

$$\sum_{i=1}^{n} f(x_i, y_i) \Delta s_i$$

Here the arc length can be approximated using a tangent vector at P_{i-1},

$$\Delta s = \sqrt{\left(\frac{\partial x}{\partial t}(t_{i-1})\right)^2 + \left(\frac{\partial y}{\partial t}(t_{i-1})\right)^2} \Delta t.$$

And we define the limit of the Riemann sum to be the following.

Definition 6.15. Let f be a function defined along a regular curve C. We define the line integral of f along C to be

$$\int_C f(x, y) ds = \lim_{n \to \infty} \sum_{i=1}^{n} f(x_i, y_i) \Delta s_i$$

$$= \int_a^b f(x(t), y(t)) \sqrt{\left(\frac{\partial x}{\partial t}\right)^2 + \left(\frac{\partial y}{\partial t}\right)^2} dt.$$

The value of the line integral does not depend on the parametrization of the curve C as we have studied that the arc length is a geometric quantity. In fact, if s is the arc length parameter, then

$$ds = \sqrt{\left(\frac{\partial x}{\partial t}\right)^2 + \left(\frac{\partial y}{\partial t}\right)^2} dt.$$

In the case when $f(x, y) = 1$, the line integral

$$\int_C ds = \int_a^b \sqrt{\left(\frac{\partial x}{\partial t}\right)^2 + \left(\frac{\partial y}{\partial t}\right)^2} dt,$$

precisely gives us the arc length of the curve.

Remark 6.16. In the case where C is the x-axis, it can be parameterized by $x = x, y = 0$, $a \leq x \leq b$, so $ds = dx$ and

$$\int_C f(x,y)ds = \int_a^b f(x,0)dx$$

becomes the single variable integral. So the line integral can be viewed as a generalization of the single variable integral. In particular, if $f \geq 0$, then the line integral represents the area of the cylindrical region between the graph of f along the curve C.

Example 6.17. Compute $\int_C (2+x^2y)ds$, where C is the upper half circle $x^2+y^2 = 1, y \geq 0$.

Solution. We may parameterize the circle using

$$x = \cos t, y = \sin t, t \in [0,\pi].$$

Then the line integral becomes,

$$\begin{aligned}\int_C (2+x^2y)ds &= \int_0^\pi (2+\cos^2 t\sin t)\sqrt{\sin^2 t+\cos^2 t}dt \\ &= \int_0^\pi (2+\cos^2 t\sin t)dt = (2t - \frac{1}{3}\cos^3 t)_0^\pi = 2\pi + \frac{2}{3}.\end{aligned}$$

If C is a piecewise-smooth curve, that is, C is a union of finitely many consecutive regular curves $C_1, C_2, \ldots, C_k$. Then the line integral of f along C is

$$\int_C f(x,y)ds = \int_{C_1} f(x,y)ds + \int_{C_2} f(x,y)ds + \cdots + \int_{C_k} f(x,y)ds.$$

Example 6.18. Compute $\int_C 2xds$, where C consists of the arc C_1 on the parabola $y = x^2, 0 \leq x \leq 1$ and a vertical segment from $(1,1)$ to $(1,2)$.

Solution. In this case, $\int_C 2xds = \int_{C_1} 2xds + \int_{C_2} 2xds$. We compute the line integral on each arc.

The parabola can be parameterized by

$$x = x, y = x^2, 0 \leq x \leq 1.$$

So

$$\int_{C_1} 2xds = \int_0^1 2x\sqrt{1+4x^2}dx = \frac{1}{4}\cdot\frac{2}{3}(1+4x^2)^{3/2}|_0^1 = \frac{5\sqrt{5}-1}{6}.$$

Now along the second arc,

$$x = 1, y = y, 1 \leq y \leq 2.$$

$$\int_{C_2} 2xds = \int_1^2 2\cdot 1\cdot\sqrt{0+1}dy = \int_1^2 2dy = 2.$$

Therefore,

$$\int_C 2xds = \frac{5\sqrt{5}-1}{6} + 2.$$

Example 6.19. Similar to the case of double and triple integral, if $\rho(x,y)$ is a density function along a line C, then we define the mass of C to be the line integral

$$m = \int_C \rho(x,y)ds$$

and the center of mass is a point with coordinate $(\bar{x}, \bar{y})$ such that

$$\bar{x} = \frac{1}{m}\int_C x\rho(x,y)ds, \bar{y} = \frac{1}{m}\int_C y\rho(x,y)ds.$$

Note that the center of mass may not be on the curve C.

Now suppose that the curve C is a regular curve in the space $\mathbb{R}^3$ given by the parametrization

$$x = x(t), y = y(t), z = z(t), a \leq t \leq b.$$

In this case the line integral of a function $f(x,y,z)$ along the curve C is

$$\int_C f(x,y,z)ds = \int_a^b f(x(t),y(t),z(t))\sqrt{\left(\frac{\partial x}{\partial t}\right)^2 + \left(\frac{\partial y}{\partial t}\right)^2 + \left(\frac{\partial z}{\partial t}\right)^2}dt.$$

Sometime we write it in a shorter form

$$\int_C f(x,y,z)ds = \int_a^b f(\mathbf{r}(t))|\mathbf{r}'(t)|dt.$$

And the arc length parameter satisfies

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

Example 6.20. Evaluate the line integral $\int_C y\sin z ds$, where C is an arc on the helix $x = \cos t, y = \sin t, z = t, 0 \leq t \leq 2\pi$.

Solution.

$$\int_C y\sin z ds = \int_0^{2\pi} \sin t \cdot \sin t\sqrt{\sin^2 t + \cos^2 t + 1}dt$$

$$= \frac{\sqrt{2}}{2}(t - \frac{1}{2}\sin 2t)|_0^{2\pi} = \sqrt{2}\pi.$$

In order to study the line integral of a vector field along a curve C. We first consider another type of the line integral for a scalar function f.

Definition 6.21. Let $f(x,y)$ be a function along a regular curve C. We define the line integral of f along C with respect to x and y to be the integrals

$$\int_C f(x,y)dx = \int_a^b f(x(t),y(t))x'(t)dt$$

$$\int_C f(x,y)dy = \int_a^b f(x(t),y(t))y'(t)dt.$$

Note that we may view

$$dx = x'(t)dt, dy = y'(t)dt.$$

Sometimes we write the line integrals with respect to x and y together. For example,

$$\int_C P(x,y)dx + \int_C Q(x,y)dy = \int_C P(x,y)dx + Q(x,y)dy.$$

Example 6.22. Find the line integral $\int_C y^2dx + xdy$ for (1)$C = C_1$ is a segment from $(-5,-3)$ to $(0,2)$; (2) $C = C_2$ is an arc of $x = 4 - y^2$ from $(-5,-3)$ to $(0,2)$.

Solution. A directional vector of C_1 is $\langle 0-(-5), 2-(-3)\rangle = \langle 5,5\rangle$. The segment C_1 can be parameterized by

$$x = 5t - 5, y = 5t - 3, 0 \le t \le 1.$$

So $dx = 5dt, dy = 5dt$.

$$\int_{C_1} y^2dx + xdy = \int_0^1 (5t-3)^2 \cdot 5 \cdot dt + (5t-5)\cdot 5 \cdot dt$$

$$= 5\int_0^1 (25t^2 - 25t + 4)dt = -\frac{5}{6}.$$

Now the parabola may be parameterized by

$$x = 4 - y^2, y = y, -3 \le y \le 2.$$

So $dx = -2ydy$ and

$$\int_{C_2} y^2dx + xdy = \int_{-3}^2 y^2(-2y)dy + (4-y^2)dy$$

$$= \int_{-3}^2 (-2y^3 - y^2 + 4)dy = 40 + \frac{5}{6}.$$

Note that here the line integrals have different values for the same function between the same endpoints, but along different paths. This suggests that the line integral depends not only on the endpoints, but also the path.

In fact, one may check that if we reverse the initial and terminal point of the curve C, then the resulting line integral will be the negative of the original one. The parametrization of$C$ determines an orientation of C. We often write $-C$ the curve consists of the same points as C but in the opposite orientation. And

$$\int_{-C} fdx = -\int_C fdx, \int_{-C} fdy = -\int_C fdy.$$

Note that if we are integrating with respect to ds then

$$\int_{-C} fds = \int_C fds,$$

which does not depend on the orientation.

The line integral with respect to x and y may also be generalized to the three variables case. In this case, we simply have

$$\int_C f(x,y,z)dx = \int_a^b f(x(t),y(t),z(t))x'(t)dt$$

$$\int_C f(x,y,z)dy = \int_a^b f(x(t),y(t),z(t))y'(t)dt$$

$$\int_C f(x,y,z)dz = \int_a^b f(x(t),y(t),z(t))z'(t)dt.$$

And we may view

$$dx = x'(t)dt, dy = y'(t)dt, dz = z'(t)dt.$$

Example 6.23. Evaluate $\int_C ydx + zdy + xdz$, where C consists of two segments C_1 from $(2, 0, 0)$ to $(3, 4, 5)$ and C_2 from $(3, 4, 5)$ to $(3, 4, 0)$.

Solution. A directional vector for C_1 is $\langle 1, 4, 5\rangle$, so

$$x = 2 + t, y = 4t, z = 5t, 0 \le t \le 1.$$

$$\int_{C_1} ydx + zdy + xdz = \int_0^1 4tdt + 5t \cdot 4dt + (2 + t) \cdot 5dt = 24 + \frac{1}{2}.$$

Now for C_2, a directional vector is $\langle 0, 0, -5\rangle$, so

$$x = 3, y = 4, z = 5 - 5t, 0 \le t \le 1.$$

And $dx = dy = 0$.

$$\int_{C_2} ydx + zdy + xdz = \int_0^1 3 \cdot (-5)dt = -15.$$

Therefore, the line integral

$$\int_C ydx + zdy + xdz = 9 + \frac{1}{2}.$$

Let us now discuss the line integral of vector fields. One may think of the vector field as a force field in the space and the line integral along a curve is the total work done by a particle moving along the curve in this force field. Suppose that $\mathbf{F} = \langle P(x, y, z), Q(x, y, z), R(x, y, z)\rangle$ is a vector field, and C a curve in the space with parametrization $\mathbf{r}(t)$. Let us compute the work.

