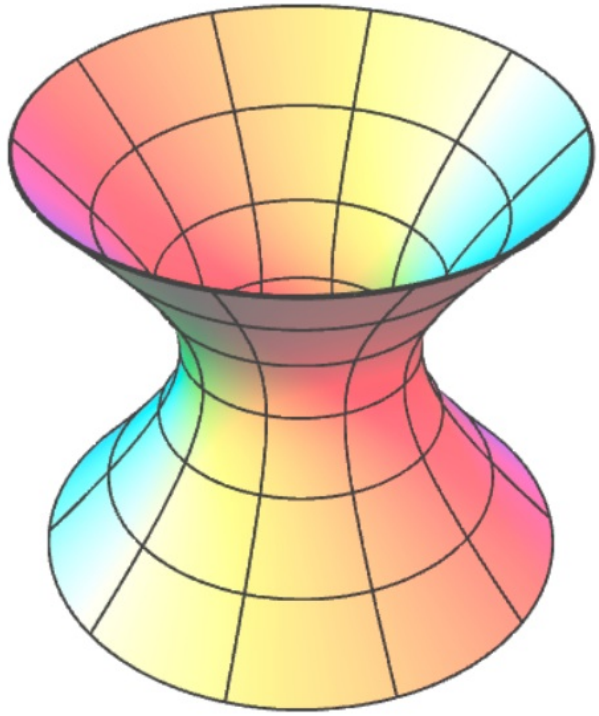


MATH 226 Multivariable Calculus

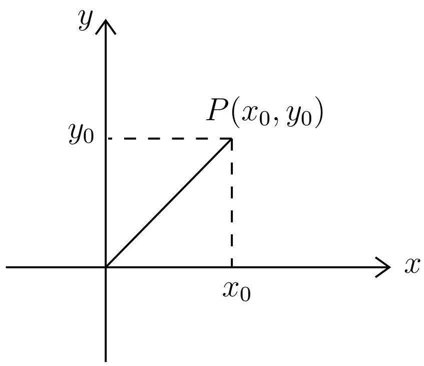
Wucheng Zhang

http://blogs.ubc.ca/wucheng/

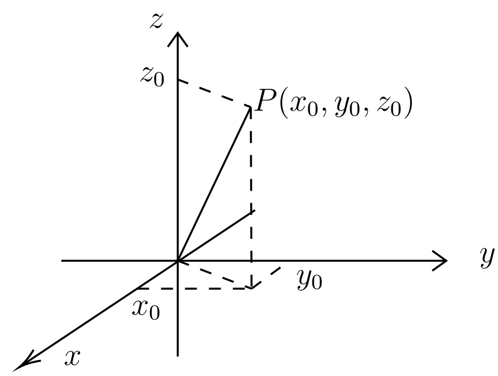
Contents

Basic Analytic Geometry

1. **Introduction and Basic Concepts**
   We wish to visualize the functions with two or three variables, for example $z = f(x, y)$ and $F(x, y, z) = 0$. Therefore we have to explore the property of not only 2 dimensional coordinates $\mathbb{R}^2$ but also 3 dimensional coordinates $\mathbb{R}^3$.



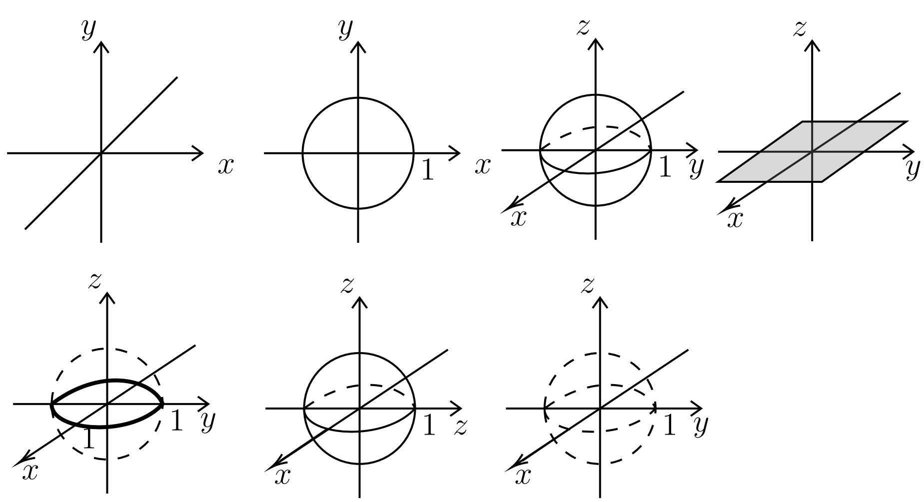
$\mathbb{R}^2$



$\mathbb{R}^3$

   First let look at some geometries in $\mathbb{R}^2$ and $\mathbb{R}^3$.

   Example 1. Geometries in $\mathbb{R}^2$ and $\mathbb{R}^3$:
   - In $\mathbb{R}^2$: $y = x$ is a straight line.
   - In $\mathbb{R}^2$: $x^2 + y^2 = 1$ is a circle.
   - In $\mathbb{R}^3$: $x^2 + y^2 + z^2 = 1$ is a sphere.
   - In $\mathbb{R}^3$: $z = 0$ is a plane.
   - In $\mathbb{R}^3$: $\begin{cases} z = 0 \\ x^2 + y^2 + z^2 = 1 \end{cases}$ is a circle.
   - In $\mathbb{R}^3$: $x^2 + y^2 + z^2 \leq 1$ is a solid ball.
   - In $\mathbb{R}^3$: $x^2 + y^2 + z^2 < 1$ is an interior of a solid ball.



In summary, we could generalize the geometry of different equations.

- In $\mathbb{R}^2$, $f(x, y) = 0$ is a curve.
- In $\mathbb{R}^3$, $F(x, y, z) = 0$ is a surface.
- In $\mathbb{R}^3$, irreducible $\begin{cases} F(x,y,z) = 0 \\ G(x,y,z) = 0 \end{cases}$ is a curve.
- In $\mathbb{R}^3$, $F(x, y, z) \leq 0$ or $F(x, y, z) \geq 0$ is a solid region.
- In $\mathbb{R}^3$, $F(x, y, z) > 0$ or $F(x, y, z) < 0$ is the interior of a solid region.

Then we wish to have a further discussion on solid regions and interiors of solid regions, which requires a general definition of open set and closed set.

> **Definition 1.** *Open set* is a set $S \subset \mathbb{R}^n$ such that every point $P \in S$ has a neighbour $B_r(P) := \{Q \in \mathbb{R}^n | |PQ| < r\}$ for some positive $r > 0$. *closed set* is set $S \subset \mathbb{R}^n$ such that its complement $S^c = \mathbb{R}^n - S$ is open.

Let's look at some examples.

Example 2. Examples of open sets and closed sets.

(a) $S = \{x \in \mathbb{R} | 0 < x < 1\}$ is an open set. $S = \{x \in \mathbb{R} | 0 \leq x \leq 1\}$ is a closed set. $S = \{x \in \mathbb{R} | 0 \leq x < 1\}$ is neither a closed set or open set.

(b) $S = \{(x, y) \in \mathbb{R}^2 | 0 < x < 1, 0 < y < 1\}$ is an open set. $S = \{(x, y) \in \mathbb{R}^2 | 0 \leq x \leq 1, 0 \leq y \leq 1\}$ is an closed set. $S = \{(x, y) \in \mathbb{R}^2 | 0 \leq x \leq 1, 0 < y < 1\}$ is neither a closed set or open set.

(c) $S = \{(x, y, z) \in \mathbb{R}^3 | x^2 + y^2 + z^2 = 4\}$ is a closed set. $S = \{(x, y, z) \in \mathbb{R}^3 | x^2 + y^2 + z^2 \leq 4\}$ is a closed set. $S = \{(x, y, z) \in \mathbb{R}^3 | x^2 + y^2 + z^2 < 4\}$ is an open set.

For those solids, we also have a definition for boundaries.

> **Definition 2.** *Boundary* of S is a set of all points P such that every neighbourhood of P contains both points in S and points in S^c.

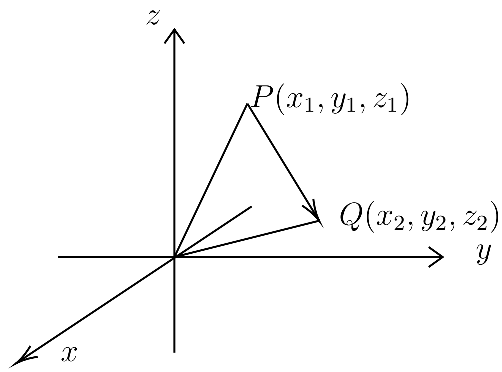
For example, the boundary of $S = \{(x, y, z) \in \mathbb{R}^3 | x^2 + y^2 + z^2 \leq 4\}$ is $\{(x, y, z) \in \mathbb{R}^3 | x^2 + y^2 + z^2 = 4\}$.

2. **Vector**

   (a) **Definitions**

   To explore the geometries in $\mathbb{R}^2$ and $\mathbb{R}^3$, we need to study the basic elements of $\mathbb{R}^2$ and $\mathbb{R}^3$ - vectors.

   > **Definition 3.** *Vector* is defined as a geometric object that has magnitude (or length) and direction. Mathematically, vector $\overrightarrow{PQ} = \langle x_2 - x_1, y_2 - y_1, z_2 - z_1 \rangle$ if $P = \langle x_1, y_1, z_1 \rangle$ and $Q = \langle x_2, y_2, z_2 \rangle$. An alternative way to represent vector is $\vec{v} = \langle a, b, c \rangle = a\vec{i} + b\vec{j} + c\vec{k}$.



$$\overrightarrow{PQ} = \langle\ x_2 - x_1, y_2 - y_1, z_2 - z_1\rangle$$

Note that if we only need the direction information of a vector we could normalize the vector. The *length* of the vector $\vec{v} = \langle a, b, c\rangle$ is $|\vec{v}| = \sqrt{a^2 + b^2 + c^2} \geq 0$. If $|\vec{v}| > 0$, then the *unit vector* (also called *direction vector*) of $\vec{v}$ is

$$\vec{u} = \frac{1}{|\vec{v}|}\vec{v} = \frac{\vec{v}}{|\vec{v}|}$$

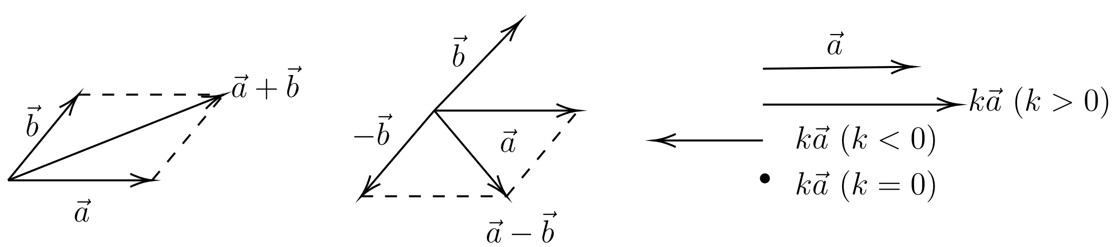
whose length is $|\vec{u}| = 1$. And the *unit angles* are $\cos\alpha = \dfrac{a}{|\vec{v}|}$, $\cos\beta = \dfrac{b}{|\vec{v}|}$ and $\cos\gamma = \dfrac{c}{|\vec{v}|}$. Notice $\cos^2\alpha + \cos^2\beta + \cos^2\gamma = 1$ and $\langle\cos\alpha, \cos\beta, \cos\gamma\rangle = \dfrac{\vec{v}}{|\vec{v}|} = \vec{u}$. If $|\vec{v}| = 0$, if is impossible to find the direction vector of $\vec{v}$ since $\vec{v} = \vec{0}$ is just a dot.

(b) **Operations** Let $\vec{a} = \langle x_1, y_1, z_1\rangle$ and $\vec{b} = \langle x_2, y_2, z_2\rangle$.

i. **Addition:** $\vec{a} + \vec{b} = \langle x_1 + x_2, y_1 + y_2, z_1 + z_2\rangle$

ii. **Subtraction:** $\vec{a} - \vec{b} = \langle x_1 - x_2, y_1 - y_2, z_1 - z_2\rangle$

iii. **Scalar Multiplication:** $k\vec{a} = \langle kx_1, ky_1, kz_1\rangle$



iv. **Dot Product**

A. **Definition:** $\vec{a}\cdot\vec{b} = x_1x_2 + y_1y_2 + z_1z_2$
Note: $\vec{a}\cdot\vec{b}$ could be abbreviated as $\vec{a}\vec{b}$.

B. **Linear Properties:**

- $\vec{v}\cdot\vec{v} = v_1^2 + v_2^2 + v_3^2 = |\vec{v}|^2$
- $\vec{v}\cdot\vec{w} = \vec{w}\cdot\vec{v}$

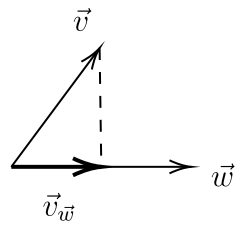
- $\vec{v} \cdot (c\vec{w}) = c(\vec{v} \cdot \vec{w})$
- $\vec{u} \cdot (\vec{v} + \vec{w}) = \vec{u} \cdot \vec{v} + \vec{u} \cdot \vec{w}$

C. **Corollary:** $\vec{a} \cdot \vec{b} = |\vec{a}||\vec{b}| \cos\theta$ where $0 \leq \theta \leq \pi$ is the angle between $\vec{a}$ and $\vec{b}$.
Proof. By cosine rule, $|\vec{a} - \vec{b}|^2 = |\vec{a}^2 + \vec{b}^2| - 2|\vec{a}|\vec{b}| \cos\theta$. By the linear property of dot product, $|\vec{a} - \vec{b}|^2 = (\vec{a} - \vec{b})^2 = |\vec{a}|^2 + |\vec{b}|^2 - 2\vec{a} \cdot \vec{b}$. Therefore, we have $\vec{a} \cdot \vec{b} = |\vec{a}||\vec{b}| \cos\theta$.

D. **Projection:** Let $\vec{v}, \vec{w} \in \mathbb{R}^n$ and $\vec{w} \neq \vec{0}$. We wish to decompose $\vec{v}$ such that

$$\vec{v} = \text{vector parallel to } \vec{w} \ + \ \text{vector perpendicular to } \vec{w}$$

We say the vector parallel to $\vec{w}$ is the projection of $\vec{v}$ on $\vec{w}$ and note it as $\vec{v}_{\vec{w}}$.



Then we have

$$\vec{v}_{\vec{w}} = s\frac{\vec{w}}{|\vec{w}|}$$

where s is the scalar projection $s = \dfrac{\vec{v} \cdot \vec{w}}{|\vec{w}|}$.
Proof. We know that $\vec{v}_{\vec{w}} \parallel \vec{w}$ since

$$\vec{v}_{\vec{w}} = \underbrace{\frac{s}{|\vec{w}|}}_{\text{scalar}} \vec{w}$$

and $(\vec{v} - \vec{v}_{\vec{w}}) \perp \vec{w}$ since

$$(\vec{v} - \vec{v}_{\vec{w}}) \cdot \vec{w} = \vec{v} \cdot \vec{w} - \frac{\vec{v} \cdot \vec{w}}{|\vec{w}|} \frac{\vec{w}}{|\vec{w}|} \cdot \vec{w} = \vec{v} \cdot \vec{w} - \vec{v} \cdot \vec{w} \frac{\cancel{\vec{w}^2}}{\cancel{|\vec{w}|^2}} = 0$$

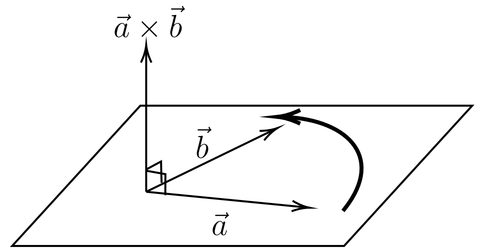
v. **Cross Product**

A. **Definition:** The cross product of $\vec{a}$ and $\vec{b}$ is defined as

$$\vec{a} \times \vec{b} = \begin{vmatrix} \vec{i} & \vec{j} & \vec{k} \\ x_1 & y_1 & z_1 \\ x_2 & y_2 & z_2 \end{vmatrix} = \vec{i} \begin{vmatrix} y_1 & z_1 \\ y_2 & z_2 \end{vmatrix} - \vec{j} \begin{vmatrix} x_1 & z_1 \\ x_2 & z_2 \end{vmatrix} + \vec{k} \begin{vmatrix} x_1 & y_1 \\ x_2 & y_2 \end{vmatrix}$$

Consequently, we know $\vec{a} \times \vec{b} = \langle y_1 z_2 - z_1 y_2, z_1 x_2 - x_1 z_2, x_1 y_2 - x_2 y_1 \rangle$.

B. **Length:** $|\vec{a} \times \vec{b}| = |\vec{a}||\vec{b}| \sin\theta$ where $0 \leq \theta \leq \pi$ is the angle between $\vec{a}$ and $\vec{b}$.
Direction: perpendicular to $\vec{a}$ and $\vec{b}$ and determined by right hand rule.

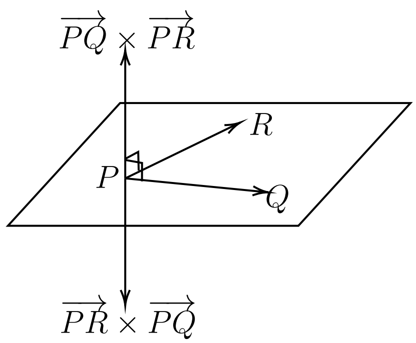


C. **Linear Properties:**

- $\vec{a} \times \vec{b} = -\vec{b} \times \vec{a}$
- $\vec{a} \times \vec{b} \perp \vec{a}$, $\vec{a} \times \vec{b} \perp \vec{b}$
- $\vec{a} \times \vec{b} = 2S_{\triangle}$ where the triangle $\triangle$ is spanned by $\vec{a}$ and $\vec{b}$.
- $\vec{a} \times \vec{b} = 0 \Leftrightarrow \vec{a} \parallel \vec{b}$

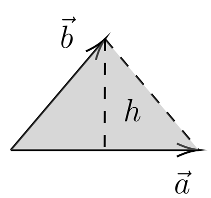
D. **Application**

- Find a normal vector $\vec{v}$ such that $\vec{v}$ is perpendicular to the plane spanned by $\overrightarrow{PQ}$ and $\overrightarrow{PR}$. This requires $\vec{n} \perp \overrightarrow{PQ}$ and $\vec{n} \perp \overrightarrow{PR}$.