We divide C into subarcs $P_{i-1}P_i$ of arc length Δs_i and on each arc, we pick a sample point (x_i^*, y_i^*, z_i^*) with parameter t_i^*. When Δs_i is small, as the particle moves from P_{i-1} to P_i, the trajectory can be approximated by a tangent line in the direction of $\mathbf{r}'(t_i^*)$. Let $\mathbf{T}(t_i^*)$ be the unit tangent vector in the direction of $\mathbf{r}'(t_i^*)$. Then the work is approximately

$$\mathbf{F}(x_i^*, y_i^*, z_i^*) \cdot (\Delta s_i \mathbf{T}(t_i^*)) = \mathbf{F}(x_i^*, y_i^*, z_i^*) \cdot \mathbf{T}(t_i^*)\Delta s_i.$$

So that the total work is approximately

$$W \approx \sum_i \mathbf{F}(x_i^*, y_i^*, z_i^*) \cdot \mathbf{T}(t_i^*)\Delta s_i.$$

Definition 6.24. Let $\mathbf{F}$ be a vector field whose components are continuous functions and C a curve with parametrization $\mathbf{r}(t)$. We define the line integral of $\mathbf{F}$ along C to be

$$\int_C \mathbf{F} \cdot \mathbf{T} ds,$$

where $\mathbf{T}$ is the unit tangent vector to the curve $\mathbf{r}(t)$.

Note that since

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

and

$$ds = |\mathbf{r}'(t)|dt,$$

we also write

$$\int_C \mathbf{F}\cdot\mathbf{T}ds = \int_a^b \mathbf{F}\cdot\frac{\mathbf{r}'(t)}{|\mathbf{r}'(t)|}|\mathbf{r}'(t)|dt = \int_a^b \mathbf{F}\cdot\mathbf{r}'(t)dt = \int_C \mathbf{F}\cdot d\mathbf{r}.$$

Example 6.25. Find the work done by the force field $\mathbf{F} = x^2\mathbf{i} - xy\mathbf{j}$ in moving a particle along the quarter circle $\mathbf{r} = \langle \cos t, \sin t\rangle$, $0 \le t \le \frac{\pi}{2}$.

Solution. Since along the curve, $x = \cos t, y = \sin t$, the vector field

$$\mathbf{F} = \cos^2 t\mathbf{i} - \cos t\sin t\mathbf{j}.$$

$$\mathbf{r}'(t) = -\sin t\mathbf{i} + \cos t\mathbf{j}.$$

Therefore, the work is

$$\int_C \mathbf{F}\cdot d\mathbf{r} = \int_0^{\pi/2} (-\cos^2 t\sin t - \cos^2 t\sin t)dt$$

$$= \int_0^{\pi/2} -2\cos^2 t\sin tdt = \frac{2}{3}\cos^3 t|_0^{\pi/2} = -\frac{2}{3}.$$

Example 6.26. Find $\int_C \mathbf{F}\cdot d\mathbf{r}$, where $\mathbf{F} = xy\mathbf{i}+yz\mathbf{j}+xz\mathbf{k}$ and C is a curve with parametrization $x = t, y = t^2, z = t^3, 0 \le t \le 1$.

Solution. Along the curve C,

$$\mathbf{F} = t^3\mathbf{i} + t^5\mathbf{j} + t^4\mathbf{k}.$$

And

$$\mathbf{r}'(t) = \mathbf{i} + 2t\mathbf{j} + 3t^2\mathbf{k}.$$

So

$$\int_C \mathbf{F}\cdot d\mathbf{r} = \int_0^1 (t^3 + 5t^6)dt = \frac{27}{28}.$$

Finally, let us investigate the relation between the line integral of a vector field and its component functions. If $\mathbf{F} = P\mathbf{i} + Q\mathbf{j} + R\mathbf{k}$, then

$$\int_C \mathbf{F}\cdot d\mathbf{r} = \int_a^b \mathbf{F}\cdot\mathbf{r}'dt$$

$$= \int_a^b Px'(t) + Qy'(t) + Rz'(t)dt$$

$$= \int_a^b Pdx + Qdy + Rdz.$$

In other words, this is the sum of line integrals of the components with respect to x, y, z. Note that from the last expression, we obtain that if we reverse the orientation of the curve C, then

$$\int_{-C} \mathbf{F}\cdot d\mathbf{r} = -\int_C \mathbf{F}\cdot d\mathbf{r}.$$

This does not contradicts to the fact that

$$\int_C \mathbf{F}\cdot d\mathbf{r} = \int_C \mathbf{F}\cdot\mathbf{T}ds,$$

where the second integral is with respect to ds, since on $-C$, the unit tangent would become $-\mathbf{T}$.

6.3. **The fundamental theorem for line integrals.** The fundamental theorem of calculus states that

$$\int_a^b F'(t)dt = F(b) - F(a).$$

This result can be generalized to line integrals.

Theorem 6.27 (Fundamental theorem for line integrals)**.** *Let C be a regular curve with a parametrization $\mathbf{r}(t)$, $a \leq t \leq b$. Suppose that f is a differentiable function with continuous gradient ∇f along C. Then*

$$\int_C \nabla f \cdot d\mathbf{r} = f(\mathbf{r}(b)) - f(\mathbf{r}(a)).$$

Proof. We will use the fundamental theorem of calculus to proof this formula.

$$\int_C \nabla f \cdot d\mathbf{r} = \int_a^b \nabla f \cdot \mathbf{r}'(t)dt$$

$$= \int_a^b \left(\frac{\partial f}{\partial x} x'(t) + \frac{\partial f}{\partial y} y'(t) + \frac{\partial f}{\partial z} z'(t) \right) dt$$

$$= \int_a^b \frac{d}{dt} f(\mathbf{r}(t))dt$$

$$= f(\mathbf{r}(b)) - f(\mathbf{r}(a)).$$

□

Note that C is a regular curve connecting two points $A = \mathbf{r}(a)$ and $B = \mathbf{r}(b)$, the formula can be written as

$$\int_C \nabla f \cdot d\mathbf{r} = f(B) - f(A).$$

Example 6.28. Find the work done by the gravitational field $\mathbf{F}(\mathbf{x}) = -\frac{mMG}{|\mathbf{x}|^3}\mathbf{x}$ in moving a particle of mass m from $(3, 4, 12)$ to $(2, 2, 0)$.

Solution. We have computed before that the gravitational force field is conservative. In other words, $\mathbf{F} = \nabla f$ for some potential function f, where

$$f = \frac{mMG}{\sqrt{x^2 + y^2 + z^2}}.$$

Therefore, using the fundamental theorem for line integrals,

$$W = \int_C \mathbf{F} \cdot d\mathbf{r} = \int_C \nabla f \cdot d\mathbf{r}$$

$$= f(2, 2, 0) - f(3, 4, 12) = mMG(\frac{1}{2\sqrt{2}} - \frac{1}{13}).$$

As a result of the fundamental theorem, the line integral of the gradient of some function f only depends on the value of the function on the initial and terminal point of the path, but does not depend on the path itself. In this case, we say the line integral is independent of path.

Definition 6.29. If $\mathbf{F}$ is a continuous vector field on a region D, we say the line integral $\int_C \mathbf{F} \cdot d\mathbf{r}$ is independent of path if

$$\int_{C_1} \mathbf{F} \cdot d\mathbf{r} = \int_{C_2} \mathbf{F} \cdot d\mathbf{r},$$

for any C_1, C_2 that connects the same initial and terminal points.

With this notation, the line integral of a conservative vector field is independent of path. In fact, this property further implies that the line integral of a conservative vector field along a closed loop is 0. Indeed, we say a curve is closed, if its initial point is the same as the terminal point. If C is a closed curve, let us pick two points A, B on the curve C. The two points divide C into two arcs C_1 from A to B and C_2 from B to A. So

$$\int_C \mathbf{F} \cdot d\mathbf{r} = \int_{C_1} \mathbf{F} \cdot d\mathbf{r} + \int_{C_2} \mathbf{F} \cdot d\mathbf{r} = \int_{C_1} \mathbf{F} \cdot d\mathbf{r} - \int_{-C_2} \mathbf{F} \cdot d\mathbf{r} = 0,$$

since C_1 and $-C_2$ have the same initial and terminal points. Therefore, we obtained

Theorem 6.30. *The line integral $\int_C \mathbf{F} \cdot d\mathbf{r}$ is independent of path if and only if for any closed (piecewise smooth) curve C, $\int_C \mathbf{F} \cdot d\mathbf{r} = 0$.*

The physical interpretation of the independent of path is that the work done by a conservative force field only depends on the initial and terminal position of the particle. And if the path is closed, then the work is 0.

A natural question is if the line integral of a vector field $\mathbf{F}$ is independent of path, then is it true that it is conservative? That is, $\mathbf{F} = \nabla f$ for some f? This is true if the region D is "connected". We say a region D is connected, or more precisely, path connected, if for any two point in D, there is a path in D joins them.

In this case, we may use the line integral to recover the potential function of a vector field, if the integral is independent of path.

Theorem 6.31. *Let $\mathbf{F}$ be a continuous vector field on an open connected region D. If $\int_C \mathbf{F} \cdot d\mathbf{r}$ is independent of path in D, then $\mathbf{F}$ is conservative. In other words, there is a function f such that $\mathbf{F} = \nabla f$.*

Proof. Let us proof the theorem in the case of two variables. The proof for functions of three variable is the same. The idea is to build a function f by performing the line integral and use the path independence to check that this function is the desired potential function.

We fix a point $A = (a, b)$ in D. Define a function

$$f(x, y) = \int_{(a,b)}^{(x,y)} \mathbf{F} \cdot d\mathbf{r}.$$

Since the integral is independent of path, it does not matter which path we choose. We will show that $\nabla f = \mathbf{F}$. Assume that $\mathbf{F} = P\mathbf{i} + Q\mathbf{j}$.

We compute the $\frac{\partial f}{\partial x}$ in the following way. Since D is open, we may pick a point (x_1, y) near (x, y) with the same y coordinate. We connect (a, b) and (x_1, y) using a path C_1, and (x_1, y), (x, y) using a horizontal segment C_2. Then

$$f(x, y) = \int_{C_1} \mathbf{F} \cdot d\mathbf{r} + \int_{C_2} \mathbf{F} \cdot d\mathbf{r}.$$

Note that when x is changing in a small open region, we may fix the point x_1, so that the path C_1 can be fixed. This means that the first integral can be picked to be a constant k. Thus,

$$\frac{\partial f}{\partial x} = \frac{\partial}{\partial x}k + \frac{\partial}{\partial x}\int_{C_2} \mathbf{F}\cdot d\mathbf{r} = 0 + \frac{\partial}{\partial x}\int_{C_2} \mathbf{F}\cdot d\mathbf{r}.$$

Now along the path C_2, y is constant, so

$$\int_{C_2} \mathbf{F}\cdot d\mathbf{r} = \int_{C_2} P(x,y)dx + Q(x,y)dy = \int_{x_1}^{x} Pdx.$$

Therefore,

$$\frac{\partial f}{\partial x} = \frac{\partial}{\partial x}\int_{x_1}^{x} P(x,y)dx = P(x,y).$$

Similarly, we may obtain that $\frac{\partial f}{\partial y} = Q(x,y)$. Thus,

$$\mathbf{F} = P\mathbf{i} + Q\mathbf{j} = \nabla f.$$

□

Example 6.32. Let $\mathbf{F} = (3+2xy)\mathbf{i}+(x^2-3y^2)\mathbf{j}$. Find a potential function f such that $\nabla f = \mathbf{F}$. Evaluate the line integral $\int_C \mathbf{F}\cdot d\mathbf{r}$, there C is the curve given by $\mathbf{r} = e^t \sin t\mathbf{i} + e^t \cos t\mathbf{j}$, $0 \le t \le \pi$.

Solution. To find the potential function, note that if f is a potential function, then

$$f_x = 3 + 2xy$$
$$f_y = x^2 - 3y^2$$

If we integrate the first equation with respect to x, then

$$f(x,y) = 3x + x^2y + g(y).$$

Here $g(y)$ can be any function of y. To find $g(y)$, we take the $y-$ partial derivative,

$$f_y = x^2 + g'(y)$$

and we compare this with $f_y = x^2 - 3y^2$.

$$g'(y) = -3y^2 \to g(y) = -y^3 + k$$

Therefore, a potential function is

$$f(x,y) = 3x + x^2y - y^3 + k,$$

where k can be any constant.

Now to find the path integral, since we have the potential function f, the initial and terminal points are $\mathbf{r}(0) = (0,1)$ and $\mathbf{r}(\pi) = (0,-e^\pi)$. So

$$\int_C \mathbf{F}\cdot d\mathbf{r} = f(0,-e^\pi) - f(0,1) = e^{3\pi} + 1.$$

From the above example, we see that if $\mathbf{F}$ is an arbitrary vector field, we can still integrate its x or y component. But if $\mathbf{F}$ is not conservative, when we take the partial derivative, the result will not match with the other component. One useful way to check whether a vector field is conservative is that if P, Q are both differentiable, since

$$f_x = P, f_y = Q,$$

$$\frac{\partial P}{\partial y} = f_{xy} = f_{yx} = \frac{\partial Q}{\partial x}.$$

In other words,

Theorem 6.33. *If* $\mathbf{F} = P\mathbf{i} + Q\mathbf{j}$ *is a conservative vector field on a region* D*, then*

$$\frac{\partial P}{\partial y} = \frac{\partial Q}{\partial x}$$

One may ask the converse of this result is true or not. In fact, the converse of this theorem depends on the type of the region D. We say a region D is simply-connected, if for any closed curve in D, one can continuously deform the curve to a point with in D.

Theorem 6.34. *Let* $\mathbf{F} = P\mathbf{i} + Q\mathbf{j}$ *on an open, connected and simply-connected region* D*. If*

$$\frac{\partial P}{\partial y} = \frac{\partial Q}{\partial x},$$

then $\mathbf{F}$ *is conservative.*

Example 6.35. Determine whether the following vector fields are conservative or not.
(1). $\mathbf{F} = (x - y)\mathbf{i} + (x - 2)\mathbf{j}$.
We compute that $\frac{\partial P}{\partial y} = -1$ while $\frac{\partial Q}{\partial x} = 1$. So $\mathbf{F}$ is not conservative.
(2). $\mathbf{F} = (3 + 2xy)\mathbf{i} + (x^2 - 3y^2)\mathbf{j}$.
In fact, this is the vector field in the previous example. We check that

$$\frac{\partial P}{\partial y} = \frac{\partial Q}{\partial x} = 2x,$$

also the domain of $\mathbf{F}$ is the entire $\mathbb{R}^2$ which is open, connected and simply-connected. So $\mathbf{F}$ is conservative.

Example 6.36. Consider a vector field $\mathbf{F} = \frac{-y\mathbf{i}+x\mathbf{j}}{x^2+y^2}$. In this case

$$\frac{\partial P}{\partial y} = \frac{\partial Q}{\partial x} = \frac{-x^2 + y^2}{(x^2 + y^2)^2}.$$

But we cannot conclude since the domain is not simply-connected. In fact, it is not conservative, since if we integrate along the unit circle $\mathbf{r} = \cos t\mathbf{i} + \sin t\mathbf{j}$, $0 \le t \le 2\pi$,

$$\int_C \mathbf{F} \cdot d\mathbf{r} = \int_C P dx + Q dy = 2\pi \neq 0.$$

So it is not path independent.

One may check that in this case, we can even write down a "potential" function $f = \arctan(y/x)$. But the domain of this function excludes the y-axis.

Example 6.37. Let $\mathbf{F} = y^2\mathbf{i} + (2xy + e^{3z})\mathbf{j} + 3ye^{3z}\mathbf{k}$. Find a potential function such that $\nabla f = \mathbf{F}$.

Solution. If f is a potential function, then

$$f_x = y^2, f_y = 2xy + e^{3z}, f_z = 3ye^{3z}.$$

If we integrate f_x with respect to x, then

$$f(x, y, z) = xy^2 + g(y, z),$$

where $g(y,z)$ is a constant with respect to x. To find the function $g(y,z)$, we differentiate it with respect to y.

$$f_y = 2xy + g_y \Rightarrow g_y = e^{3z}.$$

If we integrate g_y with respect to y, then

$$g(y,z) = ye^{3z} + h(z) \Rightarrow f = xy^2 + ye^{3z} + h(z)$$

Now differentiate with respect to z,

$$f_z = 3ye^{3z} + h'(z) \Rightarrow h'(z) = 0 \Rightarrow h = k$$

Therefore, a potential function is $f = xy^2 + ye^{3z} + k$, where k is a constant.

Remark 6.38. For functions of three variables, the condition in Theorem 6.33, using Clairaut's relation, is given by

$$\frac{\partial P}{\partial y} = \frac{\partial Q}{\partial x}, \frac{\partial P}{\partial z} = \frac{\partial R}{\partial x}, \frac{\partial Q}{\partial z} = \frac{\partial R}{\partial y}.$$

Example 6.39 (Conservation of Energy)**.** Let us consider moving an object along a path $C : \mathbf{r} = \mathbf{r}(t)$ in a force field $\mathbf{F}$. Newton's Second Law of Motion states that the force is related to the acceleration $\mathbf{a} = \mathbf{r}''$ by

$$\mathbf{F} = m\mathbf{r}''.$$

The work done by the force is

$$\begin{aligned} W = \int_C \mathbf{F} \cdot d\mathbf{r} &= \int_a^b m\mathbf{r}'' \cdot \mathbf{r}' dt \\ &= \frac{m}{2} \int_a^b \frac{d}{dt}(\mathbf{r}' \cdot \mathbf{r}') dt \\ &= \frac{m}{2}(|\mathbf{r}'(b)|^2 - |\mathbf{r}'(a)|^2). \end{aligned}$$

The quantity $\frac{m}{2}|\mathbf{v}|^2$ is called the kinetic energy, so we write

$$W = K(B) - K(A).$$

On the other hand, if $\mathbf{F}$ is a conservative force field, then $\mathbf{F} = \nabla f$ and we define the potential energy to be $P = -f$.

$$W = \int_C \mathbf{F} \cdot d\mathbf{r} = f(B) - f(A) = P(A) - P(B).$$

By comparing both equations we obtain that

$$K(A) + P(A) = K(B) + P(B),$$

which indicates that if an object moves from A to B under a conservative force field, then the sum of its potential energy and kinetic energy remains constant. This is called the Law of Conservation of Energy.

6.4. **Green's Theorem.** As a generalization of the fundamental theorem of calculus, Green's theorem gives the relation between a line integral around a oriented simple closed curve and the double integral over the region bounded by this curve. Here, a simple curve means it has no self-intersections.

Definition 6.40. For a simple closed curve, we say that it is positively oriented, if there is a parametrization $\mathbf{r}(t)$, $a \le t \le b$ such that the region enclosed by the curve is always on the left as $\mathbf{r}(t)$ traverses C. If we view the curve on the Cartesian coordinate system, then a positive orientation refers to a counterclockwise traversal of C. Similarly, if the region is on the right side of $\mathbf{r}(t)$, or $\mathbf{r}(t)$ clockwise traverses C, then we say the curve is negatively oriented.

Theorem 6.41 (Green's Theorem)**.** *Let C be a positively oriented, piecewise-smooth, simple, closed curve in the plane, and D be the region bounded by C. If P and Q are functions with continuous partial derivatives, then*

$$\int_C Pdx + Qdy = \iint_D \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y} \right) dA.$$

Sometimes we use the notation

$$\oint_C Pdx + Qdy$$

to denote the line integral along a positively oriented closed curve C. Since the curve C is the boundary of the region D, if we use ∂ to denote the boundary, then Green's theorem can be also written as

$$\oint_{\partial D} Pdx + Qdy = \iint_D \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y} \right) dA.$$

Before we proceed to the proof of Green's theorem, let us first investigate a few examples.

Example 6.42. Evaluate $\oint_C x^4 dx + xy dy$, where C is the triangle consisting the line segments from $(0,0)$ to $(1,0)$, then to $(0,1)$ and back to $(0,0)$.

Solution. This line integral can be computed in two different ways. Let us first use the formula of the line integrals to evaluate this integral. In this case, we write the line integral as the sum of three integrals along the sides of the triangle.

$$\int_C x^4 dx + xy dy = \int_{C_1} x^4 dy + xy dy + \int_{C_2} x^4 dy + xy dy + \int_{C_3} x^4 dy + xy dy.$$

Along C_1, $y = 0, 0 \le x \le 1$,

$$\int_{C_1} x^4 dx + xy dy = \int_0^1 x^4 dx = \frac{1}{5}.$$

Along C_2, $y = (1-x)$, while x is from 1 to 0 in this orientation. So

$$\int_{C_2} x^4 dx + xy dy = \int_1^0 x^4 dx + x(1-x)(-1) dx = -\frac{1}{30}$$

Along C_3, $x = 0$, y from 1 to 0,

$$\int_{C_3} x^4 dx + xy dy = 0.$$

Now

$$\int_C x^4 dx + xydy = \frac{1}{5} - \frac{1}{30} + 0 = \frac{1}{6}.$$

Let us verify Green's theorem by computing the double integral over the region enclosed by C. Note that C is positively oriented.

$$\int_C x^4 dx + xydy = \iint_D \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y}\right) dA = \int_0^1 \int_0^{1-x} (y-0)dydx$$

$$= \frac{1}{2}\int_0^1 (1-x)^2 dx$$

$$= -\frac{1}{6}(1-x)^3|_0^1 = \frac{1}{6}.$$

Sometimes when the line integral is too complicated, we may use Green's theorem to compute a double integral instead.

Example 6.43. Evaluate $\oint_C (3y - e^{\sin x})dx + (7x + \sqrt{y^4+1})dy$, where C is the circle $x^2 + y^2 = 9$.