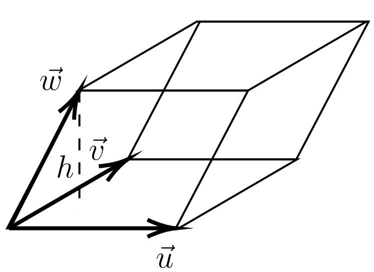
$$\vec{n} = \overrightarrow{PQ} \times \overrightarrow{PR}$$
or
$$\vec{n} = \overrightarrow{PR} \times \overrightarrow{PQ}$$

- Find the area of triangle spanned by $\vec{a}$ and $\vec{b}$.



$$S = \frac{1}{2}|\vec{a}|h = \frac{1}{2}|\vec{a}||\vec{b}|\sin\theta = \frac{1}{2}|\vec{a} \times \vec{b}|$$

- Find the volume of a parallelepiped spanned by $\vec{u}$, $\vec{v}$ and $\vec{w}$.



$$V = Sh = |\vec{u} \times \vec{v}| \left| \vec{w} \cdot \frac{\vec{u} \times \vec{v}}{|\vec{u} \times \vec{v}|} \right| = |\vec{w} \cdot (\vec{u} \times \vec{v})\,|$$

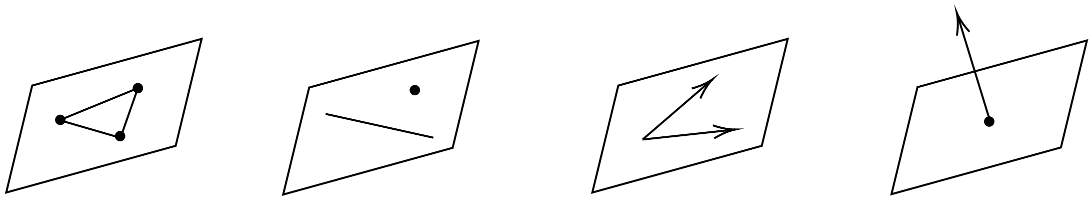
$$\Downarrow$$

$$V = |\vec{w} \cdot (\vec{u} \times \vec{v})\,| = \begin{vmatrix} x_1 & y_1 & z_1 \\ x_2 & y_2 & z_2 \\ x_3 & y_3 & z_3 \end{vmatrix}$$

3. **Curves and Surfaces in $\mathbb{R}^3$**

(a) **Plane**

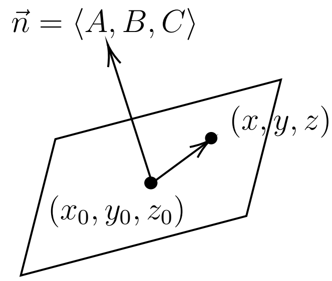
Basically, there are four ways to define a plane: using three points, using a point and line, using two vectors spanning and using one point and a normal vector.



After defining the plane, we could represent them in the following equations:

i. **Vector+Point**

If we know the point on the plane is (x_0, y_0, z_0) and the normal vector $\langle A, B, C\rangle$. Then for any point (x, y, z) on the plane



we know $\langle A, B, C\rangle \cdot \langle x - x_0, y - y_0, z - z_0\rangle = 0$. Therefore we have the plane is

$$A(x - x_0) + B(y - y_0) + C(z - z_0) = 0 \tag{1}$$

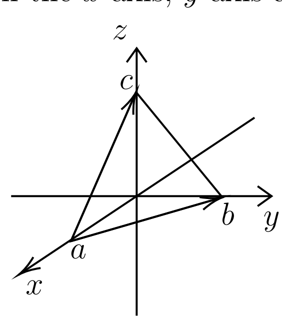
ii. **General**

By i, if we let $D \equiv -(Ax_0 + By_0 + Cz_0)$, we have the plane is

$$Ax + By + Cz + D = 0 \tag{2}$$

iii. **Intercept**

If the plane has intercepts with the x-axis, y-axis and z-axis which are a, b and c,



then we know two vectors spanning the plane are $\langle -a, 0, c\rangle$ and $\langle -a, b, 0\rangle$. Then the normal vector is $\vec{n} = \langle -a, 0, c\rangle \times \langle -a, b, 0\rangle = \langle -bc, -ac, -ab\rangle$. Then by i, we know the plane is

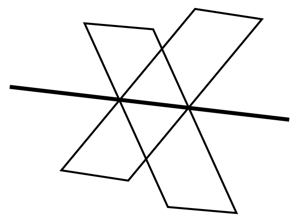
$$-bc(x - a) - acy - abz = 0$$

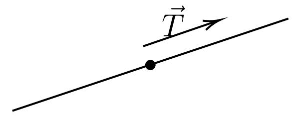
which could be simplified as

$$\frac{x}{a}+\frac{y}{b}+\frac{z}{c}=1 \tag{3}$$

(b) **Line**

Basically, there are two ways to define a straight line: using the intersection of two planes and using a point on the line and a direction vector of the line.





After defining the line, we could represent them in the following equations.

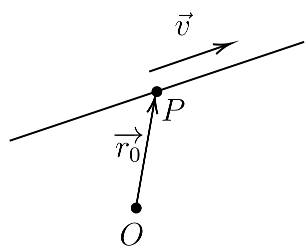
i. **General**

The general equation is

$$\begin{cases} A_1x+B_1y+C_1z+D_1=0 \\ A_2x+B_2y+C_2z+D_2=0 \end{cases} \tag{1}$$

Then we could find a point $(0,y',z')$ as a solution of the system of equations and a normal vector $\vec{n}=\langle A_1,B_1,C_1\rangle\times\langle A_1,B_1,C_1\rangle$.

ii. **Parametric equation** If we know the point $\vec{r_0}=\langle x_0,y_0,z_0\rangle$ on the line and the direction vector $\vec{v}=\langle a,b,c\rangle$,



then we know the equation would be

$$\vec{r}=\vec{r_0}+t\vec{v}=\langle x_0+ta,y_0+tb,z_0+tc\rangle \tag{2}$$

Alternatively, it is

$$\begin{cases} x=x_0+ta \\ y=y_0+tb \\ z=z_0+tc \end{cases}$$

If $a,b,c\neq 0$, we could eliminate t,

$$\mathrlap{\overbrace{\phantom{\frac{x-x_0}{a}=\frac{y-y_0}{b}}}^{\text{plane}}}\frac{x-x_0}{a}=\underbrace{\frac{y-y_0}{b}=\frac{z-z_0}{c}}_{\text{plane}} \tag{2.1}$$

If $a = 0$ and $b, c \neq 0$, we know

$$\begin{cases} x = x_0 \\ y = y_0 + tb \\ z = z_0 + tc \end{cases} \tag{2.2}$$

If $a, b = 0$ and $c \neq 0$, we know

$$\begin{cases} x = x_0 \\ y = y_0 \end{cases} \tag{2.3}$$

(c) **Quadric Surface**

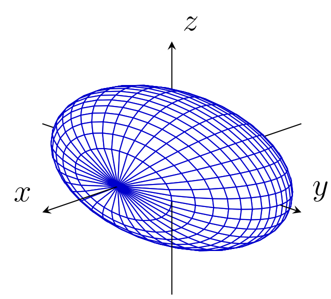
The general form of a quadric surface is

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

We usually use *3D surfaces*, *cross-sections* and *level curves*

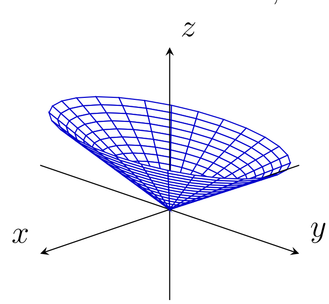
i. **Sphere:** $x^2 + y^2 + z^2 = r^2$

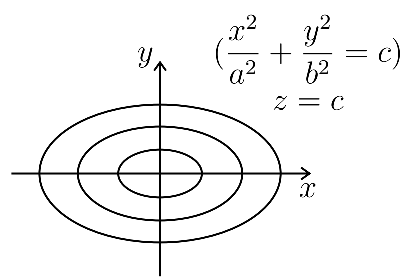
ii. **Ellipsoid:** $\dfrac{x^2}{a^2} + \dfrac{y^2}{b^2} + \dfrac{z^2}{c^2} = 1$



iii. **Cone:** $z^2 = \dfrac{x^2}{a^2} + \dfrac{y^2}{b^2}$

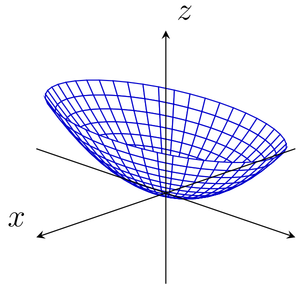
If $a = b$, we have a circular cone; if $a \neq b$, we have a elliptical cone.

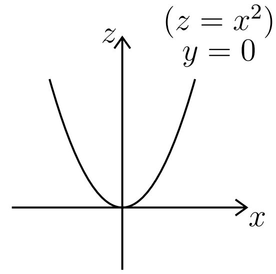


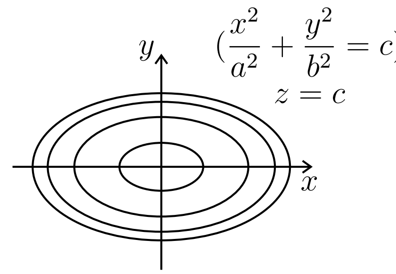


iv. **Paraboloid:** $z = \dfrac{x^2}{a^2} + \dfrac{y^2}{b^2}$

If $a = b$, we have a circular paraboloid; if $a \neq b$, we have a elliptical paraboloid.

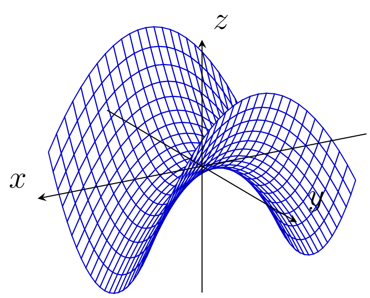


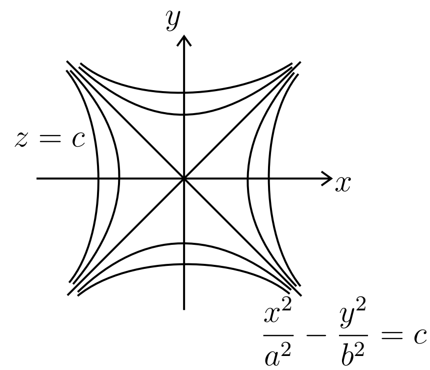


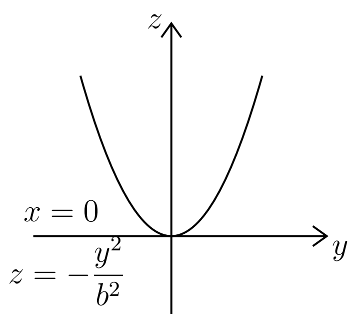


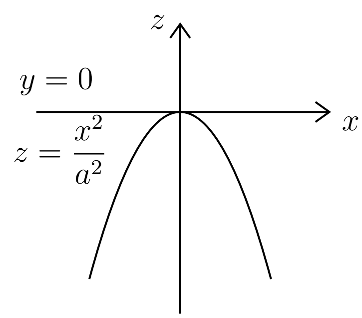
v. **hyperbolic paraboloid:** $z = \dfrac{x^2}{a^2} - \dfrac{y^2}{b^2}$

This means the cross-sections $z = c$ $(c = 0)$ such that $\dfrac{x^2}{a^2} - \dfrac{y^2}{b^2} = c$ are hyperbolas.



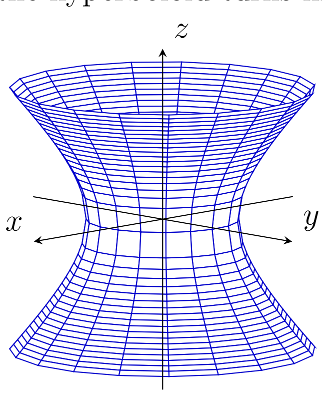




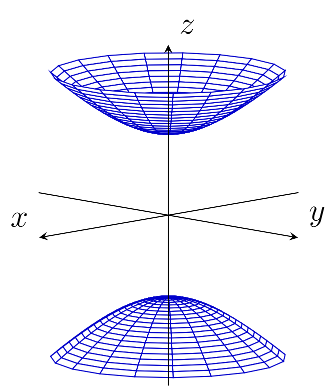


vi. **Hyperboloid:** $x^2 + y^2 - z^2 = c$ $(c \neq 0)$

Note: if $c = 0$, the hyperboloid turns into a cone.



$c = 1$: $z^2 = x^2 + y^2 - 1$



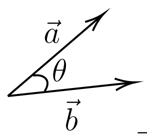
$c = -1$: $z^2 = x^2 + y^2 + 1$

Notice that larger $|c|$ will make the surfaces further away from the center.

4. **Measurement by Vector Calculation**

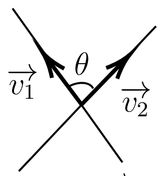
(a) **Angle**

There are four types of angles: the angle between two vectors, the angle between two lines, the angle between two planes and the angle between a line and a plane.



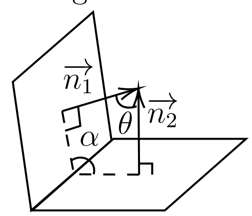
$$\cos\theta = \frac{\vec{a}\cdot\vec{b}}{|\vec{a}|\cdot|\vec{b}|}$$

$$\theta \in [0, \pi]$$



$$\cos\theta = \frac{\overrightarrow{v_1}\cdot\overrightarrow{v_2}}{|\overrightarrow{v_1}|\cdot|\overrightarrow{v_2}|}$$

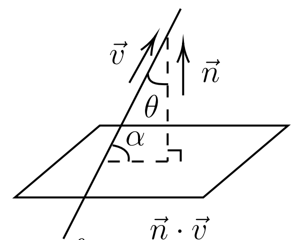
$$\theta \in \left[0, \frac{\pi}{2}\right]$$



$$\cos\theta = \frac{\overrightarrow{n_1}\cdot\overrightarrow{n_2}}{|\overrightarrow{n_1}|\cdot|\overrightarrow{n_2}|}$$

$$\alpha = \theta \text{ or } \pi - \theta$$

$$\alpha \in \left[0, \frac{\pi}{2}\right]$$



$$\cos\theta = \frac{\vec{n}\cdot\vec{v}}{|\vec{n}|\cdot|\vec{v}|}$$

$$\alpha = \frac{\pi}{2} - \theta \text{ or } \theta - \frac{\pi}{2}$$

$$\alpha \in \left[0, \frac{\pi}{2}\right]$$

(b) **Distance**

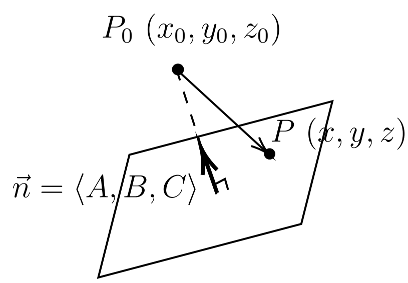
i. Distance between two points $P_1\,(x_1, y_1, z_1)$ and $P_2\,(x_2, y_2, z_2)$ is

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

ii. Distance between the point $P_0(x_0, y_0, z_0)$ and the plane $Ax + By + Cz + D = 0$ is

$$d = \frac{|Ax_0 + By_0 + Cz_0 + D|}{\sqrt{A^2 + B^2 + C^2}}$$

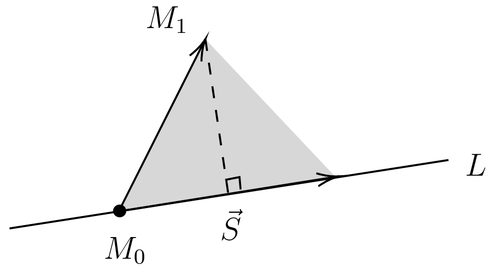
Proof. Let $P\,(x, y, z)$ on the plane such that $Ax + By + Cz + D = 0$. Then we have the distance is $d = \left|\overrightarrow{P_0P}\cdot\frac{\vec{n}}{|\vec{n}|}\right| = \left|\langle x - x_0, y - y_0, z - z_0\rangle \cdot \frac{\langle A, B, C\rangle}{\sqrt{A^2 + B^2 + C^2}}|\right| = \frac{|A(x - x_0) + B(y - y_0) + C(z - z_0) + D|}{\sqrt{A^2 + B^2 + C^2}} = \frac{|Ax_0 + By_0 + Cz_0 + D|}{\sqrt{A^2 + B^2 + C^2}}$.



iii. Distance between the point M_1 and the line defined by point M and direction vector $\overrightarrow{S}$ is

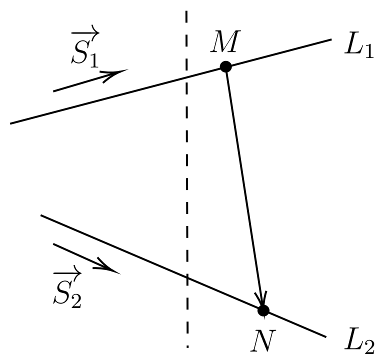
$$d = \frac{|\overrightarrow{M_0M_1} \times \overrightarrow{S}|}{|\overrightarrow{S}|}$$

Proof. Consider the area of the triangle, we know $\frac{1}{2}|\overrightarrow{M_0M_1} \times \overrightarrow{S}| = \frac{1}{2}|\overrightarrow{S}|d$.



iv. The distance between two lines defined by points M, N and direction vectors $\overrightarrow{S_1}$, $\overrightarrow{S_2}$ is

$$d = \frac{|(\overrightarrow{S_1} \times \overrightarrow{S_2}) \cdot \overrightarrow{MN}|}{|\overrightarrow{S_1} \times \overrightarrow{S_2}|}$$



Multivariable Functions and Limits

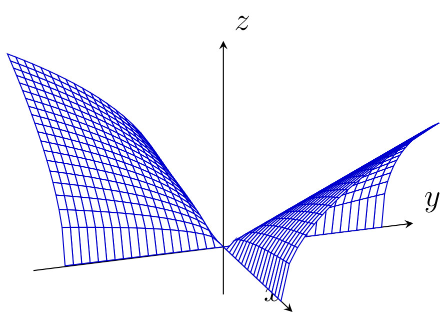
1. **Basic Concepts**
   Previously, we discussed real-valued functions of a single variable $y = f(x)$. From now on, we are going to start the discussion of real-valued functions of multiple variables.