Solution. The region bounded by C is the disk D of radius 3. By Green's theorem,

$$\oint_C (3y - e^{\sin x})dx + (7x + \sqrt{y^4+1})dy = \iint_D (7-3)dA$$

$$= 4\iint_D dA = 36\pi.$$

We may also use Green's theorem to evaluate a double integral using the line integral along the boundary.

Example 6.44. Find the area enclosed by the ellipse $\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1$.

Solution. The area can be computed using a double integral $\iint_D dA$, where D is the region bounded by the ellipse. If a vector $\mathbf{F} = \langle P, Q\rangle$ is such that $\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y} = 1$, we may pick $Q = \frac{1}{2}x, P = -\frac{1}{2}y$. So using the parametrization $x = a\cos t, y = b\sin t$,

$$\iint_D dA = \frac{1}{2}\oint_{\partial D} -ydx + xdy$$

$$= \frac{1}{2}\int_0^{2\pi} ab\cos^2 t + ab\sin^2 t dt = \pi ab.$$

Remark 6.45. There are several ways of writing the area of a region using a line integral. In the above example, we picked $P = -\frac{1}{2}y, Q = \frac{1}{2}x$. In fact, we may use anyone of the following depending on the expression.

$$\iint_D dA = \oint_{\partial D} xdy = -\oint_{\partial D} ydx = \frac{1}{2}\oint_{\partial D} -ydx + xdy.$$

Let us now proof Green's theorem for a simple case.

Proof. Suppose that the region D is both Type I and Type II. We call such regions simple regions. Let C be the boundary of D, oriented positively. We will show that

$$\int_C Pdx = -\iint_D \frac{\partial P}{\partial y} dA, \int_C Qdy = \iint_D \frac{\partial Q}{\partial x} dA.$$

As a Type I region, assume that $D = \{(x, y)|a \leq x \leq b, g_1(x) \leq y \leq g_2(x)\}$. Then

$$\iint_D \frac{\partial P}{\partial y} dA = \int_a^b \int_{g_1}^{g_2} \frac{\partial P}{\partial x} dydx = \int_a^b P(x, g_2) - P(x, g_1)dx.$$

On the other hand, let us denote the four arcs, counterclockwise, on C using C_1, C_2, C_3, C_4, where C_1 is the graph of $y = g_1(x)$. Then

$$\int_{C_1} Pdx = \int_a^b P(x, g_1)dx,$$

$$\int_{C_3} Pdx = \int_a^b -P(x, g_2)dx.$$

Now along C_2, C_4, the variable x is constant, so $dx = 0$.

$$\int_{C_2} Pdx = \int_{C_4} Pdx = 0.$$

Therefore, by adding these integrals together,

$$\int_C Pdx = \iint_D -\frac{\partial P}{\partial y} dA.$$

Similarly, one can show $\int_C Qdy = \iint_D \frac{\partial Q}{\partial x} dA$ when D is a Type II region using the same method. □

We may extend Green's theorem for simple regions to the general case. In this case, we can try to write the region as a finite union of simple regions by adding two vertical or horizontal segments in the opposite orientation. In this case, if $D = D_1 \cup D_2$ such that the boundary of D is $C_1 \cup C_2$ and the boundary of D_1 is $C_1 \cup C_3$, and the boundary of D_2 is $(-C_3) \cup C_2$, then

$$\begin{aligned}
\iint_{D=D_1\cup D_2} \frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y} dA &= \iint_{D_1} \frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y} dA + \iint_{D_2} \frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y} dA \\
&= \int_{C_1\cup C_3} Pdx + Qdy + \int_{(-C_3\cup C_2)} Pdx + Qdy \\
&= \int_{C=C_1\cup C_2} Pdx + Qdy.
\end{aligned}$$

Example 6.46. Evaluate $\oint_C y^2dx + 3xydy$, where C is the boundary of the region D in the upper half plane between $x^2 + y^2 = 1$ and $x^2 + y^2 = 4$.

Solution. The region D can be described using polar coordinates,

$$D = \{(r, \theta)|1 \leq r \leq 2, 0 \leq \theta \leq \pi\}.$$

Using Green's theorem

$$\oint_C y^2 dx + 3xy dy = \iint_D y dA = \int_0^\pi \int_1^2 r \sin\theta r dr d\theta = \frac{14}{3}.$$

Using the above idea, we may further extend Green's theorem to the case when the region D is not simply-connected. If D is a region bounded in between two curves C_1 and C_2, assume that C_1 is the outer curve and C_2 is the inner curve. If both curves are positively oriented, that is, the region D is on the left side of the curve then they are traversed. For the outer curve C_1, this is counterclockwise direction but for C_2, it is clockwise. As a result, if we denote by $-C_2$ the curve in counterclockwise orientation, then Green's theorem is

$$\iint_D \frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y} dA = \int_{C_1} P dx + Q dy - \int_{-C_2} P dx + Q dy.$$

Example 6.47. Let $\mathbf{F} = (-y\mathbf{i} + x\mathbf{j})/(x^2 + y^2)$. Show that $\int_C \mathbf{F} \cdot d\mathbf{r} = 2\pi$ for every positively oriented close curve that encloses the origin.

Solution. Suppose that C is a positively oriented close curve that encloses the origin. Let C' be a circle of radius a lies inside C, positively oriented. The region D between C and C' has boundary $C \cup (-C')$. So Green's theorem gives

$$\int_{C\cup(-C')} P dx + Q dy = \iint_D \frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y} dA = \iint_D 0 dA = 0$$

$$\Rightarrow \int_C P dx + Q dy = \int_{C'} P dx + Q dy.$$

Now the circle C' can be parameterized by $x = a\cos t, y = a\sin t, 0 \le t \le 2\pi$.

$$\int_{C_1} \mathbf{F} \cdot d\mathbf{r} = \int_0^{2\pi} \frac{a^2 \sin^2 t + a^2 \cos^2 t}{a^2 \cos^2 t + a^2 \sin^2 t} dt = \int_0^{2\pi} dt = 2\pi$$

Green's theorem allows to proof Theorem 6.34.

Proof. Recall that Theorem 6.34 states if $\frac{\partial Q}{\partial x} = \frac{\partial P}{\partial y}$ on a simply-connected region D, then $\mathbf{F}$ is conservative. Indeed, if C is a closed circle in D, then D being simply-connected implies C bounds a disk R. Now by Green's theorem,

$$\int_C P dx + Q dy = \iint_R 0 dA = 0.$$

So the line integral of $\mathbf{F}$ along any closed curve C is 0 which implies that it is conservative. $\square$

The expression $\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y}$ in Green's theorem is similar to the component of $\operatorname{curl} \mathbf{F}$. We may in fact, express Green's theorem using the vector operators div and curl. Recall that

$$\operatorname{curl} \mathbf{F} = \nabla \times \mathbf{F} = det \begin{pmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ \frac{\partial}{\partial x} & \frac{\partial}{\partial y} & \frac{\partial}{\partial z} \\ P & Q & 0 \end{pmatrix} = \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y} \right) \mathbf{k}.$$

So

$$\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y} = \operatorname{curl} \mathbf{F} \cdot \mathbf{k}.$$

Then Green's theorem can be written as

$$\int_C \mathbf{F} \cdot d\mathbf{r} = \iint_D \operatorname{curl} \mathbf{F} \cdot \mathbf{k} dA.$$

Another way of writing Green's theorem is the following. In the expression of the line integral $\int_C \mathbf{F} \cdot d\mathbf{r} = \int_C \mathbf{F} \cdot \mathbf{T} ds$, where $\mathbf{T} = (x'(t)\mathbf{i} + y'(t)\mathbf{j})/|\mathbf{r}'|$ is the tangent vector. If we write the normal vector $\mathbf{n} = (y'(t)\mathbf{i} - x'(t)\mathbf{j})/|\mathbf{r}'|$, we may form a different line integral

$$\int_C \mathbf{F} \cdot \mathbf{n} ds = \int_C P dy - Q dx = \iint_D \frac{\partial P}{\partial x} + \frac{\partial Q}{\partial y} dA = \iint \operatorname{div} \mathbf{F} dA.$$

6.5. **Parametric surfaces.** We have described a curve using a vector valued function $\mathbf{r}(t)$. In the same way, we may describe a surface by a vector function $\mathbf{r}(u, v)$ of two parameters u and v. Suppose that

$$\mathbf{r}(u, v) = \langle x(u, v), y(u, v), z(u, v) \rangle$$

is a vector valued function defined on a region D. The functions

$$x = x(u, v), y = y(u, v), z = z(u, v)$$

are called component functions of $\mathbf{r}(u, v)$. The set of points in $\mathbb{R}^3$ whose (x, y, z) coordinates satisfy these equations are called a parametric surface. The above equations are also called parametric equation of a surface.

Example 6.48. Identify the parametric surface $\mathbf{r}(u, v) = \langle 2\cos u, v, 2\sin u \rangle$.

Solution. From the vector valued function, we obtain that the parametric equation of the surface is

$$x = 2\cos u, y = v, z = 2\sin u.$$

Thus for any point x, y, z on the surface,

$$x^2 + z^2 = 4.$$

Since $y = v$, and there is no restriction on v, the surface is a cylinder with radius 2 which is symmetric about the y-axis.

On a given parametric surface S with parametric equation $\mathbf{r}(u, v)$, there are two families for curve are very important–the curves when u is a constant or when v is a constant. These two families of curves, when $u = u_0$, the curve $\mathbf{r}(u_0, v)$ or $v = v_0$, the curve $\mathbf{r}(u, v_0)$ are called grid curves.

Example 6.49. Find a parametric equation of the sphere $x^2+y^2+z^2 = 4$ using the spherical coordinates and identify the grid curves.

Solution. The surface is a sphere of radius 2. Putting $\rho = 2$ in the spherical coordinates we obtain that

$$x = 2\sin\phi\cos\theta, y = 2\sin\phi\sin\theta, z = 2\cos\phi.$$

The corresponding vector equation is

$$\mathbf{r}(\phi, \theta) = 2\sin\phi\cos\theta\mathbf{i} + 2\sin\phi\sin\theta\mathbf{j} + 2\cos\phi\mathbf{k}.$$

The domain is $0 \le \phi \le \pi$, $0 \le \theta \le 2\pi$. The grid curves when ϕ =constant are latitude (circles). And when θ =constant are meridians (semi-circles) connecting the north and south pole.

Example 6.50. Find a parametric equation for $z = x^2 + 2y^2$.