> **Definition 4.** *Real-valued multivariable function* is a function
>
> $$\begin{aligned} f: \quad \mathbb{R}^n &\longrightarrow \mathbb{R} \\ (x_1, \ldots, x_n) &\longmapsto f(x_1, \ldots, x_n) \end{aligned}$$
>
> The *domain* of f is $\mathscr{D}(f) = \{(x_1, \ldots, x_n) \in \mathbb{R} | \text{defined under } f\}$. The *range* of f is $\{z \in \mathbb{R} | z = f(x_1, \ldots, x_n), (x_1, \ldots, x_n) \in \mathscr{D}(f)\}$. The *graph* of f is $\{(x_1, \ldots, x_n, x_{n+1}) \in \mathbb{R}^{n+1} | x_{n+1} = f(x_1, \ldots, x_n), (x_1, \ldots, x_n) \in \mathscr{D}(f)\}$. *Level curves* are graphs $\{(x_1, \ldots, x_n) \in \mathscr{D}(f) | f(x_1, \ldots, x_n) = c\}$.

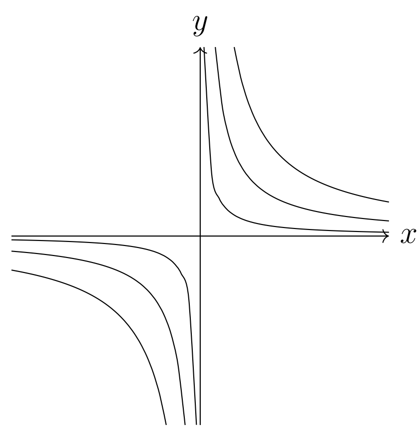
It's not hard to find that if $n = 2$, the level curves are curves in 2D space; if $n = 3$, the level curves are surfaces in 3D spaces.

Let's look at an example.

Example 3. Consider $f(x, y) = \sqrt{xy}$ where domain is $\mathscr{D}(f) = \{(x, y) \in \mathbb{R} | xy > 0\}$. The range is $(-\infty, \infty)$. The graph is



And the level curves are $xy = c^2$. The following level curves are $c = 1$, $c = 2$, $c = 3$.



2. **Limit**

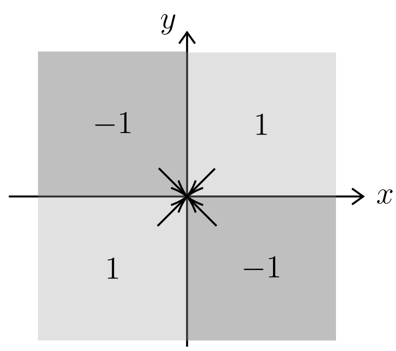
(a) **Definition**

We want to define the limit of multivariable functions and from now on the discussion will be focused on functions of two variables as an example. The key point of limit is "approaching" - getting closer. This indicates that the distance getting smaller which is characterized as $0 < |x-a| < \delta$ and $|f(a) - L| < \varepsilon$. $|f(a) - L| < \varepsilon$ is still valid but we have to change the distance $|x-a|$ to $|\vec{x} - \vec{a}|$ such as $|(x, y) - (a, b)| = \sqrt{(x-a)^2 + (y-b)^2}$.

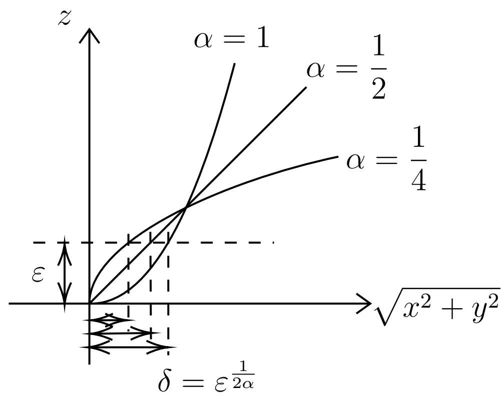
> **Definition 5.** We say $\lim\limits_{(x,y)\to(a,b)} f(x, y) = L$ if:
>
> - every neighbourhood of (a, b) contains points of $\mathscr{D}(f)$ different from (a, b),
> - $\forall\varepsilon > 0$, $\exists\delta > 0$, when $0 < \sqrt{(x-a)^2 + (y-b)^2} < \delta$, $|f(x, y) - L| < \varepsilon$.

Note that the point (a, b) is not actually required to be in the domain of f. Also the definition only requires the decrease of distance but has nothing to do with the direction of (x, y) approaching (a, b). Therefore if the limit exists, it should be the same for all the direction (x, y) approaches (a, b). Let's look at some examples.

Example 4. Let $f(x, y) = \begin{cases} -1 & xy > 0 \\ 1 & xy < 0 \end{cases}$. We want to check whether $\lim\limits_{(x,y)\to(0,0)} f(x, y)$ exits. Though $(0, 0)$ is not in the domain of $f(x, y)$, it would be the evidence that the limit does not exist. We could check the limit of (x, y) approaching $(0, 0)$ in different direction. We have $\lim\limits_{\substack{(x,y)\to(0,0)\\ x=y}} f(x, y) = 1$ and $\lim\limits_{\substack{(x,y)\to(0,0)\\ x=-y}} f(x, y) = -1$. Therefore, the limit does not exist.



Example 5. Let $f(x, y) = (x^2 + y^2)^\alpha$ $(\alpha > 0)$. We want to show that $\lim\limits_{(x,y)\to(0,0)} (x^2 + y^2)^\alpha = 0$. Let $\varepsilon > 0$, we need to find δ such that if $\sqrt{x^2 + y^2} < \delta$, then $|f(x, y)^\alpha - 0| < \varepsilon$. This implies $|(x^2 + y^2)^\alpha| < \varepsilon$ and then $\sqrt{x^2 + y^2} < \varepsilon^{\frac{1}{2\alpha}} = \delta$. Therefore, the required $\delta = \varepsilon^{\frac{1}{2\alpha}}$.



Example 6. Let $f(x,y) = \dfrac{xy}{x^2+y^2}$ $((x,y) \neq (0,0))$. We want to show that $\lim\limits_{(x,y)\to(0,0)} f(x,y)$ does not exist. So we are going to try different directions. Along $x = 0$, $\lim\limits_{\substack{(x,y)\to(0,0)\\ x=0}} f(x,y) = 0$. Along $y = 0$, $\lim\limits_{\substack{(x,y)\to(0,0)\\ y=0}} f(x,y) = 0$. However, along $y = x$, $\lim\limits_{\substack{(x,y)\to(0,0)\\ y=x}} f(x,y) = \lim\limits_{x\to 0} \dfrac{x^2}{2x^2} = \dfrac{1}{2} \neq 0$. Therefore, the limit does not exist.

Example 7. Let $f(x,y) = \dfrac{x^4y^4}{(x^2+y^4)^3}$ $((x,y) \neq (0,0))$. We want to show that $\lim\limits_{(x,y)\to(0,0)} f(x,y)$ does not exist. So we are going to try different directions. Along $x = 0$, $\lim\limits_{\substack{(x,y)\to(0,0)\\ x=0}} f(x,y) = 0$. Along $y = 0$, $\lim\limits_{\substack{(x,y)\to(0,0)\\ y=0}} f(x,y) = 0$. Along $y = x$, $\lim\limits_{\substack{(x,y)\to(0,0)\\ y=x}} f(x,y) = \lim\limits_{x\to 0} \dfrac{x^8}{(x^2+x^4)^3} = \lim\limits_{x\to 0} \dfrac{x^2}{(x^2+1)^3} = 0$. Along $y = mx$, $\lim\limits_{\substack{(x,y)\to(0,0)\\ y=mx}} f(x,y) = \lim\limits_{x\to 0} \dfrac{x^8(m^4+1)}{(x^2+m^4x^4)^3} = \lim\limits_{x\to 0} \dfrac{x^2(m^4+1)}{(m^4x^2+1)^3} = 0$. However, along $y = x^\alpha$, $\lim\limits_{\substack{(x,y)\to(0,0)\\ y=x^\alpha}} f(x,y) = \lim\limits_{x\to 0} \dfrac{x^{4+4\alpha}}{(x^2+x^{4\alpha})^3}$. If $\alpha = \dfrac{1}{2}$, $\lim\limits_{x\to 0} \dfrac{x^{4+4\alpha}}{(x^2+x^{4\alpha})^3} = \dfrac{1}{8}$. Therefore, the limit does not exist.

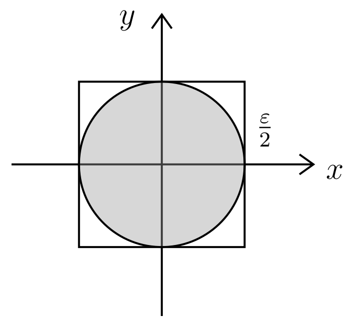
Example 8. Let $f(x,y) = \dfrac{x^3+y^3}{x^2+y^2}$ $((x,y) \neq (0,0))$. We want to show that $\lim\limits_{(x,y)\to(0,0)} f(x,y) = 0$. Let $\varepsilon > 0$, we want to find $\delta > 0$ such that $\left|\dfrac{x^3+y^3}{x^2+y^2}\right| < \varepsilon$ if $\sqrt{x^2+y^2} < \delta$. This suffices if

$$\left|\frac{x^3}{x^2+y^2}\right| < \frac{\varepsilon}{2}, \ \left|\frac{y^3}{x^2+y^2}\right| < \frac{\varepsilon}{2}$$

Since $\left|\dfrac{x^3}{x^2+y^2}\right| \le \left|\dfrac{x^3}{x^2}\right| = |x|$ and $\left|\dfrac{y^3}{x^2+y^2}\right| \le \left|\dfrac{y^3}{y^2}\right| = |y|$, this then suffices if

$$|x| < \frac{\varepsilon}{2},\ |y| < \frac{\varepsilon}{2}$$

Since we know $\{(x,y) \in \mathbb{R}^2 | |x| < \dfrac{\varepsilon}{2},\ |y| < \dfrac{\varepsilon}{2}\} \subset \{(x,y) \in \mathbb{R}^2 | \sqrt{x^2+y^2} < \dfrac{\varepsilon}{2}\}$,



we finally take $\delta = \dfrac{\varepsilon}{2}$.

(b) **Rules of limits**
Let f, g be the function of (x,y). Every neighbourhood of (a,b) contains points of $\mathscr{D}(f) \cap \mathscr{D}(g)$. And $\lim\limits_{(x,y)\to(a,b)} f(x,y) = A$ and $\lim\limits_{(x,y)\to(a,b)} g(x,y) = B$. Then we have

i. $\lim\limits_{(x,y)\to(a,b)} [f(x,y) \pm g(x,y)] = A \pm B$

Proof. By definition, we know $\forall \varepsilon_1 > 0$, $\exists D_1$ such that when $0 < \sqrt{(x-a)^2+(y-b)^2} < \delta_1$, $0 < |f - A| < \varepsilon_1$ and $\forall \varepsilon_2 > 0$, $\exists D_2$ such that when $0 < \sqrt{(x-a)^2+(y-b)^2} < \delta_2$, $0 < |g - B| < \varepsilon_2$. Let $\delta = \min\{\delta_1, \delta_2\}$ and $\varepsilon_1 = \varepsilon_2 = \dfrac{\varepsilon}{2}$. Then when $0 < \sqrt{(x-a)^2+(y-b)^2} < \delta$, we have

$$0 < |(f+g) - (A+B)| \le |f - A| + |g - B| < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon$$

by trigonometry inequality. Therefore $\lim_{(x,y)\to(a,b)}[f+g] = A+B$. And then when $0 < \sqrt{(x-a)^2+(y-b)^2} < \delta$, we have

$$0 < |(f-g) - (A-B)| \le |f - A| + |g - B| < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon$$

by trigonometry inequality. Therefore $\lim_{(x,y)\to(a,b)}[f-g] = A-B$.

ii. $\lim\limits_{(x,y)\to(a,b)} f(x,y)g(x,y) = AB$ *Proof.* Let $\varepsilon > 0$. By definition of the limits of g and f, $\exists D_1$ such that when $0 < \sqrt{(x-a)^2+(y-b)^2} < \delta_1$, $|f - A| < \sqrt{\dfrac{\varepsilon}{3}}$ (1). $\exists D_2$ such that when $0 < \sqrt{(x-a)^2+(y-b)^2} < \delta_2$, $|f - A| < \dfrac{\varepsilon}{3B}$ (2). $\exists D_3$ such that when $0 < \sqrt{(x-a)^2+(y-b)^2} < \delta_3$, $|g - B| < \sqrt{\dfrac{\varepsilon}{3}}$ (3). $\exists D_4$ such that when

$0 < \sqrt{(x-a)^2+(y-b)^2} < \delta_4$, $|g-B| < \dfrac{\varepsilon}{3A}$ (4). Take $\delta = \min\{\delta_1, \delta_2, \delta_3, \delta_4\}$, when $0 < \sqrt{(x-a)^2+(y-b)^2} < \delta$, we would have (1)(2)(3)(4) all are satisfied. Since

$$(f-A)(g-B) = fg - Ag - Bf + AB$$

which comes out to be

$$\begin{aligned}
|fg - AB| &= |(f-A)(g-B) + Ag + Bf - 2AB| \\
&= |(f-A)(g-B) + B(f-A) + A(g-B)| \\
&< |f-A||g-B| + B|f-A| + A|g-B| \\
&= \sqrt{\frac{\varepsilon}{3}}\sqrt{\frac{\varepsilon}{3}} + B\frac{\varepsilon}{3B} + A\frac{\varepsilon}{3A} \\
&= \varepsilon
\end{aligned}$$

So $\forall \varepsilon > 0$, we find the $\delta > 0$ such that when $0 < \sqrt{(x-a)^2+(y-b)^2} < \delta$, $|fg - AB| < \varepsilon$. By definition $\lim\limits_{(x,y)\to(a,b)} fg = AB$.

iii. $\lim\limits_{(x,y)\to(a,b)} \dfrac{f(x,y)}{g(x,y)} = \dfrac{A}{B}$

Proof. First we are going to prove a lemma $\lim\limits_{(x,y)\to(a,b)} \dfrac{1}{g} = \dfrac{1}{B}$. Let $\varepsilon > 0$. By definition, we know that there is a δ_1 such that when $0 < \sqrt{(x-a)^2+(y-b)^2} < \delta_1$, $|g-B| < \dfrac{|B|}{2}$. Therefore, $|\,|B| - |g|\,| \le |g-B| < \dfrac{|B|}{2}$ by triangle inequality which implies $|g| > \dfrac{|B|}{2}$ and then implies $0 < \dfrac{1}{|g|} < \dfrac{2}{|B|}$. Also, by definition there is δ_2 such that when $0 < \sqrt{(x-a)^2+(y-b)^2} < \delta_2$, we have $0 < |g-B| < \dfrac{|B|^2}{2}\varepsilon$. Let $\delta = \min\{\delta_1, \delta_2\}$. Then when $0 < \sqrt{(x-a)^2+(y-b)^2} < \delta$, we have

$$\begin{aligned}
|\frac{1}{g} - \frac{1}{B}| = \frac{|g-B|}{|gB|} &= \frac{1}{|g|}\frac{1}{|B|}|g-B| \\
&< \frac{2}{|B|}\frac{1}{|B|}\frac{|B|^2}{2}\varepsilon = \varepsilon
\end{aligned}$$

and also $|\dfrac{1}{g} - \dfrac{1}{B}| = \dfrac{1}{|g|}\dfrac{1}{|B|}|g-B| > 0$. Therefore, we have $\lim \dfrac{1}{g} = \dfrac{1}{B}$. By the calculate property ii, we have $\lim\limits_{(x,y)\to(a,b)} \dfrac{f}{g} = \dfrac{A}{B}$.

(c) **Continuity**

The definition of continuity of multivariable functions is similar to the definition of continuity of single variable function.

Definition 6. f is continious at (a, b) if:

- $(a, b) \in \mathscr{D}(f)$,
- $\lim\limits_{(x,y)\to(a,b)} f(x, y) = f(a, b)$ exists

Usually, the following functions are continuous:

- If for all $i \in \{1, \ldots, k\}$ when we remain $x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_k$ constant such that $g_i(x_i) = f(x_1^c, \ldots, x_{i-1}^c, x_i, x_{i+1}^c, \ldots, x_k^c)$ and g_i is continuous on I_i, then f is continuous on $\{(x_1, \ldots, x_k)|x_j \in I_j\}$,
- polynomials in $x_1, \ldots, x_k$ are continuous on $\mathbb{R}^k$,
- rational functions $\dfrac{P(x_1, \ldots, x_m)}{Q(x_1, \ldots, x_n)}$ where P and Q are polynomials are continuous when $Q(x_1, \ldots, x_n) \neq 0$ (alternatively $\{(x_1, \ldots, x_k) \in \mathbb{R}^k|Q(x_1, \ldots, x_n) \neq 0\}$ where $k = \max\{m.n\}$),
- composition of continuous functions.

Consequently, extreme value theorem and intermediate value theorem could also be applied onto the continuous multivariable functions.

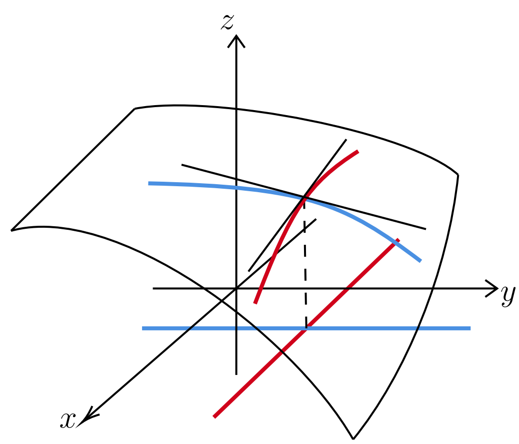
Differentiation of Multivariable Functions

1. **Partial Derivatives**

   (a) **Definition**

   We want to define the derivative of multivariable functions. Recall that the key point of derivative of single variable is find the how quick f is changing corresponding to f - the division of Δf by Δx. However, there are multiple variables to vary, how can we determine the change of f correspondingly and define the derivative correspondingly?

   An easy way to go is make only one variable changing and other variables constant. For functions of two variables, we could make y constant and find how quickly f is changing with respect to x or we could make x constant and find how quickly f is changing with respect to y.



   This means we just take the derivative of $g_1(x) = f(x, y_c)$ with respect to x and $g_2(y) = f(x_c, y)$ with respect to y. However, this kind of derivatives only tell us part of the stories. We call this kind of derivatives *partial derivatives*.

   Definition 7. The *first partial derivatives* of the function $f(x, y)$ with respect to the variables x and y are the functions $\dfrac{\partial f}{\partial x}(x, y) = \lim_{h \to 0} \dfrac{f(x+h, y) - f(x, y)}{h}$ and $\dfrac{\partial f}{\partial y}(x, y) = \lim_{h \to 0} \dfrac{f(x, y+h) - f(x, y)}{h}$ provided the limits exist.

   In general, there are multiple ways to denote partial derivative. For the first partial derivative of f with respect to x_i we have

$$\frac{\partial f}{\partial x} = \frac{\partial z}{\partial x} = D_{x_i} f = D_i f = f_i$$

   given by

$$\lim_{h \to 0} \frac{f(x_1, \ldots, x_i + h, \ldots, x_n) - f(x_1, \ldots, x_n)}{h}$$

And the value of the partial derivative at $(a_1, a_2, \ldots, a_n)$ is denoted as

$$\left.\frac{\partial f}{\partial x}\right|_{(a_1,\ldots,a_n)} = \left.\frac{\partial z}{\partial x}\right|_{(a_1,\ldots,a_n)} = D_{x_i} f(a_1, \ldots, a_n) = D_i f(a_1, \ldots, a_n) = f_i(a_1, \ldots, a_n)$$

Example 9. Consider $f(x, y) = e^x \sin(y)$. We have $\dfrac{\partial f}{\partial x}(x, y) = e^x \sin(y)$ and $\dfrac{\partial f}{\partial y}(x, y) = e^x \cos(y)$.

(b) **Higher and Mixed Partial Derivatives**

Partial derivatives of second and higher orders are calculated by taking partial derivatives of already calculated partial derivatives. For the second partial derivatives of a function with two variables, we have two pure derivatives

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

and

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

Also we have two mixed derivatives

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

and

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

Notice the order in the notations. In

$$\frac{\partial^n f}{\underleftarrow{\partial x_k \cdots \partial x_i \cdots \partial x_j}}$$

the order is from right to left. In

$$f_{\underrightarrow{\partial x_j \cdots \partial x_i \cdots \partial x_k}}, f_{\underrightarrow{j \cdots i \cdots k}}$$

the order is from left to right.

Example 10. Consider $f(x, y, z) = ze^{xy}$. We have $f_x = zye^{xy}$ and $f_z = e^{xy}$ Then we have $f_{xz} = ye^{xy}$ and $f_{zx} = e^{xy}$.

You may notice that the two mixed partial derivatives are the same. This is not always true but we have a theorem to clarify it.

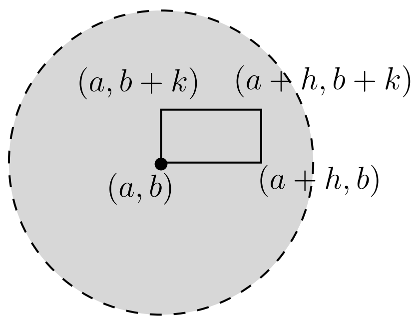
Theorem 1. (Equality of Mixed Partials) Let $f(x_11, \ldots, x_n)$ be the function domain on $\mathscr{D}$, if:

- all partial derivatives of order smaller than n are continuous of neighbour of P,
- two partials of order n with the same differentiation are continuous at P,

then these two partials are equal at P.

Let's consider the special case of this theorem: the function of two variables.

Theorem*: Let function $f(x, y)$ domain on $\mathscr{D}$. If all first and second partial derivatives are continuous in a neighbourhood at P, then $f_{xy}(P) = f_{yx}(P)$.
Proof. Let $P = (a, b)$ and the neighbourhood contains a rectangle with vertex (a, b), $(a, b + k)$, $(a + h, b)$ and $(a + h, b + k)$.



Let $Q = f(a + h, b + k) - f(a, b + k) - f(a + h, b) + f(a, b)$. There are ways to group Q. We could have

$$Q = (\underbrace{f(a + h, b + k) - f(a, b + k)}_{v(b+k)}) - (\underbrace{f(a + h, b) - f(a, b)}_{v(b)})$$

and

$$Q = (\underbrace{f(a + h, b + k) - f(a + h, b)}_{u(a+h)}) - (\underbrace{f(a, b + k) - f(a, b)}_{u(a)})$$

where

$$v(y) = f(a + h, y) - f(a, y), u(x) = f(x, b + k) - f(x, b)$$

Then by mean value theorem, we know there exists $0 < \theta_1 < 1$ $(b < b + k\theta_1 < b + k)$ such that

$$Q = v(b + k) - v(b) = kv'(b + k\theta_1) = k[f_y(a + h, b + k\theta_1) - f_y(a, b + k\theta_1)]$$

Then by mean value theorem again, we know there exists $0 < \theta_2 < 1$ $(a < a + h\theta_2 < a + h)$ such that

$$k[f_y(a + h, b + k\theta_1) - f_y(a, b + k\theta_1)] = khf_{yx}(a + h\theta_2, b + k\theta_1)$$

Similarly, by mean value theorem, there exists $0 < \theta_3, \theta_4 < 1$ such that

$$Q = khf_{xy}(a + h\theta_4, b + k\theta_3)$$

Therefore, by equating Q, we have

$$khf_{yx}(a + h\theta_2, b + k\theta_1) = khf_{xy}(a + h\theta_4, b + k\theta_3)$$

Cancelling kh, we have

$$f_{yx}(a + h\theta_2, b + k\theta_1) = f_{xy}(a + h\theta_4, b + k\theta_3)$$

We take the limit $h \to 0$ and $k \to 0$ and then we have

$$f_{yx}(a, b) = f_{xy}(a, b)$$

2. **Linearization and Differentiation**

   (a) **Review of Linearization of Single Variable Function**

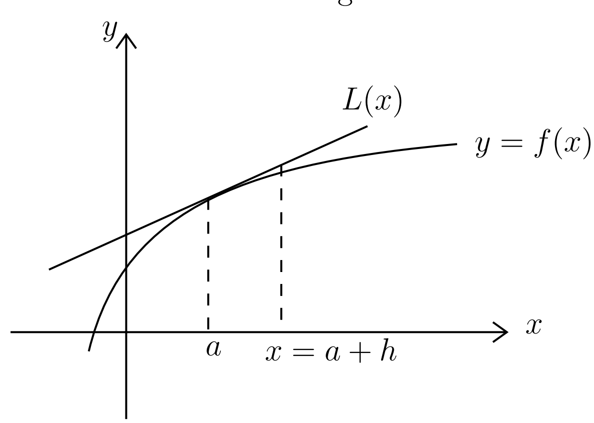
   Recall that the linearization of $f(x)$ at $x = a$ is basically to use a tangent line to approximate

$$L(x) = f(a) + f'(a)(x - a)$$

   We could find that being able to find the non-vertical tangent line is equivalent to being differentiable at a and is equivalent to the existence of $f'(a)$. Also they could imply that $f(x)$ is continuous.

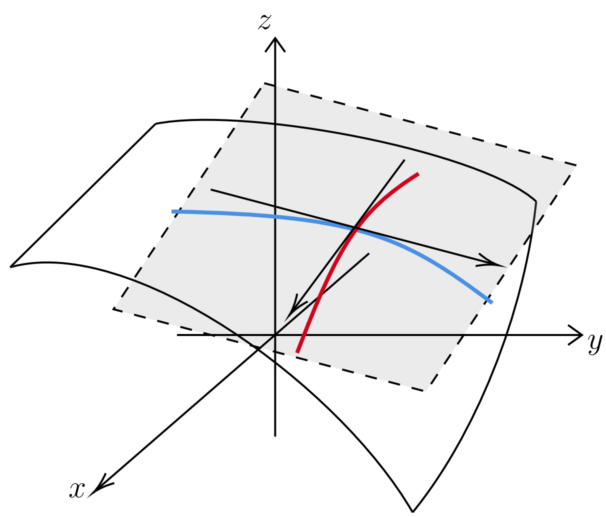
$$\begin{tikzcd}
\exists f'(a) \arrow[rr, Leftrightarrow] \arrow[dr, Leftrightarrow] & & \begin{array}{c}\text{graph } y = f(x) \text{ has a vertical} \\ \text{tangent line at } x = a\end{array} \arrow[dl, Leftrightarrow] \\
& \begin{array}{c} f(x) \text{ is differentiable at} \\ x = a \end{array} \arrow[d, Rightarrow] & \\
& \begin{array}{c} f(x) \text{ is continuous at} \\ x = a \end{array} &
\end{tikzcd}$$

   Let's consider the differentiation and the tangent line in a more careful way.



By definition, we have $f'(a) = \lim_{x\to a} \dfrac{f(x)-f(a)}{x-a}$. This is equivalent to, by rearranging the equation, $0 = \lim_{x\to a}\left(\dfrac{f(x)-f(a)}{x-a} - f'(a)\right) = \lim_{x\to a} \dfrac{f(x) - (\overbrace{f(a)+f'(a)(x-a)}^{L(x)})}{x-a} = \lim_{x\to a}\dfrac{f(x)-L(x)}{x-a}$. This also be interpreted as the error $E(x) = f(x) - L(x)$ is much smaller than the distance $\Delta x = x - a$ when $x \to a$. This is the requirement of differentiablity and the existence of non-vertical tangent line.

(b) **Linearization of Functions of Two Variables and Tangent Plane**
To linearize $z = f(x,y)$ at (a,b), we want to find the tangent plane instead of tangent line. To find a plane, we need to find a point on the plane and find the direction of the plane. It not hard to see that point $(a,b,f(a,b))$ is for sure on the plane. The direction normal vector could be found by two vector spanning the plane. These two vector could be found by partial derivative along x and along y.



The vector along x is $\vec{v}_1 = \langle 1, 0, \dfrac{\partial f}{\partial x}\rangle$ and the vector along y is $\vec{v}_2 = \langle 0, 1, \dfrac{\partial f}{\partial y}\rangle$. Then the normal vector is

$$\vec{n} = \vec{v}_1 \times \vec{v}_2 = \langle -\frac{\partial f}{\partial x}, -\frac{\partial f}{\partial y}, 1\rangle$$

Therefore the equation of the plane is

$$-\left.\frac{\partial f}{\partial x}\right|_{(a,b)}(x-a) - \left.\frac{\partial f}{\partial x}\right|_{(a,b)}(y-b) + (z - f(a,b)) = 0$$

which could be simplified as

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

This is the tangent plane of $f(x,y)$ at (a,b) if exists. Then the linearization of $f(x,y)$ at (a,b) is

$$L(x,y) = \left.\frac{\partial f}{\partial x}\right|_{(a,b)} (x-a) + \left.\frac{\partial f}{\partial x}\right|_{(a,b)} (y-b) + f(a,b)$$

However, the existence of linear approximation function does not imply the existence of the actual tangent plane.

(c) **Differentiation of Functions of Two Variables**

Borrowing the idea of differentiablity on single variable function, we define the differentiablity of functions of two variables by taking the limit of $\dfrac{f(x,y)-L(x,y)}{|(x,y)-(a,b)|}$.

Definition 8. We say that the function $f(x,y)$ is *differentiable* at the point (a,b) if

$$\lim_{(x,y)\to(a,b)} \frac{f(x,y)-L(x,y)}{\sqrt{(x-a)^2+(y-b)^2}} = 0$$

where $L(x,y) = f_1(a,b)(x-a) + f_2(a,b)(y-b) + f(a,b)$ is the linear approximation.

In a more general case, let $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ at $\vec{a} = \langle a_1, \ldots, a_n \rangle$. $f(x)$ is linearly approximate to $L(\vec{x}) = f(\vec{a}) + f_1(\vec{a})(x_1 - a_1) + \cdots + f_n(\vec{a})(x_n - a_n)$. Then the differentiablity of $f(\vec{x})$ is defined as $\displaystyle\lim_{\vec{x}\to\vec{a}} \frac{f(\vec{x}) - L(\vec{x})}{|\vec{x}-\vec{a}|} = 0$.

(d) **Consequences**

i. **Mean Value Theorem**

Similar to single variable functions, functions with two variables also mean value theorem.

Theorem 2. (Mean Value Theorem) If $f_1(x,y)$ and $f_2(x,y)$ are continuous in a neighbourhood of the point (a,b), and if the absolute values of h and k are sufficiently small, then there exist numbers θ_1 and θ_2, each between 0 and 1, such that

$$f(a+h, b+k) - f(a,b) = hf_1(a+h\theta_1, b+k) + kf_2(a, b+k\theta_2)$$

Proof. Reconstruct $f(a+h,b+k) - f(a,b)$,

$$f(a+h,b+k) - f(a,b) = \underbrace{f(a+h,b+k) - f(a,b+k)}_{\Delta u(x)} + \underbrace{f(a,b+k) - f(a,b)}_{\Delta v(y)}$$

where $u(x) = f(x, b+k)$ and $v(y) = f(a,y)$. Then by mean value theorem on single variable, we know there exists $0 < \theta_1 < 1$ such that $u(a+h) - u(a) =$

$hu'(a+h\theta_1) = hf_1(a+h\theta_1, b+k)$. And similarly there exists $0 < \theta_2 < 1$ such that $v(b+k) - v(b) = kv'(b+k\theta_2) = kf_2(a, b+k\theta_2)$. Therefore, we have

$$f(a+h, b+k) - f(a,b) = hf_1(a+h\theta_1, b+k) + kf_2(a, b+k\theta_2)$$

ii. **Claim:** If f_1 and f_2 are continuous in a neighbourhood of the point (a,b), then f is differentiable at (a,b).
Proof. We know, by mean value theorem of multivariables,

$$\begin{aligned}
& \left|\frac{f(a+h,b+k) - f(a,b) - f_1(a,b)h - f_2(a,b)k}{\sqrt{h^2+k^2}}\right| \\
= & \left|\frac{hf_1(a+h\theta_1, b+k) + kf_2(a, b+k\theta_2) - f_1(a,b)h - f_2(a,b)k}{\sqrt{h^2+k^2}}\right| \qquad 0 < \theta_1, \theta_2 < 1 \\
= & \left|\frac{h(f_1(a+h\theta_1, b+k) - f_1(a,b))}{\sqrt{h^2+k^2}} + \frac{k(f_2(a, b+k\theta_2) - f_2(a,b))}{\sqrt{h^2+k^2}}\right| \\
\leq & \left|\frac{h(f_1(a+h\theta_1, b+k) - f_1(a,b))}{\sqrt{h^2+k^2}}\right| + \left|\frac{k(f_2(a, b+k\theta_2) - f_2(a,b))}{\sqrt{h^2+k^2}}\right| \\
\leq & |(f_1(a+h\theta_1, b+k) - f_1(a,b))| + |(f_2(a, b+k\theta_2) - f_2(a,b))|
\end{aligned}$$

When we take the limit $h \to 0$ and $k \to 0$, we know $|(f_1(a+h\theta_1, b+k) - f_1(a,b))| + |(f_2(a, b+k\theta_2) - f_2(a,b))| \to 0$. Therefore

$$\lim_{(h,k)\to(0,0)} \left|\frac{f(a+h,b+k) - f(a,b) - f_1(a,b)h - f_2(a,b)k}{\sqrt{h^2+k^2}}\right| = 0$$

and $f(x,y)$ is differentiable.

iii. **Claim:** If $f(x,y)$ is differentiable at (a,b), then $\left.\frac{\partial f}{\partial x}\right|_{(a,b)}$ and $\left.\frac{\partial f}{\partial y}\right|_{(a,b)}$ exist.
Proof. They are true by definition, which requires $L(x,y) = f_1(a,b)(x-a) + f_2(a,b)(y-b) + f(a,b)$.

iv. **Claim:** If $f(x,y)$ is differentiable at (a,b), $f(x,y)$ is continuous at (a,b).
Proof. This proof would be similar to the same claim for single variable that differentiablitly implies continuity. We have $f(a+h, b+k) - f(a,b) = (f(a+h, b+k) - f(a,b) - f_1(a,b)h - f_2(a,b)k) + (f_1(a,b)h + f_2(a,b)k) = \sqrt{h^2+k^2} \times \frac{f(a+h,b+k) - f(a,b) - f_1(a,b)h - f_2(a,b)k}{\sqrt{h^2+k^2}} + (f_1(a,b)h + f_2(a,b)k)$. Therefore we have

$$\begin{aligned}
& \lim_{(h,k)\to(0,0)} (f(a+h,b+k) - f(a,b)) \\
= & \lim_{(h,k)\to(0,0)} \sqrt{h^2+k^2} \lim_{(h,k)\to(0,0)} \frac{f(a+h,b+k) - f(a,b) - f_1(a,b)h - f_2(a,b)k}{\sqrt{h^2+k^2}} \\
& + \lim_{(h,k)\to(0,0)} f_1(a,b)h + f_2(a,b)k \\
= & \ 0
\end{aligned}$$

Therefore, $f(x,y)$ is continuous at (a,b).

(e) **Summary**
In summary, we have the following diagram:

$$\begin{tikzcd}
\text{(i) } f_x, f_y \text{ are continuous} \arrow[d, Rightarrow] & \text{(iii) exists a non-vertical tangent plane at } (a,b) \arrow[dl, Leftrightarrow] \\
\text{(ii) } f \text{ is differentiable at } (a,b) \arrow[d, Rightarrow] \arrow[dr, Rightarrow] & \\
\text{(iv) } f_x(a,b),\ f_y(a,b) \text{ exists} & \text{(v) } f \text{ is continuous at } (a,b)
\end{tikzcd}$$

Notice that: (ii) can't imply (i), (iv) can't imply (ii) and (v), (v) can't imply (iv) and (ii), (ii) can't imply (i). Here are some counter examples.

Example 11. Consider $f(x,y) = \sqrt{x^2+y^2}$. Since $\lim\limits_{(x,y)\to(0,0)} \sqrt{x^2+y^2} = 0 = f(0,0)$, $f(x,y)$ is continuous at $(0,0)$. However, $f_x(0,0) = \lim\limits_{x\to 0} \dfrac{\sqrt{x^2}-0}{x-0} = \lim\limits_{x\to 0} \dfrac{|x|}{x}$ does not exists. This example shows that: continuity of $f(x,y)$ does not imply the existence of first partial derivative (v does not imply iv).