Solution. Sometimes we may just regard x, y as parameters. In this case the parametric equations are

$$x = x, y = y, z = x^2 + 2y^2.$$

Example 6.51. An important class of surfaces we will study is called surface of revolution. A surface of revolution can be obtained by, for example, rotating the curve $y = f(x)$ about the x-axis. In this case, a parametric equation can be

$$x = x, y = f(x)\cos\theta, z = f(x)\sin\theta.$$

Let us now find the tangent of a parametric surface S with a vector equation $\mathbf{r}(u, v)$. Consider a point $\mathbf{r}(u_0, v_0)$ on the surface. To find the tangent plane that passes through $P = \mathbf{r}(u_0, v_0)$, we first find two tangent vectors.

We may use the tangent vectors to the grid curves that passes through P. When $u = u_0$, the grid curve has a parametrization given by $\mathbf{r}(u_0, v)$ with parameter v, so the tangent vector is

$$\mathbf{r}_v = \langle \frac{\partial x}{\partial v}, \frac{\partial y}{\partial v}, \frac{\partial z}{\partial v} \rangle.$$

At the point P, this is $\mathbf{r}_v(u_0, v_0)$. Similarly, the tangent vector to the grid curve $v = v_0$ or $\mathbf{r}(u, v_0)$ is $\mathbf{r}_u(u_0, v_0)$. Now the normal vector to the tangent plane is

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

and one may write down the equation of a tangent plane using the normal vector.

Definition 6.52. We say a parametric surface $\mathbf{r}(u, v)$ is smooth if the vector

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

Example 6.53. Find the tangent plane to the surface $x = u^2, y = v^2, z = u + 2v$ at the point $1, 1, 3$.

Solution. We first compute the tangent vectors (to the grid curves).

$$\mathbf{r}_u = 2u\mathbf{i} + \mathbf{k}$$

$$\mathbf{r}_v = 2v\mathbf{j} + 2\mathbf{k}.$$

So the normal vector is

$$\mathbf{r}_u \times \mathbf{r}_v = \det \begin{pmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ 2u & 0 & 1 \\ 0 & 2v & 2 \end{pmatrix} = -2v\mathbf{i} - 4u\mathbf{j} + 4uv\mathbf{k}.$$

At the point $(1, 1, 3)$, we have $x = u^2 = 1, y = v^2 = 1, z = u + 2v = 3$. So $u = 1, v = 1$. Therefore, a normal vector is $\langle -2, -4, 4 \rangle$ and an equation of the tangent plane is

$$-2(x-1) - 4(y-1) + 4(z-3) = 0 \text{ or } x + 2y - 2z + 3 = 0.$$

Note that one may also find $u = 1, v = 1$ first and then compute the cross product $\langle 2, 0, 1 \rangle \times \langle 0, 2, 2 \rangle$ to obtain the normal vector.

Let us now compute the surface area of a parametric surface $\mathbf{r}(u, v)$ define on a region D of (u, v). We start by subdividing the domain D into small rectangles R_{ij} of dimensions Δu and Δv. Let us call the part of the surface corresponds to R_{ij} a patch S_{ij}. Let (u_i^*, v_j^*) be the lower left corner of R_{ij}. The vectors

$$\mathbf{r}_u^* = \mathbf{r}_u(u_i^*, v_j^*), \mathbf{r}_v^* = \mathbf{r}_v(u_i^*, v_j^*)$$

are the two tangent vectors at the point $\mathbf{r}(u_i^*, v_j^*)$. The area of the patch S_{ij} can be approximated by the area of a parallelogram with sides $\mathbf{r}_u^*\Delta u$ and $\mathbf{r}_v^*\Delta v$, which is

$$\Delta A = |\Delta u\mathbf{r}_u^* \times \Delta v\mathbf{r}_v^*| = |\mathbf{r}_u^* \times \mathbf{r}_v^*|\Delta u\Delta v.$$

Now the area of the entire surface can be computed using the limit of the Riemann sum of ΔA.

Definition 6.54. If S is a smooth parametric surface with vector equation $\mathbf{r}(u, v)$ on a domain D, then the surface area of S is

$$\text{area}(S) = \iint_D |\mathbf{r}_u \times \mathbf{r}_v| dA,$$

where $\mathbf{r}_u = \langle x_u, y_u, z_u\rangle$ and $\mathbf{r}_v = \langle x_v, y_v, z_v\rangle$.

Example 6.55. Find the surface area of a sphere of radius a.

Solution. A parametric equation of the sphere is

$$x = a\sin\phi\cos\theta, y = a\sin\phi\sin\theta, z = a\cos\phi, 0 \le \phi \le \pi, 0 \le \theta \le 2\pi.$$

In this case

$$\begin{aligned}\mathbf{r}_\phi \times \mathbf{r}_\theta &= \det\begin{pmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ a\cos\phi\cos\theta & a\cos\phi\sin\theta & -a\sin\phi \\ -a\sin\phi\sin\theta & a\sin\phi\cos\theta & 0 \end{pmatrix} \\ &= a^2\sin^2\phi\cos\theta\mathbf{i} + a^2\sin^2\phi\sin\theta\mathbf{j} + a^2\sin\phi\cos\phi\mathbf{k},\end{aligned}$$

So

$$|\mathbf{r}_\phi \times \mathbf{r}_\theta| = \sqrt{a^4\sin^4\phi + a^4\sin^2\phi\cos^2\phi} = a^2\sin\phi.$$

Therefore, the surface area of a sphere is

$$\int_0^{2\pi}\int_0^\pi a^2\sin\phi d\phi d\theta = 4\pi a^2.$$

Example 6.56. Sometimes it is convenient to use the variables x, y to be the parameter of the surface which is the graph of the function $z = f(x, y)$. In this case, the parametric equation is

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

Then

$$\mathbf{r}_x = \langle 1, 0, f_x\rangle, \mathbf{r}_y = \langle 0, 1, f_y\rangle.$$

$$\mathbf{r}_x \times \mathbf{r}_y = \det\begin{pmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ 1 & 0 & f_x \\ 0 & 1 & f_y \end{pmatrix} = \langle -f_x, -f_y, 1\rangle.$$

And

$$|\mathbf{r}_x \times \mathbf{r}_y| = \sqrt{1 + f_x^2 + f_y^2}.$$

Therefore, we obtain that the surface area of a surface $z = f(x, y)$, $(x, y) \in D$ is

$$\iint_D \sqrt{1 + f_x^2 + f_y^2} dA.$$

Example 6.57. Consider the surface of revolution

$$x = x, y = f(x)\cos\theta, z = f(x)\sin\theta, a \le x \le b, 0 \le \theta \le 2\pi, f \ge 0.$$

One may compute that

$$\mathbf{r}_x = \langle 1, f'\cos\theta, f; \sin\theta\rangle, \mathbf{r}_\theta = \langle 0, -f\sin\theta, f\cos\theta\rangle.$$

So that

$$\mathbf{r}_x \times \mathbf{r}_\theta = \langle ff', -f\cos\theta, -f\sin\theta\rangle.$$

and

$$|\mathbf{r}_x \times \mathbf{r}_\theta| = f(x)\sqrt{1 + (f'(x))^2}.$$

As a result, the surface area is

$$\int_0^{2\pi} \int_a^b f\sqrt{1 + f'^2} dx d\theta = \int_a^b 2\pi f\sqrt{1 + f'^2} dx.$$

6.6. **Surface Integral.** In this section, let us study the surface integral. Suppose that a parametric surface S has a vector equation $\mathbf{r}(u, v) = \langle x(u, v), y(u, v), z(u, v)\rangle$, $(x, y) \in D$. Let $f(x, y, z)$ be a function defined on S. Let us consider the integral of f over the surface. Assume that the domain D is divided into subrectangles R_{ij} with dimensions Δu and Δv. We have seen that the surface is divided into corresponding patches S_{ij} with area ΔS_{ij}, where

$$\Delta S_{ij} \approx |\mathbf{r}_u \times \mathbf{r}_v| \Delta u \Delta v.$$

Let us pick a point P_{ij}^* in each surface patch and we define a double Riemann sum as the following.

Definition 6.58. The surface integral of f over the surface S is the limit

$$\iint_S f(x, y, z) dS = \lim_{m,n\to\infty} \sum_{i=1}^m \sum_{j=1}^n f(P_{ij}^*)\Delta S_{ij} = \iint_D f(\mathbf{r}(u, v))|\mathbf{r}_u \times \mathbf{r}_v| dA.$$

Remark 6.59. This is similar to the definition of the line integral

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

Also if the integrand $f = 1$,

$$\iint_S dS = \iint_D |\mathbf{r}_u \times \mathbf{r}_v| dA = \text{area}(S).$$

Let us investigate a few examples of the surface integrals.

Example 6.60. Evaluate the surface integral $\iint_S x^2 dS$, where S is the unit sphere $x^2 + y^2 + z^2 = 1$.

Solution. We may use the spherical coordinate to find a parametrization of S.

$$x = \sin\phi\cos\theta, y = \sin\phi\sin\theta, z = \cos\phi, 0 \le \phi \le \pi, 0 \le \theta \le 2\pi.$$

We have computed in previous examples that

$$|\mathbf{r}_\phi \times \mathbf{r}_\theta| = 1^2 \cdot \sin\phi = \sin\phi.$$

Therefore, the surface integral

$$\iint_S x^2 dS = \iint_D (\sin\phi\cos\theta)^2 |\mathbf{r}_\phi \times \mathbf{r}_\theta| dA$$

$$= \int_0^{2\pi}\int_0^{\pi} \sin^3\phi\cos^2\theta d\phi d\theta$$

$$= \int_0^{2\pi} \frac{1}{2}(1+\cos 2\theta)d\theta \cdot \int_0^{\pi}(\sin\phi - \sin\phi\cos^2\phi)d\phi = \frac{4\pi}{3}.$$

Example 6.61. We may generalize the center of mass to the case of surface integral. Suppose that $\rho(x,y,z)$ is a density function on a surface S. Then the mass is

$$m = \iint_S \rho dS.$$

The center of mass is a point $(\bar{x},\bar{y},\bar{z})$, where

$$\bar{x} = \frac{1}{m}\iint_S x\rho dS, \bar{y} = \frac{1}{m}\iint_S y\rho dS, \bar{z} = \frac{1}{m}\iint_S z\rho dS$$

Example 6.62. Evaluate $\iint_S y dS$, where S is the surface $z = x + y^2, 0 \le x \le 1, 0 \le y \le 2$.