Example 12. Consider $f(x,y) = \begin{cases} \dfrac{xy}{x^2+y^2} & (x,y) \neq (0,0) \\ 0 & (x,y) = (0,0) \end{cases}$. We have $f_x(0,0) = \lim\limits_{x\to 0} \left(\dfrac{0x}{x^2+0} - 0 \right) = 0$ and $f_y(0,0) = \lim\limits_{y\to 0} \left(\dfrac{0y}{y^2+0} - 0 \right) = 0$ exists. However we have $\lim\limits_{\substack{(x,y)\to(0,0) \\ y=x}} \dfrac{xy}{x^2+y^2} = \lim\limits_{x\to 0} \dfrac{x^2}{2x^2} = \dfrac{1}{2}$ and $\lim\limits_{\substack{(x,y)\to(0,0) \\ y=0}} \dfrac{xy}{x^2+y^2} = 0$. Therefore $\lim\limits_{(x,y)\to(0,0)} \dfrac{xy}{x^2+y^2}$ does not exists. This example shows that existence of first partial derivative does not imply the continuity of $f(x,y)$ (iv does not imply v).

Example 13. Consider $f(x,y) = \begin{cases} \dfrac{xy}{\sqrt{x^2+y^2}} & (x,y) \neq (0,0) \\ 0 & (x,y) = (0,0) \end{cases}$. Since $0 \leq |x| \dfrac{|y|}{\sqrt{x^2+y^2}} \leq |x|$, by squeeze theorem, we have $\lim_{(x,y)\to(0,0)} f(x,y) = 0$. Therefore $f(x,y)$ is continuous at $(0,0)$. We also know that $f_x(0,0) = 0$ and $f_y(0,0) = 0$ exist. However, $\lim\limits_{(x,y)\to(0,0)} \dfrac{f(x,y) - L(x,y)}{\sqrt{(x-0)^2+(y-0)^2}} = \lim\limits_{(x,y)\to(0,0)} \dfrac{xy}{x^2+y^2}$ does not exist. This example shows that: the continuity of $f(x,y)$ or the existence of first partial derivative does not imply the differentiablity of $f(x,y)$. (iv,v do not imply ii).

Example 14. Consider $f(x,y) = \begin{cases} (x^2+y^2)\sin\left(\dfrac{1}{\sqrt{x^2+y^2}}\right) & (x,y) \neq (0,0) \\ 0 & (x,y) = (0,0) \end{cases}$. We have $f_x(0,0) = \lim_{x\to 0} \dfrac{x^2 \sin\left(\frac{1}{x}\right) - 0}{x - 0} = 0$ and $f_y(0,0) = \lim_{y\to 0} \dfrac{y^2\sin\left(\frac{1}{y}\right) - 0}{y - 0} = 0$. Therefore $L(x,y) = f(0,0) + f_x(0,0)(x-0) + f_y(0,0)(y-0) = 0$. Then

$$\lim_{(x,y)\to(0,0)} \frac{f(x,y) - L(x,y)}{\sqrt{(x-0)^2 + (y-0)^2}} = \lim_{(x,y)\to(0,0)} \sqrt{x^2+y^2}\sin\left(\frac{1}{\sqrt{x^2+y^2}}\right) = 0$$

by squeeze theorem. However two first partial derivative

$$f_x(x,y) = \begin{cases} 2x\sin\left(\dfrac{1}{\sqrt{x^2+y^2}}\right) - \dfrac{x\cos\left(\frac{1}{\sqrt{x^2+y^2}}\right)}{\sqrt{x^2+y^2}} & (x,y) \neq (0,0) \\ 0 & (x,y) = (0,0) \end{cases}$$

and

$$f_y(x,y) = \begin{cases} 2y\sin\left(\dfrac{1}{\sqrt{x^2+y^2}}\right) - \dfrac{y\cos\left(\frac{1}{\sqrt{x^2+y^2}}\right)}{\sqrt{x^2+y^2}} & (x,y) \neq (0,0) \\ 0 & (x,y) = (0,0) \end{cases}$$

are not continuous since $\lim_{(x,y)\to(0,0)} f_x(x,y)$ and $\lim_{(x,y)\to(0,0)} f_y(x,y)$ diverge. This example shows the differentiablity does not imply the continuity of first partial derivatives (ii does not imply i).

3. **Chain Rule**

   (a) **Special Case:** Let $f(x,y)$ and $x = x(t)$, $y = y(t)$, we want to find the derivative of f with respect to t. We call the linearization of $f(x,y)$, we have

   $$\Delta f(x,y) = f(a+\Delta x, b+\Delta y) - f(a,b) = f_1(a,b)\Delta x + f_2(a,b)\Delta y$$

   If we divide Δt on both sides, we have

   $$\frac{\Delta f}{\Delta t} = \frac{\partial f}{\partial x}\frac{\Delta x}{\Delta t} + \frac{\partial f}{\partial y}\frac{\Delta y}{\Delta t}$$

   If we take $\Delta x \to 0$, it is plausible that

   $$\frac{df}{dt} = \frac{\partial f}{\partial x}\frac{dx}{dt} + \frac{\partial f}{\partial y}\frac{dy}{dt}$$

Theorem 3. If $z = f(x, y)$ with continuous first partial derivatives, and if x and y are differentiable functions of t, then

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

Proof. Let $f(x(t), y(t)) = g(t)$, $x = x(t)$ and $y = y(t)$. Then $\frac{dz}{dt} = \lim_{h\to 0} \frac{g(t+h) - g(t)}{h} = \lim_{h\to 0} \frac{f(x(t+h), y(t+h)) - f(x(t), y(t))}{h} = \lim_{h\to 0} \frac{f(x(t+h), y(t+h)) - f(x(t), y(t+h))}{h} + \frac{f(x(t), y(t+h)) - f(x(t), y(t))}{h}$. By the definition and the chain rule of single variable, the first part of the limit is $f_x(x, y)x'(t)$ and the second part of the limit is $f_y(x, y)y'(t)$. Therefore we have

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

(b) **General Case:** Let $z = f(x, y)$, $x = x(t, s)$ and $y = y(t, s)$. If we want to find the $\frac{\partial z}{\partial t}$ and $\frac{\partial z}{\partial s}$, we could make a similar guess as the special case.

Theorem 4. Let $z = f(x, y)$ where $x = x(s, t)$ and $y = y(s, t)$. Suppose that

- $p = x(a, b)$ and $q = y(a, b)$,
- the first partial derivatives of x and y exist at the point (a, b), and
- f is differentiable at the point (p, q).

Then $z = z(s, t) = f(x(s, t), y(s, t))$ has first partial derivatives with respect to s and t at (a, b) and

$$z_s(a, b) = f_x(p, q)x_s(a, b) + f_y(p, q)y_s(a, b)$$

$$z_t(a, b) = f_x(p, q)x_t(a, b) + f_y(p, q)y_t(a, b)$$

That is,

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

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

Proof. As the proof in single variable, we construct

$$E(h, k) = \begin{cases} 0 & (h, k) = (0, 0) \\ \dfrac{f(p+h, q+k) - f(p, q) - hf_x(p, q) - kf_y(p, q)}{\sqrt{p^2 + q^2}} & (h, k) \neq (0, 0) \end{cases}$$

which is continuous since $f(x,y)$ is differentiable (p,q). Then we have

$$f(p+h,q+k)-f(p,q)=hf_x(p,q)+kf_y(p,q)+\sqrt{p^2+q^2}E(h,k)$$

Let $h=x(a+\sigma,b)-x(a,b)$ and $k=y(a+\sigma,b)-y(a+\sigma,b)$. Then we divide σ on both sides

$$\frac{f(p+h,q+k)-f(p,q)}{\sigma}=\frac{hf_x(p,q)+kf_y(p,q)+\sqrt{p^2+q^2}E(h,k)}{\sigma}$$

Since $\lim\limits_{\sigma\to 0}\dfrac{h}{\sigma}=\lim\limits_{\sigma\to 0}=\dfrac{x(a+\sigma,b)-x(a,b)}{\sigma}=x_s(a,b)$ and similarly $\lim\limits_{\sigma\to 0}\dfrac{k}{\sigma}=y_s(a,b)$, if we take the $\sigma\to 0$, $(h,k)\to(0,0)$ and then

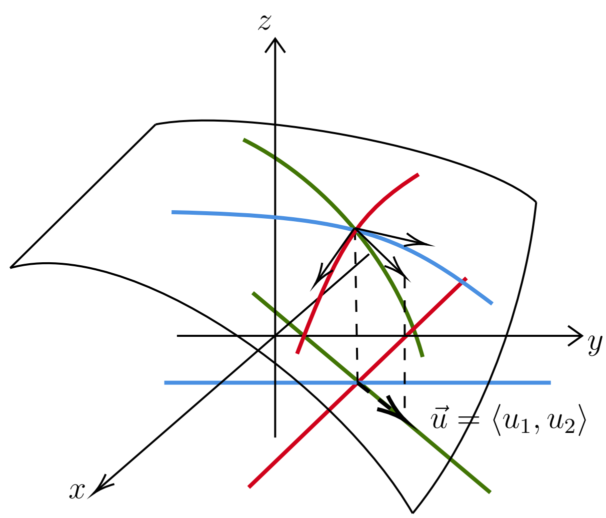
$$z_s(a,b)=f_x(p,q)x_s(a,b)+f_x(p,q)x_s(a,b)$$

Similarly, we also have $z_t(a,b)=f_x(p,q)x_t(a,b)+f_x(p,q)x_t(a,b)$.

4. **Direction Derivatives**

   (a) **Definition**

   We have partial derivatives indicating the changing rate along x and y direction. Then we want to find the changing rate along an arbitrary direction $\vec{u}=\langle u_1,u_2\rangle$



Definition 9. Let $\vec{u}=\langle u_1,u_2\rangle$ be a unit vector where $u_1^2+u_2^2=1$. The *directional derivative* of $f(x,y)$ at (a,b) in the direction of $\vec{u}$ is the rate of change of $f(x,y)$ with respect to distance measured at (a,b) along a ray in the direction of $\vec{u}$ in the xy-plane. This directional derivative is given by

$$D_{\vec{u}}f(a,b)=\lim_{h\to 0}\frac{f(a+hu_1,b+hu_2)-f(a,b)}{h}$$

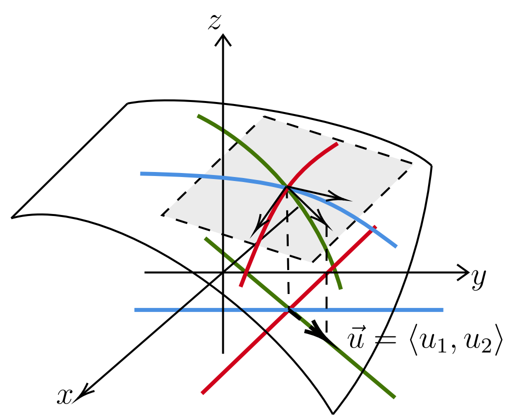
The limit is also equivalent to

$$D_{\vec{u}}f(a,b)=\left.\frac{d}{dt}\right|_{t=0}f(a+tu_1,b+tu_2)$$

if the derivative exists. Then if f is differentiable, we could evaluate the derivative by chain rule,

$$D_{\vec{u}}f(a,b) = u_1 f_1(a,b) + u_2 f_2(a,b)$$

This is plausible if we consider the linearization of f: $f(x,y) \approx f(a,b) + f_1(a,b)(x - a) + f_2(a,b)(y - b)$.



We could rearrange the equation,

$$f(a + u_1\Delta, y + u_2\Delta) - f(a,b) \approx f_1(a,b)u_1\Delta + f_2(a,b)u_2\Delta$$

Divide Δ on both sides,

$$\frac{f(a + u_1\Delta, y + u_2\Delta) - f(a,b)}{\Delta} \approx f_1(a,b)u_1 + f_2(a,b)u_2$$

Take the limit $\Delta \to 0$,

$$D_{\vec{u}}f(a,b) = u_1 f_1(a,b) + u_2 f_2(a,b)$$

(b) **Gradient**

The direction derivative could be rewritten as

$$D_{\vec{u}}f(a,b) = u_1 f_1(a,b) + u_2 f_2(a,b) = \langle u_1, u_2 \rangle \cdot \langle f_1(a,b), f_2(a,b) \rangle$$

We could see $\langle f_1(a,b), f_2(a,b) \rangle$ seems to be a general constant vector of function f. Therefore we want to define it as the gradient.

Definition 10. At any point (x,y) where the first partial derivatives of the function $f(x,y)$ exist, we define the *gradient vector* $\nabla f(x,y) = \mathbf{grad} f(x,y)$ by

$$\nabla f(x,y) = \mathbf{grad} f(x,y) = \langle f_x(x,y), f_y(x,y) \rangle$$

Then we would have a theorem to evaluate the directional derivatives.

Theorem 5. If f is differentiable at (a,b) and $\vec{u} = \langle u_1, u_2 \rangle$ is a unit vector, then the directional derivative of f at (a,b) in the direction of $\vec{u}$ is given by

$$D_{\vec{u}}f(a,b) = \vec{u} \cdot \nabla f(a,b)$$

Proof. By chain rule, we have $D_{\vec{u}}f(a,b) = \left.\frac{d}{dt}\right|_{t=0} f(a+tu_1, b+tu_2) = \vec{u}\cdot\nabla f(a,b)$.

Corollary: If $f(x,y)$ is differentiable at (a,b), then we have $D_{\vec{u}}f(a,b) = \vec{u}\cdot\nabla f(a,b) = |\nabla f(a,b)||\vec{u}|\cos\theta = |\nabla f(a,b)|\cos\theta$ where θ is the angle between $\nabla f(a,b)$ and $\vec{u}$.

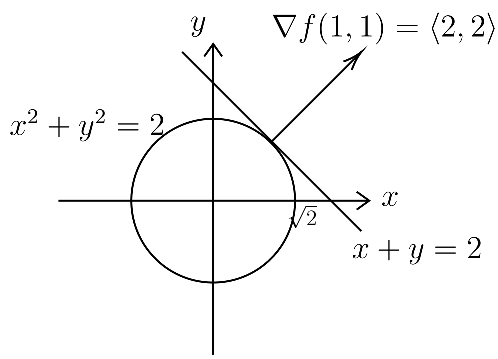
Note: If $f(x,y)$ is not differentiable at (a,b), the gradient is still defined as $\nabla f(a,b) = \langle f_x(a,b), f_y(a,b)\rangle$ and the directional derivative is still defined as the limit $D_{\vec{u}}f(a,b) = \lim_{h\to 0}\frac{f(a+hu_1, b+hu_2)-f(a,b)}{h}$ provided the limit exists. However, the theorem $D_{\vec{u}}f(a,b) = \vec{u}\cdot\nabla f(a,b)$ if not necessary to be true. Let's loot at an example.

Example 15. Consider $f(x,y) = \begin{cases} \dfrac{xy^2}{x^2+y^4} & (x,y)\neq(0,0) \\ 0 & (x,y)=(0,0) \end{cases}$. Since $\lim_{\substack{(x,y)\to(0,0)\\ x=y}} f(x,y) = \lim_{x\to 0}\frac{x^3}{x^2+y^4}$ does not exist, we know $f(x,y)$ is not continuous and therefore $f(x,y)$ is not differentiable at $(0,0)$. By definition, we still have $\nabla f(0,0) = \langle 0,0\rangle$ and $D_{\vec{u}}f(0,0) = \lim_{t\to 0}\frac{1}{t}\left[f(tu_1,tu_2)-f(0,0)\right] = \lim_{t\to 0}\frac{u_1u_2^2}{u_1^2+u_2^4t^2} = \begin{cases} \dfrac{u_2^2}{u_1} & u_1\neq 0 \\ 0 & u_1 = 0 \end{cases}$. However, $D_{\vec{u}}f(0,0) \neq \nabla f(0,0)\cdot\vec{u}$.

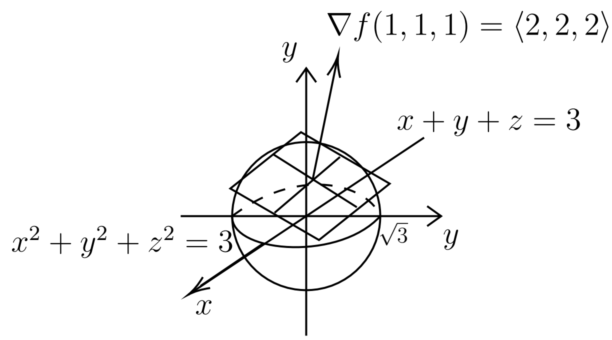
(c) **Geometry Properties**

We want to explore the geometry properties of directional vector and gradients. First of all, let's look at few examples.

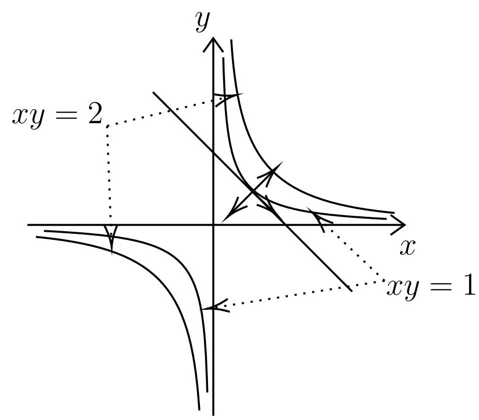
Example 16. Consider the function $f(x,y) = x^2+y^2$ with the level curve $f(x,y)=2$. The gradient of $f(x,y)$ at $(1,1)$ is $\nabla f(1,1) = \langle 2,2\rangle$ and the tangent line of the level curve $x^2+y^2=2$ at $(1,1)$ is $x+y=2$. We could see the gradient vector is perpendicular to the tangent line of the level curve at $(1,1)$.



Example 17. Consider the function $f(x,y) = x^2+y^2+z^2$ with the level curve $f(x,y,z)=3$. The gradient of $f(x,y,z)$ at $(1,1,1)$ is $\nabla f(1,1,1) = \langle 2,2,2\rangle$ and the tangent plant of the level curve $x^2+y^2+z^2=3$ is $x+y+z=3$. We could see the gradient vector is perpendicular to the tangent to the line of the level curve at $(1,1,1)$.



Example 18. Consider the function $f(x,y)=xy$ with $f(x,y)=1$ and $f(x,y)=2$. We know the gradient vector of $f(x,y)$ at $(1,1)$ is $\nabla f(1,1)=\langle 1,1\rangle$. Recall the changing rate along direction $\vec{u}$ is $D_{\vec{u}}f = |\nabla f|\cos\theta$ where θ is the angle between $\nabla f(1,1)$ and $\vec{u}$. Then $-|\nabla f| \le D_{\vec{u}}f \le |\nabla f|$. Then the fast increasing rate at $(1,1)$ is $|\nabla f(1,1)| = \sqrt{2}$ when $\theta=0$ and the fast decreasing rate at $(1,1)$ is $-|\nabla f(1,1)| = -\sqrt{2}$ when $\theta=\pi$. Notice when $\theta = \dfrac{\pi}{2}$, $D_{\vec{u}}f(1,1)=0$. This indicates that along the direction perpendicular to the gradient vector, $f(x,y)$ remains unchanged, i.e. is the level curve. This explains why the gradient vector $\nabla f(1,1)=\langle 1,1\rangle$ is perpendicular to the tangent line $x+y=2$ of the level curve $f(x,y)=1$.



In summary, if $\nabla f(x,y)\ne\vec{0}$ and $f(x,y)$ is differentiable at (a,b), we have the following geometry properties:

i. At (a,b), $f(x,y)$ increase most rapidly with the rate $|\nabla f(a,b)|$ in the direction of $\nabla f(a,b)$ and $f(x,y)$ decrease most rapidly with the rate $-|\nabla f(a,b)|$ in the direction of $\nabla f(a,b)$.
Proof. We have $D_{\vec{u}}f(a,b) = |\nabla f(a,b)|\cos\theta$. The maximum of $D_{\vec{u}}f(a,b)$ is $|\nabla f(a,b)|$ with $\theta=0$ and the minimum of $D_{\vec{u}}f(a,b)$ is $-|\nabla f(a,b)|$ with $\theta=\pi$.

ii. $\nabla f(a,b)$ is a normal vector to the level curve of f that passes through (a,b). The rate of change of $f(x,y)$ at (a,b) is zero in directions tangent to the level curve of f that passes through (a,b).
Proof. Let $\vec{r}(t)=\langle x(t),y(t)\rangle$ be the parameterized curve of the level curve and $x(0)=a$, $y(0)=b$. Therefore $f(x(0),y(0))=f(a,b)$. Take the derivative near

$t = 0$ with respect to t, we have

$$f_x(x(0), y(0)) \left.\frac{dx}{dt}\right|_{t=0} + f_x(x(0), y(0)) \left.\frac{dy}{dt}\right|_{t=0} = 0$$

This is also $\nabla f(0,0) \cdot \frac{d\vec{r}}{dt} = 0$. Therefore $\nabla f(a,b)$ is perpendicular to the tangent vector of $\frac{d\vec{r}}{dt}$ to the level curve at (a,b). Therefore, the changing rate along the tangent line of the level curve at (a,b) is $D_{\vec{u}} f(a,b) = |\nabla f(a,b)| \cos\frac{\pi}{2} = 0$.

5. **Implicit Differentiation**

(a) **Case 1:** Given $F(x,y,z) = 0$ which is locally differentiable near (x_0, y_0, z_0), there is 1 constraint equation and then 1 dependent variable and $3 - 1 = 2$ independent variables. We hope to find the derivative of z with respect to x and y if possible. This means that we assume we could find $z = z(x,y)$. First, we take the derivative with respect to x and y,

$$\begin{cases} F_1(x,y,z) + F_3(x,y,z)\dfrac{\partial z}{\partial x} = 0 \\ F_2(x,y,z) + F_3(x,y,z)\dfrac{\partial z}{\partial y} = 0 \end{cases}$$

If $F_3(x_0, y_0, z_0) \neq 0$, we could solve

$$\left.\frac{\partial z}{\partial x}\right|_{(x_0,y_0)} = -\frac{F_1(x_0,y_0,z_0)}{F_3(x_0,y_0,z_0)}$$

and

$$\left.\frac{\partial z}{\partial y}\right|_{(x_0,y_0)} = -\frac{F_2(x_0,y_0,z_0)}{F_3(x_0,y_0,z_0)}$$

And this indicates that we could find $z = z(x,y)$ at (x_0, y_0). Also, we could make a linear approximation of $z = z(x,y)$ at (x_0, y_0),

$$z(x,y) \approx L(x,y) = z_0 + \left.\frac{\partial z}{\partial x}\right|_{(x_0,y_0)} (x - x_0) + \left.\frac{\partial z}{\partial y}\right|_{(x_0,y_0)} (y - y_0)$$

with the error term $E(x,y) = z(x,y) - L(x,y)$. This would be a proper approximation if $\displaystyle\lim_{(x,y)\to(x_0,y_0)} \frac{|E(x,y)|}{|(x,y) - (x_0,y_0)|} = 0$.

Notice that if $F_2 \neq 0$, we could find $y = y(x,z)$ and the corresponding derivatives and if $F_3 \neq 0$, we could find $x = x(y,z)$ and the corresponding derivatives.

(b) **Case 2:** Given a system of equations $\begin{cases} F(x,y,u,v) = 0 \\ G(x,y,u,v) = 0 \end{cases}$ which are locally differentiable on $\mathscr{D}$ and $(x_0, y_0, u_0, v_0) \in \mathscr{D}$, there are 2 constraint equations and then 2 dependent variables and $4-2=2$ independent variables. We hope to find the derivatives of u and v with respect to x and y if possible. This means that we assume we could find $u=u(x,y)$ and $v=v(x,y)$. First, we take the derivatives with respect to x,

$$\begin{cases} F_1(x,y,u,v)+F_3(x,y,u,v)\dfrac{\partial u}{\partial x}+F_4(x,y,u,v)\dfrac{\partial v}{\partial x}=0 \\ G_1(x,y,u,v)+G_3(x,y,u,v)\dfrac{\partial u}{\partial x}+G_4(x,y,u,v)\dfrac{\partial v}{\partial x}=0 \end{cases}$$

This could be written as

$$\underbrace{\begin{pmatrix} F_3 & F_4 \\ G_3 & G_4 \end{pmatrix}}_{J}\begin{pmatrix} \dfrac{\partial u}{\partial x} \\ \dfrac{\partial v}{\partial x} \end{pmatrix}=\begin{pmatrix} -F_1 \\ -G_1 \end{pmatrix}$$

With Cramer's Rule, if $|J|\neq 0$, we know the solutions are

$$\frac{\partial u}{\partial x}=\frac{\begin{vmatrix} -F_1 & F_4 \\ -G_1 & G_4 \end{vmatrix}}{\begin{vmatrix} F_3 & F_4 \\ G_3 & G_4 \end{vmatrix}}=\frac{G_1F_4-F_1G_4}{F_3G_4-G_3F_4},\ \frac{\partial v}{\partial x}=\frac{\begin{vmatrix} F_3 & -F_1 \\ G_3 & -G_1 \end{vmatrix}}{\begin{vmatrix} F_3 & F_4 \\ G_3 & G_4 \end{vmatrix}}=\frac{G_3F_1-F_3G_1}{F_3G_4-G_3F_4}$$

Then we take the derivatives with respect to y,

$$\begin{cases} F_2(x,y,u,v)+F_3(x,y,u,v)\dfrac{\partial u}{\partial y}+F_4(x,y,u,v)\dfrac{\partial v}{\partial y}=0 \\ G_2(x,y,u,v)+G_3(x,y,u,v)\dfrac{\partial u}{\partial y}+G_4(x,y,u,v)\dfrac{\partial v}{\partial y}=0 \end{cases}$$

This could be written as

$$\underbrace{\begin{pmatrix} F_3 & F_4 \\ G_3 & G_4 \end{pmatrix}}_{J}\begin{pmatrix} \dfrac{\partial u}{\partial x} \\ \dfrac{\partial v}{\partial x} \end{pmatrix}=\begin{pmatrix} -F_2 \\ -G_2 \end{pmatrix}$$

With Cramer's Rule, if $|J|=0$, we know the solutions are

$$\frac{\partial u}{\partial x}=\frac{\begin{vmatrix} -F_2 & F_4 \\ -G_2 & G_4 \end{vmatrix}}{\begin{vmatrix} F_3 & F_4 \\ G_3 & G_4 \end{vmatrix}}=\frac{G_2F_4-F_2G_4}{F_3G_4-G_3F_4},\ \frac{\partial v}{\partial x}=\frac{\begin{vmatrix} F_3 & -F_2 \\ G_3 & -G_2 \end{vmatrix}}{\begin{vmatrix} F_3 & F_4 \\ G_3 & G_4 \end{vmatrix}}=\frac{G_3F_2-F_3G_2}{F_3G_4-G_3F_4}$$

Therefore, we could find out $u=u(x,y)$ and $v=v(x,y)$. Also, we could make a linear approximation of $u(x,y)$ and $v(x,y)$ around (x_0,y_0),

$$\begin{pmatrix} u \\ v \end{pmatrix}\approx\begin{pmatrix} u_0 \\ v_0 \end{pmatrix}+\left.\begin{pmatrix} \dfrac{\partial u}{\partial x} & \dfrac{\partial u}{\partial y} \\ \dfrac{\partial v}{\partial x} & \dfrac{\partial v}{\partial y} \end{pmatrix}\right|_{(x_0,y_0)}\begin{pmatrix} x-x_0 \\ y-y_0 \end{pmatrix}=\overrightarrow{L}(x,y)$$

where the error term is

$$\overrightarrow{E}(x,y) = \begin{pmatrix} u(x,y) \\ v(x,y) \end{pmatrix} - \overrightarrow{L}(x,y)$$

and this is a proper approximation if $\displaystyle\lim_{(x,y)\to(x_0,y_0)} \frac{|\overrightarrow{E}(x,y)|}{|(x,y)-(x_0,y_0)|} = 0.$

Notice, it is also possible find other solutions to the system of equations and the derivatives if the corresponding determinants are non-zero. For example, we could have

$$\begin{cases} x = x(u.v) \\ y = y(u,v) \end{cases} \tag{A}$$

and

$$\begin{cases} x = x(y.v) \\ u = u(y,v) \end{cases} \tag{B}$$

For the derivative of (A), we use $\left(\dfrac{\partial x}{\partial v}\right)_u$ to indicate u is independent and held unchanged. For the derivative of (B), we use $\left(\dfrac{\partial x}{\partial v}\right)_y$ to indicate y is independent and held unchanged. This could reduce the ambiguity.

(c) **General Case**

Recall the determinants with form as $\begin{vmatrix} F_3 & F_4 \\ G_3 & G_4 \end{vmatrix}$ are quite common. We want to give a general definition of them.

Definition 11. The *Jacobian determinant* (or simply *Jacobian*) of the n functions, $F^{(1)}(x_1,\ldots,x_m)=0,\ \ldots,\ F^{(n)}(x_1,\ldots,x_m)=0$ where $m \geq n$, with respect to n variables $x_{m_1},\ \ldots, x_{m_n}$ where $m_1,\ldots,m_n \in \{1,\ldots,m\}$ is the determinant

$$\frac{\partial(F^{(1)},\ldots,F^{(n)})}{\partial(x_{m_1},\ \ldots,x_{m_n})} = \begin{vmatrix} F^{(1)}_{m_1} & \cdots & F^{(1)}_{m_n} \\ \vdots & \ddots & \vdots \\ F^{(n)}_{m_1} & \cdots & F^{(n)}_{m_n} \end{vmatrix}$$

Besides, let's explore how the number of equations and the number of variables determine the number of dependent and independent variables. Let $m \geq n$ and

$$\begin{cases} F^{(1)}(x_1,\cdots,x_m) = 0 \\ F^{(2)}(x_1,\cdots,x_m) = 0 \\ \qquad\vdots \\ F^{(n)}(x_1,\cdots,x_m) = 0 \end{cases}$$

There are n equations and m variables. Then we have n constraint equations and n dependent variables. Then there are $m-n$ independent variables.

With the above preparation, we could have the implicit function theorem.

Theorem 6. (Implicit Function Theorem) Consider a system of n equations in $n+m$ variables

$$\begin{cases} F^{(1)}(x_1,\cdots,x_m,y_1,\cdots,y_n)=0 \\ F^{(2)}(x_1,\cdots,x_m,y_1,\cdots,y_n)=0 \\ \vdots \\ F^{(n)}(x_1,\cdots,x_m,y_1,\cdots,y_n)=0 \end{cases}$$

and a point $P_0=(a_1,\ldots,a_m,b_1,\ldots,b_n)$ that satisfies the system. Suppose each of the functions $F^{(i)}$ has continuous first partial derivatives with respect to each of the variables x_j and y_k ($i=1,\ldots,n$, $j=1,\ldots,m$, $k=1,\ldots,n$), near P_0. Finally, suppose that

$$\frac{\partial(F^{(1)},\ldots,F^{(n)})}{\partial(y_1,\ldots,y_n)}\neq 0$$

Then the system can be solved for $y_1,\ldots,y_n$ as functions of $x_1,\ldots,x_m$ near P_0. That is, there exist functions

$$\phi_1(x_1,\ldots,x_m),\ldots,\phi_n(x_1,\ldots,x_m)$$

such that

$$\phi_j(a_1,\ldots,a_m)=b_j \ (j=1,\ldots,n)$$

and such that

$$\begin{cases} F^{(1)}(x_1,\cdots,\phi_1(x_1,\ldots,x_m),\ldots,\phi_n(x_1,\ldots,x_m))=0 \\ F^{(2)}(x_1,\cdots,\phi_1(x_1,\ldots,x_m),\ldots,\phi_n(x_1,\ldots,x_m))=0 \\ \vdots \\ F^{(n)}(x_1,\cdots,\phi_1(x_1,\ldots,x_m),\ldots,\phi_n(x_1,\ldots,x_m))=0 \end{cases}$$

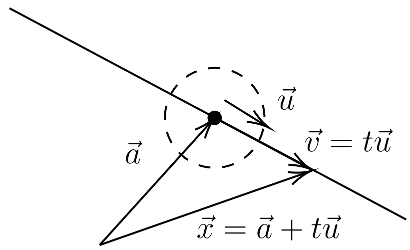
hold for all $(x_1,\ldots,x_m)$ sufficiently near $(a_1,\ldots,a_m)$. Moreover,

$$\frac{\partial\phi_i}{\partial x_j}=\left(\frac{\partial y_i}{\partial x_j}\right)_{x_1,\ldots,x_{j-1},x_j,\ldots,x_m}=-\frac{\dfrac{\partial(F^{(1)},\ldots,F^{(n)})}{\partial(y_1,\ldots,x_i,\ldots,y_n)}}{\dfrac{\partial(F^{(1)},\ldots,F^{(n)})}{\partial(y_1,\ldots,y_j,\ldots,y_n)}}$$

6. **Taylor Theorem**

   (a) **General Case**

   Let $f(x_1,\ldots,x_n)$ and the 1st, 2nd, 3rd, ..., kth partial derivatives are continuous on a neighbourhood of $\vec{a}=(x_1^0,\ldots,x_n^0)$. Consider a direction unit vector $\vec{u}$ and the neighbour points of $\vec{a}$ along the direction of $\vec{u}$ could be written as $\vec{x} = \vec{a} + t\vec{u}$.



Then we have $f(\vec{x}) = f(\vec{a} + t\vec{u}) = g(t)$. By Taylor's theorem of single variable,

$$g(t) \approx g(0) + g'(0)t + \frac{1}{2}g''(0)t^2 + \cdots + \frac{1}{k!}g^{(k)}t^k$$

To expand $g(t)$, we have to find $g^{(i)}(t)$ $(i \in \{1, \ldots, k\})$. We know $g'(t)$ is just the directional derivative

$$g'(t) = \frac{d}{dt}f(\vec{a} + t\vec{u}) = \nabla f(\vec{a} + t\vec{u}) \cdot \vec{u}$$

Then we let $\vec{u}\cdot\nabla = u_1\frac{\partial}{\partial x_1} + \cdots + u_n\frac{\partial}{\partial x_n}$ be a operator applied on function $f(x_1, \ldots, x_n)$. Then it is not hard to exam that

$$g^{(k)}(t) = (\vec{u} \cdot \nabla)^k f(\vec{a} + t\vec{u}) = (\vec{u} \cdot \nabla)^k f(\vec{x})$$

Therefore we find out that

$$\begin{aligned} g(t) &\approx f(\vec{a}) + (\vec{u} \cdot \nabla)f(\vec{a})t + \frac{1}{2}(\vec{u} \cdot \nabla)^2 f(\vec{a})t^2 + \cdots + \frac{1}{k!}(\vec{u} \cdot \nabla)^k f(\vec{a})t^k \\ &= f(\vec{a}) + (\vec{v} \cdot \nabla)f(\vec{a}) + \frac{1}{2}(\vec{v} \cdot \nabla)^2 f(\vec{a}) + \cdots + \frac{1}{k!}(\vec{v} \cdot \nabla)^k f(\vec{a}) = P_k(\vec{x}) \end{aligned}$$

,which is the kth order of Taylor polynomial. Then the error term, by Taylor's theorem of single variable is

$$E(\vec{x}) - P_k(\vec{x}) = \frac{g^{(k+1)}(s)}{(k+1)!}t^{k+1} = \frac{1}{(k+1)!}(\vec{v} \cdot \nabla)^{k+1} f(\vec{a} + s\vec{u})$$

where $0 \leq s \leq k$. There the Taylor theorem converges if and only if $\lim_{t\to 0} \frac{E(\vec{x})}{t^k} = 0$. This requires the error $|E(\vec{x})| \leq |\vec{v}|^{k+1}$. In summary, we state the Taylor theorem of multivariables.

Theorem 7. Let $f(x_1, \ldots, x_n)$ be a $k+1$ times continuously differentiable function at $\vec{a} \in \mathbb{R}^n$. Then for $\vec{x} = \vec{a} + t\vec{u}$ where $\vec{u}$ is the unit directional vector, we have

$$f(x) = P_k(\vec{x}) + E(\vec{x})$$

where

$$f(\vec{x}) = f(\vec{a}) + (\vec{v} \cdot \nabla) f(\vec{a}) + \frac{1}{2}(\vec{v} \cdot \nabla)^2 f(\vec{a}) + \cdots + \frac{1}{k!}(\vec{v} \cdot \nabla)^k f(\vec{a})$$

and

$$E(\vec{x}) = \frac{1}{(k+1)!}(\vec{v} \cdot \nabla)^{k+1} f(\vec{a} + s\vec{u}) \ \ (0 \leq s \leq t)$$

and $\lim_{t \to 0} \frac{E(\vec{x})}{t^k} = 0$.