Solution. In this case, we may pick x, y to be the parameters and

$$\mathbf{r} = \langle x, y, x + y^2\rangle.$$

So

$$\mathbf{r}_x = \langle 1,0,1\rangle, \mathbf{r}_y = \langle 0,1,2y\rangle.$$

$$|\mathbf{r}_x \times \mathbf{r}_y| = \sqrt{1+1+4y^2}.$$

And therefore,

$$\iint_S y dS = \int_0^1\int_0^2 y\sqrt{2+4y^2}dydx$$

$$= \sqrt{2}\int_0^1 dx \cdot \int_0^2 y\sqrt{1+2y^2}dy = \frac{12\sqrt{2}}{3}.$$

Note that in general if S is the graph of $z = f(x,y), (x,y) \in D$, then

$$\mathbf{r}(x,y) = \langle x, y, f(x,y)\rangle$$

is a parametrization of S. So in this case,

$$\mathbf{r}_x = \langle 1, 0, f_x\rangle, \mathbf{r}_y = \langle 0, 1, f_y\rangle,$$

$$\mathbf{r}_x \times \mathbf{r}_y = \langle -f_x, -f_y, 1\rangle.$$

Therefore,

$$|\mathbf{r}_x \times \mathbf{r}_y| = \sqrt{1 + f_x^2 + f_y^2},$$

$$\iint_S g(x,y)dS = \iint_D g(x,y)\sqrt{1+f_x^2+f_y^2}dA.$$

Because the surface integral is also defined using the limit of a Riemann sum, we obtain the following property which is similar to other type of integrals. If S is a piece-wise smooth surface, in other words, a union of smooth surfaces $S_1, S_2, \ldots, S_n$, where S_i intersects S_j possibly along their boundary, then

$$\iint_S f dS = \iint_{S_1} f dS + \cdots + \iint_{S_n} f dS.$$

Example 6.63. Evaluate $\iint_S z dS$, where S is the boundary of a solid between the cylinder $S_1 : x^2 + y^2 = 1$, the disk $S_2 : z = 0, x^2 + y^2 \leq 1$, and an elliptical disk $z = 1 + x$ above S_2.

Solution. Let us first find the integral over S_1. We may parameterize S_1 using

$$x = \cos\theta, y = \sin\theta, z = z.$$

To find the domain of the parameters θ, z, note that

$$0 \leq \theta \leq 2\pi, 0 \leq z \leq 1 + x \Rightarrow 0 \leq z \leq 1 + \cos\theta.$$

In this case,

$$\mathbf{r}_\theta \times \mathbf{r}_z = \det\begin{pmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ -\sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{pmatrix} = \langle \cos\theta, \sin\theta, 0 \rangle.$$

So $|\mathbf{r}_\theta \times \mathbf{r}_z| = 1$ and

$$\begin{aligned} \iint_{S_1} z dS &= \iint_D z |\mathbf{r}_\theta \times \mathbf{r}_z| dA \\ &= \int_0^{2\pi} \int_0^{1+\cos\theta} z dz d\theta \\ &= \int_0^{2\pi} \frac{1}{2}(1 + \cos\theta)^2 d\theta = \frac{3\pi}{2}. \end{aligned}$$

Now on the surface S_2, since $z = 0$,

$$\iint_{S_2} z dS = \iint_{S_2} 0 dS = 0.$$

Finally for S_3, the surface is a part of the plane $z = 1 + x$. Therefore, we can pick

$$x = x, y = y, z = 1 + x, x^2 + y^2 \leq 1.$$

The integral

$$\iint_{S_3} z dS = \iint_D (1 + x)\sqrt{1 + 1 + 0} dA$$

$$= \int_0^{2\pi} \int_0^1 (1 + r\cos\theta)\sqrt{2} r dr d\theta = \sqrt{2} \int_0^{2\pi} \frac{1}{2} + \frac{1}{3}\cos\theta d\theta = \sqrt{2}\pi.$$

Therefore,

$$\iint_S z dS = \iint_{S_1} z dS + \iint_{S_2} z dS + \iint_{S_3} z dS = (\frac{3}{2} + \sqrt{2})\pi.$$

In order to define surface integral for vector fields, let us first consider the orientation of a surface.

The orientation, in general, can be viewed as a choice of the order of the coordinate axis (or more precisely, frames, e.g, the TNB frame). On a surface, we may always pick two tangent vectors from the tangent plane. In order to form a frame, we simply need a normal vector.

Now the order of the three vectors in a frame determines the orientation up to symmetry. Note that there are only two orientations for three vectors: the left-hand orientation or the right-hand orientation. By convention, we define the orientation of a surface to be the choice of a unit normal vector $\mathbf{n}$ to the surface.

However, unlike a curve, not every surface is orientable. We say a surface is orientable, if we move a normal vector along any closed curve on the surface, the normal vector returns to itself. One can also view an orientable surface as a surface with two sides and non-orientable as a surface with one side. A famous example of a non-orientable surface is the Möbius band.

If we move a normal vector $\mathbf{n}$ along the center circle of a Möbius band, it will return to $-\mathbf{n}$.

For an orientable surface S, we say that it is oriented, if we have choose the normal vector on the surface.

If an orientable surface is the graph of $z = f(x, y)$, then a unit normal vector can be computed using

$$\mathbf{n} = \frac{\mathbf{r}_x \times \mathbf{r}_y}{|\mathbf{r}_x \times \mathbf{r}_y|} = \frac{\langle -f_x, -f_y, 1\rangle}{\sqrt{1 + f_x^2 + f_y^2}}.$$

In this case, since the z component of the normal $\mathbf{n}$ is always positive, we call this orientation an upward orientation. Similarly, if we choose the normal vector to be $-\mathbf{n}$ in the above expression, then this orientation is called a downward orientation.

If an orientable surface is represented using a vector equation $\mathbf{r}(u, v)$, then we define the orientation induced by the vector $\mathbf{r}(u, v)$ to be the choice of

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

The choice of $-\mathbf{n}$ is called the opposite orientation.

We say an orientable surface is closed, if it enclose a solid region. In other words, it is the boundary of some solid in $\mathbb{R}^3$. In this case, surface separates the space into two parts, a bounded solid and an unbounded region. The normal vector pointing towards the unbounded region is called the outward orientation, or positive orientation. And the normal vector pointing towards the bounded region is called the inward orientation or negative orientation.

Example 6.64. For the sphere of radius a with $\mathbf{r}(\phi, \theta)$,

$$x = a \sin\phi \cos\theta, y = a \sin\phi \sin\theta, z = a \cos\phi.$$

The choice of the normal vector

$$\mathbf{n} = \frac{\mathbf{r}_\phi \times \mathbf{r}_\theta}{|\mathbf{r}_\phi \times \mathbf{r}_\theta|} = \frac{1}{a}\mathbf{r}$$

is the orientation induced by the parametrization $\mathbf{r}(\phi, \theta)$. This is also the outward or positive orientation.

If we choose the opposite orientation given by $-\mathbf{n}$, then it is the inward or negative orientation.

With this notation, we may now define the surface integral of a vector field $\mathbf{F}$.

Definition 6.65. Let $\mathbf{F}$ be a continuous vector field defined on an oriented surface S with the unit normal vector $\mathbf{n}$. The surface integral of $\mathbf{F}$ over S is

$$\iint_S \mathbf{F} \cdot \mathbf{n} dS = \iint_S \mathbf{F} \cdot d\mathbf{S}.$$

This integral is also called the flux of $\mathbf{F}$ across S.

If we view the vector field $\mathbf{F}$ as the velocity field of a fluid with constant density, then $\mathbf{F} \cdot \mathbf{n}\Delta S$ measures the mass of fluid per unit time crossing the patch S_{ij} in the direction of $\mathbf{n}$. The surface integral can be interpreted as the rate of flow through the surface S.

Computational wise, since $\mathbf{n} = \frac{\mathbf{r}_u \times \mathbf{r}_v}{|\mathbf{r}_u \times \mathbf{r}_v|}$ and $dS = |\mathbf{r}_u \times \mathbf{r}_v| dA$, we can write

$$\iint_S \mathbf{F} \cdot d\mathbf{S} = \iint_S \mathbf{F} \cdot \mathbf{n} dS = \iint_D \mathbf{F}(\mathbf{r}(u, v)) \cdot (\mathbf{r}_u \times \mathbf{r}_v) dA.$$

This can be also compared with the line integral where we have

$$\int_C \mathbf{F} \cdot d\mathbf{r} = \int_a^b \mathbf{F}(\mathbf{r}(t)) \cdot \mathbf{r}' dt.$$

Example 6.66. Find the flux of the vector field $\mathbf{F} = \langle z, y, x \rangle$ across the unit sphere $x^2 + y^2 + z^2 = 1$.

Solution. We have computed for the unit sphere, if we use the parametrization

$$x = \sin\phi\cos\theta, y = \sin\phi\sin\theta, z = \cos\phi, 0 \le \phi \le \pi, 0 \le \theta \le 2\pi,$$

then

$$\mathbf{r}_\phi \times \mathbf{r}\theta = \sin^2\phi\cos\theta\mathbf{i} + \sin^2\phi\sin\theta\mathbf{j} + \sin\phi\cos\phi\mathbf{k}.$$

And since

$$\mathbf{F} = z\mathbf{i} + y\mathbf{j} + x\mathbf{k} = \cos\phi\mathbf{i} + \sin\phi\sin\theta\mathbf{j} + \sin\phi\cos\theta\mathbf{k},$$
$$\mathbf{F} \cdot (\mathbf{r}_\phi \times \mathbf{r}_\theta) = \cos\phi\sin^2\phi\cos\theta + \sin^3\phi\sin^2\theta + \sin^2\phi\cos\phi\cos\theta.$$

Therefore,

$$\begin{aligned}\iint_S \mathbf{F} \cdot d\mathbf{S} &= \iint_D \mathbf{F} \cdot (\mathbf{r}_\phi \times \mathbf{r}_\theta) dA \\ &= \int_0^{2\pi}\int_0^{\pi} (2\sin^2\phi\cos\phi\cos\theta + \sin^3\phi\sin^2\theta) d\phi d\theta \\ &= \int_0^{\pi} \sin^3\phi d\phi \cdot \int_0^{2\pi} \sin^2\theta d\theta = \frac{4\pi}{3}.\end{aligned}$$

If the surface is given by the graph of $z = f(x, y)$ with upward orientation, then

$$\mathbf{r}_x \times \mathbf{r}_y = \langle -f_x, -f_y, 1 \rangle.$$

Therefore, if $\mathbf{F} = \langle P, Q, R \rangle$, then the surface integral can be expresses using

$$\iint_S \mathbf{F} \cdot d\mathbf{S} = \iint_D (-Pf_x - Qf_y + R) dA.$$

If the surface is oriented with a downward normal vector, then we multiply the above equation by -1.

Example 6.67. Evaluate $\iint_S \mathbf{F} \cdot d\mathbf{S}$, where $\mathbf{F} = \langle y, x, z\rangle$, and S is the boundary of the solid enclosed by $z = 1 - x^2 - y^2$ and $z = 0$ with outward orientation.

Solution. We compute the surface integral in two parts. Let S_1 be the paraboloid $z = 1 - x^2 - y^2$, $x^2 + y^2 \leq 1$. The domain of x, y here can be found by putting $z = 0$ into the equation of the paraboloid, which will give us the boundary of the disk $x^2 + y^2 = 1$.