(b) **Special Case: Second Order Polynomial with Two Variables**
Let $f(x, y)$ be a 3 times continuously differentiable function at $\vec{a} = \langle a, b \rangle$ with the unit direction vector $\vec{u} = \langle u_1, u_2 \rangle$. Then the second order approximation of $f(x, y) = f(a + tu_1, b + tu_2)$ is

$$\begin{aligned} f(x,y) &\approx f(a,b) + t\vec{u} \cdot \nabla f(a,b) + \frac{1}{2}(t\vec{u} \cdot \nabla)^2 f(a,b) \\ &= f(a,b) + \underbrace{tu_1 f_1(a,b) + tu_2 f_2(a,b)}_{\text{first order}} \\ &+ \underbrace{\frac{1}{2}t^2 u_1^2 f_{11}(a,b) + \frac{1}{2}t^2 u_1 u_2 f_{12}(a,b) + \frac{1}{2}t^2 u_2 u_1 f_{21}(a,b) + \frac{1}{2}t^2 u_2^2 f(a,b)}_{\text{second order}} \end{aligned}$$

The first order term is simply $\nabla f(a,b) \cdot t\vec{u}$. And the second order term could be written as

$$\frac{1}{2}(tu_1, tu_2)\begin{pmatrix} f_{11} & f_{12} \\ f_{21} & f_{22} \end{pmatrix}\begin{pmatrix} tu_1 \\ tu_2 \end{pmatrix} = \frac{1}{2}(t\vec{u})^T H_{f_2}(t\vec{u})$$

where H_{f_2} is a symmetric matrix called Hessian matrix. Then the approximation could also be written as

$$P_2(x,y) = f(a,b) + \nabla f(a,b) \cdot \langle x - a, y - b \rangle + \frac{1}{2}\langle x - a, y - b \rangle H_{f2}\begin{pmatrix} x - a \\ y - b \end{pmatrix}$$

If the third order partial derivatives of f are continuous and bounder for (x, y) in a neighbourhood of (a, b) then the error

$$|f(x,y) - P_2(x,y)| \leq \frac{M}{3!}|(x,y) - (a,b)|^3$$

where $M \in \mathbb{R}$ depends on partial derivatives of f.

Notice that if we have more than two variable, then the Hessian matrix of the second order term is

$$H_{f_n} = \begin{pmatrix} f_{11} & \cdots & f_{1n} \\ \vdots & \ddots & \vdots \\ f_{n1} & \cdots & f_{nn} \end{pmatrix}$$

(c) **Special Case: Reducing to One Variable**

Sometimes we could reduce the multivariable function to one or a few single variable function to apply the Taylor's Theorem. There are two methods: substitution and decomposition.

Substitution usually work for the linear combination of the variables. For example, $f(x, y) = e^{2x+3y}$ at $(0, 0)$ could be reduced to $g(t) = e^t$ by substitution $t = 2x + 3y$ and expanded as $g(t) = 1 + t + \frac{1}{2!}t^2 + \frac{1}{3!}t^3 + \cdots$. And the approximation is $f(x, y) \approx 1 + (2x + 3y) + \frac{(2x + 3y)^2}{2} + \frac{(2x + 3y)^3}{6} + \cdots$.

Decomposition is more flexible. For example, we can decompose $f(x, y) = e^x \sin(y) = \left(1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots\right)\left(y - \frac{y^5}{3!} + \frac{y^5}{5!} - \cdots\right)$.

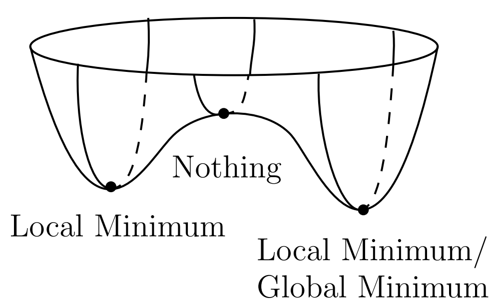
7. **Optimization**

(a) **Definition**

Recall the definitions of extreme values.

> **Definition 12.** We say $f(x, y)$ has a *local maximum* at (a, b) if $f(x, y) \leq f(a, b)$ for all points (x, y) in the domain of f sufficiently close to (a, b) and $f(x, y)$ has a *local minimum* at (a, b) if $f(x, y) \geq f(a, b)$ for all points (x, y) in the domain of f sufficiently close to (a, b). We say $f(x, y)$ has a global maximum at (a, b) if $f(x, y) \leq f(a, b)$ for all points (x, y) in the domain of $f(x, y)$ and $f(x, y)$ has a global minimum at (a, b) if $f(x, y) \geq f(a, b)$ for all points (x, y) in the domain of $f(x, y)$.

For example, we have



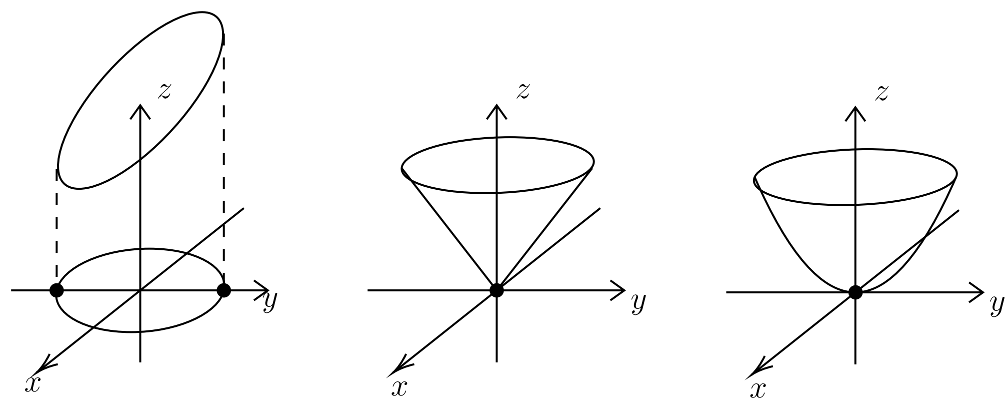
(b) **Necessary Conditions of Extreme Values and Potential Extreme Points**

As the extreme value for single variable function, it is hard to find the extreme value of multivariables directly but we could find the potential points satisfying the necessary condition.

> **Theorem 8.** If $f(\vec{x})$ has local minimum or maximum at $\vec{a}$, then we have one of the following:
>
> i. $\vec{a}$ is on the boundary of $\mathscr{D}(f)$,
>
> ii. $\nabla f(\vec{a})$ does not exist, or
>
> iii. $\nabla f(\vec{a}) = 0$.

Proof. Without loss of generality, let $f(\vec{x})$ has a local minimum at $\vec{a}$. We assume $\vec{a}$ is not on the boundary and there exists the gradient $\nabla f(\vec{a})$. Not on the boundary imply that the neighbourhood is in the domain of the function. Let $g_j(x_j) = f(a_1, \ldots, a_{j-1}, x_j, a_{j+1}, \ldots, a_n)$ which has a local minimum at a_j and the neighbourhood of a_j is in the domain. Since $g'(a_j) = \dfrac{\partial f}{\partial x_j}(a_j)$ exists, $\dfrac{\partial f}{\partial x_j}(a_j) = 0$ by interior extreme theorem. Therefore we have $\nabla f(\vec{a}) = 0$.

Example 19. Consider the following three functions.



$f(x,y) = y+3$ $(\mathscr{D}:\ x^2+y^2 \le 1)$ $\quad$ $f(x,y) = \sqrt{x^2+y^2}$ $\quad$ $f(x,y) = x^2+y^2$

The first function has a local maximum at $(0,1)$ and local minimum at $(0,-1)$, where are both on the boundary. The second function has a local minimum at $(0,0)$, where $\nabla f(0,0)$ does not exist. And the third function has a local minimum at $(0,0)$ where $\nabla f(0,0) = 0$.

Since the technique applied to three cases in the theorem are different, we want to define them separately.

Definition 13. We say $\vec{a}$ is the *critical point* of $f(\vec{x})$ if $\nabla f(\vec{a}) = 0$. We say $\vec{a}$ is the *singular point* of $f(\vec{x})$ if $\nabla f(\vec{a})$ does not exist. We say $\vec{a}$ is the *boundary point* of $f(\vec{x})$ if $\vec{a}$ is on the boundary of the domain $f(\vec{x})$.

(c) **Classification of Critical Points**

Let's first discuss how to deal with the critical points. To find out the extreme points among critical points, we have to classify the critical points.

Assume f and first, second order partial derivatives are continuous on a neighbour of $\vec{a}$. With $\nabla f(\vec{a}) = 0$, we have

$$f(\vec{x}) = f(\vec{a}) + \underbrace{\frac{1}{2}(\vec{x}-\vec{a})^T H_{f(\vec{a})}(\vec{x}-\vec{a})}_{R(\vec{x})} + O\left((\vec{x}-\vec{a})^2\right)$$

We know when $\vec{x} = \vec{a}$, $R(\vec{x})$ and $O\left((\vec{x}-\vec{a})^2\right) = 0$.
Then let $\vec{x} \neq \vec{a}$.

i. If for all $\vec{x} \in \mathbb{R}^n/\vec{a}$ sufficiently close to $\vec{a}$, $R(\vec{x}) > 0$, we can define $H_{f(\vec{a})}$ as *positive definite* and $f(\vec{x}) > f(\vec{a})$ since $O\left((\vec{x}-\vec{a})^2\right)$ is negligible. We could see $\vec{a}$ is the local minimum.
ii. Similarly, if for all $\vec{x} \in \mathbb{R}^n/\vec{a}$, $R(\vec{x}) < 0$, we can define $H_{f(\vec{a})}$ as *negative definite* and $f(\vec{x}) < f(\vec{a})$ since $O\left((\vec{x}-\vec{a})^2\right)$ is negligible. We could see $\vec{a}$ is the local maximum.
iii. If for all $\vec{x} \in \mathbb{R}^n/\vec{a}$ sufficiently close to $\vec{a}$, $R(\vec{x}) > 0$ or $R(\vec{x}) < 0$, we say $H_{f(\vec{a})}$ is *indefinite*. Since some points in the neighbour of $\vec{a}$ is larger than $f(\vec{a})$ and others are smaller, we say $\vec{a}$ is a *saddle point*.
iv. If $H_{f(\vec{a})}$ is neither positive definite, nor negative definite nor indefinite, $R(\vec{x})$ is possible to be 0 when $\vec{x} \neq \vec{a}$. Then we need more information from $O\left((\vec{x}-\vec{a})^2\right)$ and this case is inconclusive.

Let's formalize this discussion into a theorem.

Theorem 9. (Second Derivative Test) Suppose that $\vec{a} = \langle a_1, \ldots, a_n\rangle$ is a critical point of $f(\vec{x}) = f(x_1, \ldots, x_n)$ and is interior to the domain of f . Also, suppose that all the second partial derivatives of f are continuous throughout a neighbourhood of a, so that the Hessian matrix

$$H_{f(\vec{x})} = \begin{pmatrix} f_{11}(\vec{x}) & \cdots & f_{1n}(\vec{x}) \\ \vdots & \ddots & \vdots \\ f_{n1}(\vec{x}) & \cdots & f_{nn}(\vec{x}) \end{pmatrix}$$

is also continuous in that neighbourhood. Note that the continuity of the partials guarantees that H is a symmetric matrix.

i. If $H_{f(\vec{a})}$ is positive definite, $f(\vec{x})$ has a local minimum at $\vec{a}$.

ii. If $H_{f(\vec{a})}$ is negative definite, $f(\vec{x})$ has a local maximum at $\vec{a}$.

iii. If $H_{f(\vec{a})}$ is indefinite, $f(\vec{x})$ has a saddle point at $\vec{a}$.

iv. If $H_{f(\vec{a})}$ is neither positive nor negative definite, nor indefinite, this is inconclusive.

Then we need a strategy to quickly determine whether the Hessian form is positive definite, or negative definite, or indefinite or inconclusive. Let's first check a special case: diagonal matrix. Let

$$H_{f(\vec{a})} = \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{pmatrix}$$

Then $R(\vec{x}) = \dfrac{1}{2}[\lambda_1(x_1 - a_1)^2 + \cdots + (x_n - a_n)^2]$ and

$$f(\vec{x}) \approx f(a) + \frac{1}{2}[\lambda_1(x_1 - a_1)^2 + \cdots + \lambda_n(x_n - a_n)^2]$$

Then we know

i. If $\lambda_j > 0$ for all $j = \{1, \ldots, n\}$, we know $H_{f(\vec{a})}$ is positive definite and f has a local minimum at $\vec{a}$.

ii. If $\lambda_j < 0$ for all $j = \{1, \ldots, n\}$, we know $H_{f(\vec{a})}$ is positive definite and f has a local minimum at $\vec{a}$. positive definite and f has a local minimum at $\vec{a}$.

iii. If some $\lambda_j < 0$ and other $\lambda_j > 0$ for all $j = \{1, \ldots, n\}$, we know $H_{f(\vec{a})}$ is indefinite and f has a saddle point at $\vec{a}$.

iv. If some $\lambda_j \geq 0$ and other $\lambda_j \leq 0$ for all $j = \{1, \ldots, n\}$, it would be inconclusive.

Then for an arbitrary Hessian matrix, we could diagonalize it. Let $A = H_{f(\vec{a})}$ which is symmetric, then we could change coordinates to have $A = SDS^{-1}$ where D is diagonal. Also, we could use the following test. Let

$$M_1 = |f_{11}|,\ M_2 = \begin{vmatrix} f_{11} & f_{12} \\ f_{21} & f_{22} \end{vmatrix}, \ldots, M_n = \begin{vmatrix} f_{11} & \cdots & f_{1n} \\ \vdots & \ddots & \vdots \\ f_{n1} & \cdots & f_{nn} \end{vmatrix}$$

Then we have

i. If $M_1, \ldots, M_n$ are all > 0, then $H_{f(\vec{a})}$ is positive definite.

ii. If $M_1 < 0, M_2 > 0, M_3 < 0, \ldots,$, then $H_{f(\vec{a})}$ is negative definite.

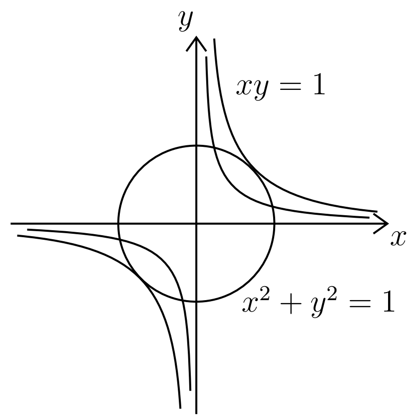
iii. If $\{M_i\}$ is in other sequences with $\det(H_{f(\vec{a})}) \neq 0$, $H_{f(\vec{a})}$ is indefinite.

iv. If $\det(H_{f(\vec{a})}) = 0$, this would be inconclusive.

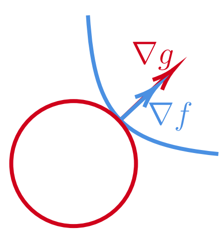
(d) **Lagrange Multipliers**

Then we want to find out how to find the extreme on the boundary, which is basically a constraint curve or surface. One way is to parameterize the curve and to plug into the target function and then to optimize with respect to the parameters. Another way is to apply Lagrange multipliers. Let's start with a single example.

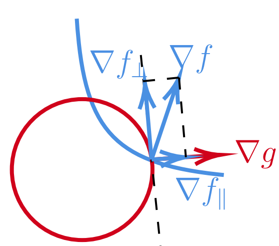
Example 20. We want to maximize $f(x, y) = xy$ subject to $x^2 + y^2 = 1$. To be more straightforward, we draw the level curve of $f(x, y)$ and the constraint equation $x^2 + y^2 = 1$.



We could see the function reaches the maximum 2 when the level curve $f(x, y) = 2$ is tangent to the constraint curve $g(x, y) = x^2 + y^2 - 1 = 0$. This means the gradient of the function $f(x, y)$ is parallel to the gradient of the constraint equation $g(x, y)$: $\nabla f = \lambda \nabla g$.



This actually makes sense. If ∇f is not parallel to ∇g when the level curve is $f(x, y) = M'$, we could decompose ∇f in the direction parallel to ∇g and in the direction perpendicular to ∇g.



Then if we go along the direction of $\nabla f_\perp$ which is tangent to the constraint curve, $f(x, y)$ would increase and M' is not the maximum.

We want to generalize the idea in the example. We construct the *Lagrange function* $\mathcal{L}(x, y, \lambda) = f(x, y) - \lambda g(x, y)$ and the extreme value is possible to reached when $\nabla\mathcal{L} = \vec{0}$ if the gradient ∇f and ∇g exist and are nonzero.

Theorem 10. (Lagrange Multipliers) Suppose that f and g have continuous first partial derivatives near the point $P_0(x_0, y_0)$ on the curve $\mathcal{C}$ with equation $g(x, y) = 0$. Suppose also that, when restricted to points on the curve $\mathcal{C}$, the function $f(x, y)$ has a local maximum or minimum value at P_0. Finally, suppose that

i. P_0 is not an endpoint of $\mathcal{C}$,

ii. $\nabla g(P_0) \neq \vec{0}$.

Then there exists a number λ_0 called *Lagrange multiplier* such that (x_0, y_0, λ_0) is a critical point of the Lagrange function $\mathcal{L}(x, y, \lambda) = f(x, y) - \lambda g(x, y)$.

Proof. First, (i) and (ii) guarantee f and g are smooth enough to have tangent lines and gradients. We assume that ∇f is not parallel to ∇g. Then we decompose ∇f in the direction perpendicular to ∇g and parallel to ∇g. By assumption, $\nabla f_{\perp} \neq \vec{0}$ and this is also the direction of the tangent line. Therefore, if we go along the curve, f would increase or decrease and at this point f is not the maximum or minimum. Therefore, ∇f is parallel to ∇g. And since $\nabla g \neq \vec{0}$, there exists a number λ_0 such that $\nabla f = \lambda_0 \nabla g$. And $\nabla\mathcal{L} \neq \vec{0}$ is just $f_x - \lambda_0 g_x = 0$, $f_y - \lambda_0 g_y = 0$ and $g(x, y) = 0$. The last one is just the constraint equation and the first two are $\nabla f = \lambda_0 \nabla g$.