On S_1, the vector field is $\mathbf{F} = \langle y, x, 1 - x^2 - y^2\rangle$ so

$$\begin{aligned}
\iint_{S_1} \mathbf{F} \cdot d\mathbf{S} &= \iint_D (-y(-2x) - x(-2y) + 1 - x^2 - y^2) dA \\
&= \iint_D (1 + 4xy - x^2 - y^2) dA \\
&= \int_0^{2\pi} \int_0^1 (1 + 4r^2 \cos\theta \sin\theta - r^2) r dr d\theta \\
&= \frac{\pi}{2}.
\end{aligned}$$

Now let S_2 be the portion of the surface $z = 0$, $x^2 + y^2 \leq 1$. Note that the surface is oriented with downward normal vector $\mathbf{n} = \langle 0, 0, -1\rangle$, so

$$\begin{aligned}
\iint_{S_2} \mathbf{F} \cdot d\mathbf{S} &= \iint_{S_2} \mathbf{F} \cdot \langle 0, 0, -1\rangle dS \\
&= \iint_D -z dA = 0.
\end{aligned}$$

As a result we obtain that $\iint_S \mathbf{F} \cdot dS = \frac{\pi}{2}$.

Example 6.68. In physics, surface integral is also used to compute the net charge. If $\mathbf{E}$ is an electric field on a surface S, then the integral

$$\iint_S \mathbf{E} \cdot d\mathbf{S}$$

is called the electric flux of $\mathbf{E}$ through S. The Gauss's law states that the net charge enclosed by a surface S is

$$Q = \varepsilon_0 \iint_S \mathbf{E} \cdot d\mathbf{S},$$

where $\varepsilon_0 \approx 8.854 \times 10^{-12} C^2/N \cdot m^2$ is a constant.

Example 6.69. Another application of the surface integral appears in the study of heal flow. The heat flow is a vector field

$$-K\nabla u,$$

where $u(x, y, z)$ is the temperature, and K is a constant which is called the conductivity of the substance. The rate of heat flow across the surface S is given by

$$-K \iint_S \nabla u \cdot d\mathbf{S}.$$

6.7. **Stokes' theorem.** Stokes' theorem can be viewed as a generalization of Green's theorem to the case of surface integrals. Recall that if we write Green's theorem using the vector operator curl. Then it states that

$$\int_C \mathbf{F} \cdot d\mathbf{r} = \iint_D \operatorname{curl} \mathbf{F} \cdot \mathbf{k} dA.$$

Note that the vector $\mathbf{k}$ is the unit normal vector to the xOy-plane, and the integral $\iint_D \operatorname{curl} \mathbf{F} \cdot \mathbf{k} dA = \iint_D \operatorname{curl} \mathbf{F} \cdot d\mathbf{S}$.

Theorem 6.70 (Stokes' Theorem)**.** *Let S be an oriented piecewise smooth surface with a boundary C. Assume that C is a simple, closed, piecewise smooth curve with positive orientation. Let $\mathbf{F}$ be a smooth vector field defined on an open region which contains S. Then*

$$\int_C \mathbf{F} \cdot d\mathbf{r} = \iint_S \operatorname{curl} \mathbf{F} \cdot d\mathbf{S}.$$

The positive orientation of the curve C sometimes is also called the induced orientation from the surface S. In particular, the normal vector of the surface S and the orientation of the curve C satisfies the right-hand rule. Since the positively oriented curve C is the boundary of the surface S, Stokes' theorem sometimes is written as

$$\int_{\partial S} \mathbf{F} \cdot d\mathbf{r} = \iint_S \operatorname{curl} \mathbf{F} \cdot d\mathbf{S}.$$

Example 6.71. Evaluate the line integral $\int_C \mathbf{F} \cdot d\mathbf{r}$, where $\mathbf{F} = \langle -y^2, x, z^2 \rangle$ and C is the curve of intersection of the plane $y + z = 2$ and the cylinder $x^2 + y^2 = 1$. C is oriented counterclockwise when viewed from above.

Solution. Let us compute the integral in two different ways to verify Stokes' theorem. If we apply Stokes' theorem, then the curve C is the boundary of the region S in the plane $y + z = 2$ bounded within the cylinder with upward orientation. In this case, we may parameterize S using

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

Therefore,

$$\operatorname{curl} \mathbf{F} = \det \begin{pmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ \frac{\partial}{\partial x} & \frac{\partial}{\partial y} & \frac{\partial}{\partial z} \\ -y^2 & x & z^2 \end{pmatrix} = (1 + 2y)\mathbf{k}.$$

By Stokes' theorem,

$$\begin{aligned}
\int_C \mathbf{F} \cdot d\mathbf{r} = \iint_S \operatorname{curl} \mathbf{F} \cdot d\mathbf{S} &= \iint_D (1 + 2y) dA \\
&= \int_0^{2\pi} \int_0^1 (1 + 2r \sin\theta) r dr d\theta \\
&= \int_0^{2\pi} \frac{1}{2} + \frac{2}{3} \sin\theta d\theta = \pi.
\end{aligned}$$

Now let us compute the line integral directly to verify Stokes' theorem. We consider a parametrization of the curve C.

$$x = \cos t, y = \sin t, z = 2 - \sin t, 0 \leq t \leq 2\pi.$$

Therefore,

$$\int_C \mathbf{F}\cdot d\mathbf{r} = \int_0^{2\pi} (-\sin^2 t)(-\sin t) + \cos^2 t + (2-\sin t)^2(-\cos t)dt$$

$$= \int_0^{2\pi} \cos^2 t dt = \pi.$$

Example 6.72. Evaluate the integral $\iint_S \operatorname{curl}\mathbf{F}\cdot d\mathbf{S}$, where $\mathbf{F} = \langle xz, yz, xy\rangle$ and S is the part of the sphere $x^2+y^2+z^2=4$ inside the cylinder $x^2+y^2=1$ above the xOy-plane with upward orientation.

Solution. We compute this integral using Stokes' theorem. The boundary of the surface S is given by the intersection of $x^2+y^2+z^2=4$ and $x^2+y^2=1$. Therefore, C is a circle $z=\sqrt{3}$ and $x^2+y^2=1$ oriented counterclockwise when viewed from above. A parametrization of the curve C is

$$x=\cos t, y=\sin t, z=\sqrt{3}, 0\le t\le 2\pi.$$

$$\iint_S \operatorname{curl}\mathbf{F}\cdot d\mathbf{S} = \int_C \mathbf{F}\cdot d\mathbf{r}$$

$$= \int_0^{2\pi} \sqrt{3}\cos t\cdot(-\sin t) + \sqrt{3}\sin t\cdot\cos t + \cos t\sin t\cdot 0dt$$

$$\int_0^{2\pi} 0dt = 0.$$

Stokes' theorem in fact tells us that if S_1 and S_2 are two different surfaces that shares the same boundary C with the same orientation. Then

$$\iint_{S_1} \operatorname{curl}\mathbf{F}\cdot d\mathbf{S} = \iint_{S_2} \operatorname{curl}\mathbf{F}\cdot d\mathbf{S}.$$

In fluid dynamics, if $\mathbf{F}$ is the velocity vector field of some fluid flow, then the line integral

$$\int_C \mathbf{F}\cdot d\mathbf{r} = \int_C \mathbf{F}\cdot\mathbf{T}ds$$

is called the circulation of $\mathbf{F}$ around C. Let us now consider a point P in the fluid. Let S_a be a small disk of radius a centered at P. We approximate the curl of $\mathbf{F}$ on S_a using $\operatorname{curl}\mathbf{F}(P)$ at the point P.

By Stokes' theorem,

$$\int_{C_a} \mathbf{F}\cdot d\mathbf{r} = \iint_{S_a} \operatorname{curl}\mathbf{F}\cdot d\mathbf{S}$$

$$\approx \iint_{S_a} \operatorname{curl}\mathbf{F}(P)\cdot\mathbf{n}dS = \operatorname{curl}\mathbf{F}(p)\cdot\mathbf{n}\pi a^2.$$

This gives us an interpretation of the curl of a vector field,

$$\operatorname{curl}\mathbf{F}(P)\cdot\mathbf{n} \approx \lim_{a\to 0}\frac{1}{\pi a^2}\int_{C_a}\mathbf{F}\cdot d\mathbf{r},$$

which indicates that the quantity $\operatorname{curl}\mathbf{F}\cdot\mathbf{n}$ is measuring the circulation of a vector field $\mathbf{F}$ about the axis $\mathbf{n}$. The direction which maximizes the circulation effect is in the direction of $\operatorname{curl}\mathbf{F}$.

Finally, Stokes' theorem allows us to immediately obtain Theorem 6.34. Recall that Theorem 6.34 states if a vector field satisfies $\frac{\partial Q}{\partial x} = \frac{\partial P}{\partial y}$ on a simply-connected region then it is conservative. Note that this can be generalized to

Theorem 6.73. *If* $\mathbf{F}$ *is a vector field such that* $\operatorname{curl}\mathbf{F} = \mathbf{0}$ *on a simply-connected region in* $\mathbb{R}^3$*, then* $\mathbf{F}$ *is conservative.*

Proof. Indeed, if the region in simply-connected, then any smooth closed curve C in the region bounds a disk S. Then by Stokes' theorem, the line integral

$$\int_C \mathbf{F}\cdot d\mathbf{r} = \iint_S \operatorname{curl}\mathbf{F}\cdot d\mathbf{S} = 0,$$

which implies that $\mathbf{F}$ is conservative. □

6.8. **Divergence Theorem.** Another way to write Green's theorem using vector operators is the following

$$\int_C \mathbf{F}\cdot\mathbf{n}ds = \iint_S \operatorname{div}\mathbf{F}dA.$$

The line integral on the left hand side is with respect to the unit normal vector. This can be generalized to the case of surface integral and we have the following theorem

Theorem 6.74 (Gauss's theorem, Divergence theorem)**.** *Let* E *be a simple solid region in* $\mathbb{R}^3$ *and* S *the boundary of* E *with outward orientation. Let* $\mathbf{F}$ *be a vector field with continuous partial derivatives defined on an open region containing* E*. Then*

$$\iint_S \mathbf{F}\cdot d\mathbf{S} = \iiint_E \operatorname{div}\mathbf{F}dV.$$

Let us investigate a few examples of Divergence theorem.

Example 6.75. Find the flux of the vector field $\mathbf{F} = \langle z, y, x\rangle$ over the unit sphere $x^2+y^2+z^2 = 1$.

Solution. Note that we have computed the flux in Example 6.66, which is $\frac{4}{3}\pi$. Let us now verify this result using the Divergence theorem.

$$\operatorname{div}\mathbf{F} = 0+1+0 = 1.$$

The unit sphere is the boundary of the unit ball E of radius 1.

$$\iint_S \mathbf{F}\cdot d\mathbf{S} = \iiint_E \operatorname{div}\mathbf{F}dV = \operatorname{vol}(E) = \frac{4\pi}{3}.$$

Example 6.76. Evaluate the integral $\iint_S \mathbf{F}\cdot d\mathbf{S}$, where $\mathbf{F} = \langle xy, y^2+e^{xz}, \sin xy\rangle$. And S is the surface of the region E bounded by $z = 1-x^2$, $z = 0$, $y = 0$ and $y+z = 2$ with outward orientation.

Solution. The surface integral is very difficult to compute directly. Using the Divergence theorem, we first find $\operatorname{div}\mathbf{F}$.

$$\operatorname{div}\mathbf{F} = y+2y+0 = 3y.$$

Now the region E can be expressed as a Type 3 region,

$$\{(x,y,z)| -1\le x\le 1, 0\le z\le 1-x^2, 0\le y\le 2-z\}.$$

Therefore,

$$\begin{aligned}\iint_S \mathbf{F}\cdot d\mathbf{S} &= \iiint_E \operatorname{div}\mathbf{F} dV = \iiint_E 3y dV \\ &= 3\int_{-1}^{1}\int_0^{1-x^2}\int_0^{2-z} y dy dz dx \\ &= 3\int_{-1}^{1}\int_0^{1-x^2}\frac{(2-z)^2}{2} dz dx \\ &= 3\int_{-1}^{1} -\frac{1}{6}((x^2+1)^3-8) dx = \frac{184}{35}.\end{aligned}$$

Similar to Green's theorem, the Divergence theorem can be generalized to the case where the solid region E is between two surfaces S_1 and S_2. Suppose that S_1 lies inside S_2, in this case, the orientation for the surface S_1 is the inward orientation. If we denote by $\iint_{S_1}\mathbf{F}\cdot d\mathbf{S}$ the surface integral with the outward normal, then the Divergence theorem states that

$$\iiint_E \operatorname{div}\mathbf{F} dV = -\iint_{S_1}\mathbf{F}\cdot d\mathbf{S} + \iint_{S_2}\mathbf{F}\cdot d\mathbf{S}.$$

In the next example, we will use this generalized Divergence theorem to obtain Gauss's Law for a single charge.

Example 6.77. Consider the electric field generated by a single charge at the origin

$$\mathbf{E} = \frac{\varepsilon Q}{|\mathbf{x}|^3}\mathbf{x},$$

where Q is the electric charge and ε is a constant, and $\mathbf{x} = \langle x, y, z\rangle$ is a position vector.

We will use the divergence theorem to show that the electric flux of $\mathbf{E}$ through any closed surface S that encloses the origin is

$$\iint_S \mathbf{E}\cdot d\mathbf{S} = 4\pi\varepsilon Q.$$

Note that if S is a surface that encloses the origin, we may always find a sphere S_a of small radius a centered at the origin and contained inside S. Let R be the region bounded between S and S_a, then

$$\iiint_R \operatorname{div}\mathbf{E} dV = \iint_S \mathbf{E}\cdot d\mathbf{S} - \iint_{S_a}\mathbf{E}\cdot d\mathbf{S}.$$

Now the divergence of $\mathbf{E}$ is $\operatorname{div}\mathbf{E} = 0$. Therefore,

$$\iint_S \mathbf{E}\cdot d\mathbf{S} = \iint_{S_a}\mathbf{E}\cdot d\mathbf{S}.$$

On the other hand, since S_a is a sphere centered at the origin, the unit normal vector is

$$\mathbf{n} = \frac{\mathbf{x}}{|\mathbf{x}|}.$$

$$\begin{aligned}\iint_{S_a}\mathbf{E}\cdot d\mathbf{S} &= \iint_{S_a}\mathbf{E}\cdot\mathbf{n} dS = \iint_{S_a}\frac{\varepsilon Q}{a^2} dS \\ &= \frac{\varepsilon Q}{a^2}4\pi a^2 = 4\pi\varepsilon Q.\end{aligned}$$

Similar to Stokes' theorem and curl $\mathbf{F}$, the Divergence theorem provides us a way to interpret div $\mathbf{F}$. Let us consider a fixed point P in the space and a small ball B_a around P. Then the divergence of $\mathbf{F}$ on B_a can be approximated by the divergence at P.

$$\iint_{S_a} \mathbf{F} \cdot d\mathbf{S} = \iiint_{B_a} \operatorname{div} \mathbf{F} dV \approx \iiint_{B_a} \operatorname{div} \mathbf{F}(P) dV = \operatorname{div} \mathbf{F}(P) \operatorname{vol}(B_a).$$

This implies that

$$\operatorname{div} \mathbf{F}(P) = \lim_{a \to 0} \frac{1}{\operatorname{vol}(B_a)} \iint_{S_a} \mathbf{F} \cdot d\mathbf{S}.$$

In other words, the divergence can be viewed as the net rate of outward flux per unit volume at P.

6.9. **Differential Forms (Not required.)** Let $f(x, y, z)$ be a differentiable function. We define the differential of f to be

$$df = \frac{\partial f}{\partial x} dx + \frac{\partial f}{\partial y} dy + \frac{\partial f}{\partial z} dz.$$

We call the expression $\omega = gdf$ the 1-form.

Let ω, η, ξ be 1-forms, we define an anti-symmetric product $\wedge$ to be a bilinear product satisfies

$$(f\omega) \wedge \eta = f(\omega \wedge \eta)$$
$$(\omega + \xi) \wedge \eta = \omega \wedge \eta + \xi \wedge \eta$$
$$(\omega \wedge \xi) \wedge \eta = \omega \wedge (\xi \wedge \eta)$$
$$\omega \wedge \eta = -\eta \wedge \omega$$

Example 6.78. $f = x^2 + y^2$ is a function. $df = 2xdx + 2ydy$ is a 1-form or the differential of f. If $g = xyz$, then $dg = yzdx + zxdy + xydz$.

$$\begin{aligned} df \wedge dg &= (2xdx + 2ydy) \wedge (yzdx + zxdy + xydz) \\ &= 2xyzdx \wedge dx + 2x^2zdx \wedge dy + 2x^2ydx \wedge dz \\ &\quad +2y^2zdy \wedge dx + 2xyzdy \wedge dy + 2xy^2dy \wedge dz \end{aligned}$$

Note that by anti-symmetry, $dx \wedge dx = -dx \wedge dx = 0$.

$$\begin{aligned} df \wedge dg &= 0 + 2x^2zdx \wedge dy + 2x^2ydx \wedge dz \\ &\quad -2y^2zdx \wedge dy + 0 + 2xy^2dy \wedge dz \\ &= (2x^2z - 2y^2z)dx \wedge dy + 2xy^2dy \wedge dz + 2x^2ydx \wedge dz. \end{aligned}$$

If ω, η are 1- forms, the expression $\omega \wedge \eta$ is called a 2-form. In general a k-form is

$$\sum_i f_i dx_{i_1} \wedge dx_{i_2} \wedge \cdots \wedge dx_{i_k},$$

where each dx_{i_j} is the differential of the coordinate functions x_{i_j}.

If ω is a k-form and η is a p-form, then the wedge product between ω and η satisfies

$$\omega \wedge \eta = (-1)^{kp} \eta \wedge \omega$$

The set of all differential forms with wedge product forms an algebra which is called Grassmann algebra or exterior algebra

$$\Omega = \bigoplus_k \Omega^k, \Omega^k = \{\omega | \omega \text{ is a k-form.}\}$$

This algebra is generated by functions f and 1-forms dg.

Let f be a function, dg a 1-form, ω a k-form. There is a unique derivative operator d on the exterior algebra, which is called the exterior derivative, that satisfies

$$d(f) = df$$

$$d(dg) = 0$$

$$d(\omega \wedge \eta) = d\omega \wedge \eta + (-1)^k \omega \wedge d\eta.$$

Example 6.79. Let $\omega = xdy - ydx$ be a 1-form. Then

$$d\omega = dx \wedge dy - dy \wedge dx = 2dx \wedge dy$$

is a 2-form.

Example 6.80. If $f = x^2 ydy \wedge dz + ydx \wedge dy$, then

$$df = (2xydx + x^2 dy) \wedge dy \wedge dz + dy \wedge dx \wedge dy = 2xydx \wedge dy \wedge dz.$$

In general if $f = f(x, y, z)$ is a function, then

$$df = \frac{\partial f}{\partial x}dx + \frac{\partial f}{\partial y}dy + \frac{\partial f}{\partial z}dz.(\nabla)$$

If $\omega = \omega_1 dx + \omega_2 dy + \omega_3 dz$ is a 1-form, then

$$d\omega = d\omega_1 \wedge dx + d\omega_2 \wedge dy + d\omega_3 \wedge dz$$

$$= (\frac{\partial \omega_1}{\partial x}dx + \frac{\partial \omega_1}{\partial y}dy + \frac{\partial \omega_1}{\partial z}dz) \wedge dx + (\frac{\partial \omega_2}{\partial x}dx + \frac{\partial \omega_2}{\partial y}dy + \frac{\partial \omega_2}{\partial z}dz) \wedge dy + (\frac{\partial \omega_3}{\partial x}dx + \frac{\partial \omega_3}{\partial y}dy + \frac{\partial \omega_3}{\partial z}dz) \wedge dz$$

$$= (\frac{\partial \omega_2}{\partial x} - \frac{\partial \omega_1}{\partial y})dx \wedge dy + (\frac{\partial \omega_3}{\partial y} - \frac{\partial \omega_2}{\partial z})dy \wedge dz + (\frac{\partial \omega_1}{\partial z} - \frac{\partial \omega_3}{\partial x})dz \wedge dx.(\text{curl})$$

If $\eta = \eta_1 dx \wedge dy + \eta_2 dy \wedge dz + \eta_3 dz \wedge dx$, then

$$d\eta = d\eta_1 \wedge dx \wedge dy + d\eta_2 \wedge dy \wedge dz + d\eta_3 \wedge dz \wedge dx$$

$$= (\frac{\partial \eta_1}{\partial z} + \frac{\partial \eta_2}{\partial x} + \frac{\partial \eta_3}{\partial y})dx \wedge dy \wedge dz.(\text{div})$$

If we define the integral of a k-form over a k-dimensional region to be the usual multi-variable integral, for example, let $f(x, y)$ be a function defined on a region D in $\mathbb{R}^2$, if $\omega = f dx \wedge dy$, we define

$$\int_D \omega = \iint_D f(x, y) dxdy.$$

Then in this case, the fundamental theorem of calculus can be written as

$$\int_D d\omega = \int_{\partial D} \omega.$$

This formula is known as the general Stokes' theorem. Here d is the exterior derivative of a differential form and ∂D is the boundary of the region D with the induced orientation.