Note: This theorem only state the necessary condition of a extreme value point such that $\nabla g \neq 0$ and the point is not on the curve. However, when we are seeking for the extrema, we first assume the existence of the solution of $\nabla\mathcal{L} = 0$ and we want to solve the equations

$$\begin{cases} \dfrac{\partial \mathcal{L}}{\partial x} = \dfrac{\partial f}{\partial x} - \lambda \dfrac{\partial g}{\partial x} = 0 \\ \dfrac{\partial \mathcal{L}}{\partial y} = \dfrac{\partial f}{\partial y} - \lambda \dfrac{\partial g}{\partial y} = 0 \\ \dfrac{\partial \mathcal{L}}{\partial \lambda} = g(x, y) = 0 \end{cases}$$

where the third equation is just the constraint equation and satisfied by default. However it is not, in fact, necessary to have a solution.

Note: This theorem only discuss the extreme points which is not the the endpoint of the curve, ∇g, ∇f exist and ∇g is non-zero. Therefore, in total, we have to check: the points such that

i. $\nabla\mathcal{L} = \vec{0}$,

ii. $\nabla g = 0$,

iii. ∇g or ∇f does not exist,

iv. are endpoints of the constraint curve.

Let's check an example.

Example 21. We want to maximize $f(x, y) = 2xy$ subject to $x^2 + 4y^2 = 4$. Let the Lagrange equation to be

$$\mathcal{L}(x, y, \lambda) = xy - \lambda(x^2 + 4y^2 - 4)$$

There is no endpoints and ∇f, ∇g exist and $\nabla g \neq 0$. Then we only have to solve the equations

$$\begin{cases} \dfrac{\partial \mathcal{L}}{\partial x} = 2y + 2\lambda x = 0 & (A) \\ \dfrac{\partial \mathcal{L}}{\partial y} = 2x + 8\lambda y = 0 & (B) \\ \dfrac{\partial \mathcal{L}}{\partial \lambda} = x^2 + 4y^2 - 4 = 0 & (C) \end{cases}$$

Since (C) is satisfied by default, we only have to deal with (A) and (B). Rewrite (A) and (B) as $2\lambda x = 2y$ and $8\lambda y = 2x$ and divide (A) by (B). We could finally get $x^2 = 4y^2$ and plug it into the constraint equation, which gives the solution $\left(\sqrt{2}, \dfrac{\sqrt{2}}{2}\right)$, $\left(\sqrt{2}, -\dfrac{\sqrt{2}}{2}\right)$, $\left(-\sqrt{2}, \dfrac{\sqrt{2}}{2}\right)$ and $\left(-\sqrt{2}, -\dfrac{\sqrt{2}}{2}\right)$. Since first two points have the value 2 and last two points have the value -2. we know minimum of $f(x, y)$ subject to $x^2 + 4y^2 = 4$ is 2 at $\left(\sqrt{2}, \dfrac{\sqrt{2}}{2}\right)$ and $\left(-\sqrt{2}, -\dfrac{\sqrt{2}}{2}\right)$.

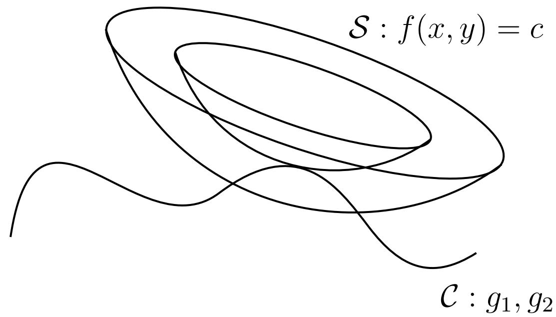
Note: We could generalize this method further more. First, we could have more variables. For example, we could optimize $f(x, y, z)$ subject to $g(x, y, z) = 0$. Also, we could have more constraint equations. For example, we could optimize $f(x, y, z)$ subject to $g_1(x, y, z) = 0$ and $g_2(x, y, z) = 0$. The Lagrange equation is

$$\mathcal{L}(x, y, z, \lambda) = f(x, y, z) - \lambda_1 g_1(x, y, z) - \lambda_2 g_2(x, y, z)$$

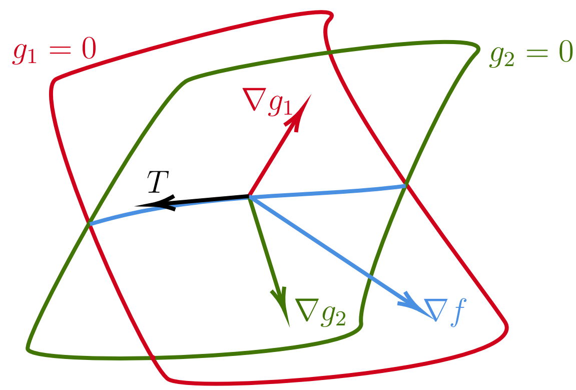
And for $\nabla \mathcal{L} = \vec{0}$, we have to solve

$$\begin{cases} \dfrac{\partial \mathcal{L}}{\partial x} = \dfrac{\partial f}{\partial x} - \lambda_1 \dfrac{\partial g_1}{\partial x} - \lambda_2 \dfrac{\partial g_2}{\partial x} = 0 & (A) \\ \dfrac{\partial \mathcal{L}}{\partial y} = \dfrac{\partial f}{\partial y} - \lambda_1 \dfrac{\partial g_1}{\partial y} - \lambda_2 \dfrac{\partial g_2}{\partial y} = 0 & (B) \\ \dfrac{\partial \mathcal{L}}{\partial z} = \dfrac{\partial f}{\partial z} - \lambda_1 \dfrac{\partial g_1}{\partial z} - \lambda_2 \dfrac{\partial g_2}{\partial z} = 0 & (C) \\ \dfrac{\partial \mathcal{L}}{\partial \lambda_1} = g_1(x, y, z) = 0 & (D) \\ \dfrac{\partial \mathcal{L}}{\partial \lambda_2} = g_2(x, y, z) = 0 & (E) \end{cases}$$

Notice that (D) and (E) are just constraint equations and are satisfied by default. The geometric meaning is that the level curve (c) is tangent to the constraint curve determined by two constraint equations.



To have that, we need $\nabla f = \lambda_1 \nabla g_1 + \lambda_2 \nabla g_2$ as a linear combination of the gradients of constraint equations.



However, this is based on the assumption that $\overrightarrow{T}$ exists and is non-zero, ∇f exists and the extreme point is not an endpoint. Therefore, in general, we have to check the points such that

i. $\nabla \mathcal{L} = \vec{0}$,
ii. $T = \nabla g_1 \times \nabla g_2 = 0$ (one of gradients is zero or two gradients are parallel),
iii. ∇f, or ∇g_1 or ∇g_2 does not exist, and
iv. are endpoints of the constraint curve.

Finally, we want to explore a little bit further more: what does the the Lagrange multiplier mean? Let $f(x, y)$ maximize to M at (x_0, y_0) subject to $g(x, y) = c$, then the rate of M with respect to c is

$$\frac{\partial M}{\partial c} = \lambda_0$$

It's a very surprising result. Let's prove it. Since we want to vary c, then $x_0(c)$, $y_0(c)$ and $\lambda_0(c)$ are all functions of c. Since the Lagrange function is defined as

$$\mathcal{L}(x, y, \lambda) = f(x, y) - \lambda(g(x, y) - c)$$

it is also a function of c. Then at the extreme point, we write it as

$$\mathcal{L}(x_0(c), y_0(c), \lambda_0(c), c) = f(x_0(c), y_0(c)) - \lambda_0(c)(g(x_0(c), y_0(c)) - c)$$

On one aspect, $g(x_0(c), y_0(c)) - c = 0$ and $f(x_0(c), y_0(c)) = M$, then

$$\mathcal{L}(x_0(c), y_0(c), \lambda_0(c), c) = M$$

On another aspect, we have let

$$\mathcal{L}^*(c) = \mathcal{L}(x_0(c), y_0(c), \lambda_0(c), c)$$

as a single variable function. Then by chain rule, we have

$$\frac{d\mathcal{L}^*}{dc} = \frac{\partial \mathcal{L}}{\partial y_0}\frac{dy_0}{dc} + \frac{\partial \mathcal{L}}{\partial \lambda_0}\frac{d\lambda_0}{dc} + \frac{\partial \mathcal{L}}{\partial x_0}\frac{dx_0}{dc} + \frac{\partial \mathcal{L}}{\partial c}\frac{dc}{dc}$$

Since $\nabla \mathcal{L} = \vec{0}$, we have $\dfrac{\partial \mathcal{L}}{\partial x_0} = 0$, $\dfrac{\partial \mathcal{L}}{\partial y_0} = 0$, $\dfrac{\partial \mathcal{L}}{\partial \lambda_0} = 0$. Also, $\dfrac{dc}{dc} = 1$. Therefore, we have

$$\frac{d\mathcal{L}^*}{dc} = \cancel{\frac{\partial \mathcal{L}}{\partial y_0}\frac{dy_0}{dc}} + \cancel{\frac{\partial \mathcal{L}}{\partial \lambda_0}\frac{d\lambda_0}{dc}} + \cancel{\frac{\partial \mathcal{L}}{\partial x_0}\frac{dx_0}{dc}} + \frac{\partial \mathcal{L}}{\partial c}\cancel{\frac{dc}{dc}} = \frac{\partial \mathcal{L}}{\partial c}$$

Also, we have

$$\frac{\partial \mathcal{L}}{\partial c} = \frac{\partial}{\partial c}\left(f(x_0, y_0) - \lambda_0(g(x_0, y_0) - c)\right) = \lambda_0$$

Therefore we have

$$\frac{dM}{dc} = \frac{d\mathcal{L}^*}{dc} = \frac{\partial \mathcal{L}}{\partial c} = \lambda_0$$

(e) **Extreme Values on Compact Domains**

Let's summarize the entire discussion above. We have discussed how to find extreme values on an open domain and how to find extreme values on a constraint curves but we have not discuss the existence of extreme values. The following theorem discuss the existence of extreme value on a *compact*, which means bounded and closed, domain.

> **Theorem 11.** If $\mathscr{D} \subset \mathbb{R}^n$ is compact and f is continuous on $\mathscr{D}$, then f attains global minimum and maximum on $\mathscr{D}$.

Notice that boundedness guarantees that f is bounded and closedness guarantees that the extreme values could be attained on $\mathscr{D}$.

Then let's summarize the procedure of finding extreme values on a compact domain.

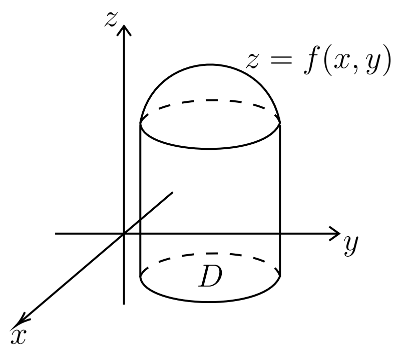
i. Find ∇f.
ii. Find all points where $\nabla f = 0$ or does not exist and check whether they are local extreme.
iii. Find the local extremes of f on the boundary of $\mathscr{D}$ by Lagrange multipliers.
iv. Compare all the local extremes to find out the global extremes.

Integration of Multivariable Functions

1. **Double Integral**

   (a) **Definition**

   As the area under the curve $y = f(x)$, we are also interested what the volume under the surface $z = f(x, y)$.



   Then we are going to imitate the definition of integral of one variable over $[a, b]$ to give the definition of double integral of two variables over $D \subset \mathbb{R}^2$.

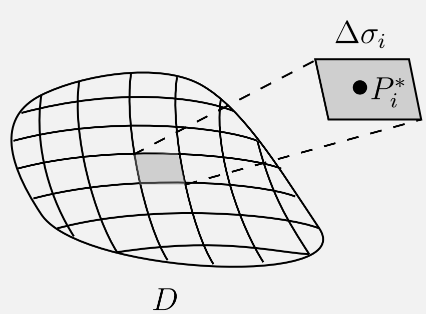
Definition 14. Let D to be closed and bounded and $f(x, y)$ is defined and bounded on D. Then we define the *double integral* in next four steps:

1° Partition the domain D into n subdomains $\Delta\sigma_1, \dots, \Delta\sigma_n$ such that $\Delta\sigma_1 \cup \cdots \cup \Delta\sigma_n = \emptyset$ and $\Delta\sigma_1 \cap \cdots \cap \Delta\sigma_n = D$

2° Pick up all the sample (representative) points $P_i^*(x_i^*, y_i^*) \in \Delta\sigma_i^*$

3° Add up all the small volumes $\displaystyle\sum_{i=1}^{n} f(x_i^*, y_i^*)\Delta\sigma_i$ which is called *Riemann sum*

4° Let Δx_i to be the largest difference in x of $\Delta\sigma_i$ and Δy_i to be the largest difference in y of $\Delta\sigma_i$. Define $\lambda = \max\limits_i \sqrt{\Delta_i^2 + \Delta y_i^2}$. Then take the limit $\displaystyle\lim_{\lambda\to 0} \sum_{i=1}^{n} f(x_i^*, y_i^*)\Delta\sigma_i$.



Then we say f is integrable over the domain D and has *double integral*

$$I = \iint_D f(x, y)\, dA$$

if the limit $\lim_{\lambda \to 0} \sum_{i=1}^{n} f(x_i^*, y_i^*) \Delta\sigma_i = I$ and does not depend on the choice of $\Delta\sigma_i$ and P_i^*.

(b) **Existence:**
Similar to the integral of single variable, if $f(x, y)$ is continuous on D, then $f(x, y)$ is integrable on D. Also, we could tolerate some finite discontinuity to have the function integrable.

Theorem 12. If $f(x, y)$ is bounded on D and $f(x, y)$ is continuous on D except finite numbers of curves of finite length, then $f(x, y)$ is integrable on D.

(c) **Properties**
Double integrals have many very similar properties as the single-variable integral. Let D, D_1, $D_2 \subset \mathbb{R}^2$ and $f(x, y)$, $g(x, y)$ integrable over D.

i. **Zero Domain:** $\iint_D f(x, y)\, dA = 0$ if $D = \emptyset$.

ii. **Domain Composition:** $\iint_{D_1+D_2} f(x, y)\, dA = \iint_{D_1} f(x, y)\, dA + \iint_{D_2} f(x, y)\, dA$ if $D_1 \cap D_2 = \emptyset$.

iii. **Integral Over One:** $\iint_D 1\, dA = A$ where A is the area of D.
Note: In general, the integral over one of a domain is the measure of a domain.

iv. **Linear Combination:** Let L, M be the constants, $\iint_D (Lf(x, y) + Mg(x, y))\, dA = L \iint_D f(x, y)\, dA + M \iint_D g(x, y)\, dA$.

v. **Sign-preserving:** If $f(x, y) \geq 0$ over D, then $\iint_D f(x, y)\, dA \geq 0$.
Corollary: If $f(x, y) \geq g(x, y)$ over D, then $\iint_D f(x, y) dA \geq \iint_D g(x, y)\, dA$.
Corollary: If $|f(x, y)|$ is integrable over D, then $\left| \iint_D f(x, y)\, dA \right| \leq \iint_D |f(x, y)|\, dA$.
Corollary: If $m \leq f(x, y) \leq M$ over D, then $mA \leq \iint_D f(x, y)\, dA \leq MA$ where A is the area of domain D.
Corollary (Mean Value Theorem): If the function $f(x, y)$ is continuous on a

closed, bounded, connected set D in the xy-plane, then there exists a point (x_0, y_0) in D such that $\iint_D f(x,y)\,dA = f(x_0,y_0)A$ where A is the area of D.

vi. **Symmetry**

A. Let D be symmetric along $x=0$. If $f(-x,y) = -f(x,y)$, $\iint_D f(x,y)dA = 0$. If $f(-x,y) = f(x,y)$, $\iint_D f(x,y)dA = 2\iint_{D_1} f(x,y)dA$ where D_1 is right half of D.

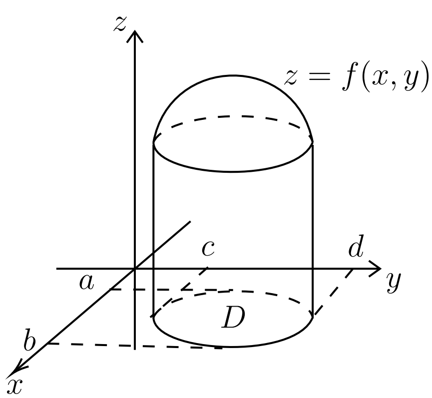
B. Let D be symmetric along $y=0$. If $f(x,-y) = -f(x,y)$, $\iint_D f(x,y)dA = 0$. If $f(x,-y) = f(x,y)$, $\iint_D f(x,y)dA = 2\iint_{D_1} f(x,y)dA$ where D_1 is top half of D.

C. Let D be symmetric along $y = x$, then $\iint_D f(x,y)dA = \iint_D f(y,x)dA$.

(d) **Integral Techniques**

i. **Iteration in Cartesian Coordinates**

Let $f(x,y)$ integrable on D. We want to find out the integral $\iint_D f(x,y)\,dA$.



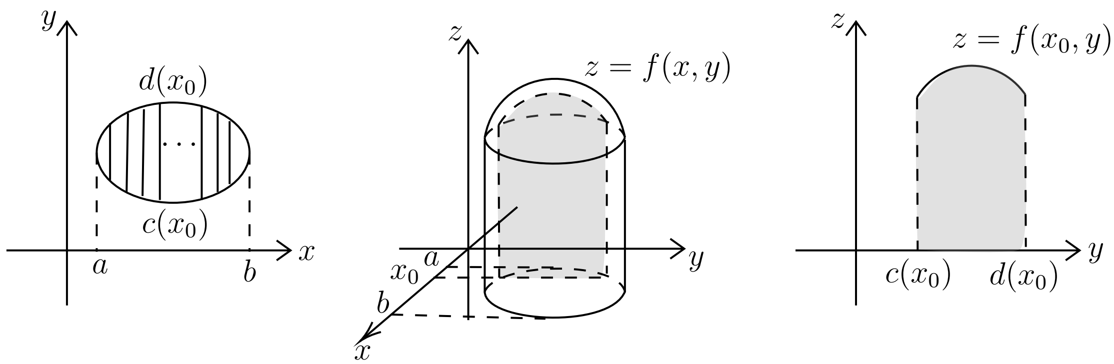
A good way is to reduce the double integral to the integral of single variable. This means we have to reduce the domain D to lines. We could first divide the domain into lines along y direction, integrate along these lines and then integrate along x direction.

Then we want to define domain as

$$D = \{(x,y) | a \le x \le b,\ c(x) \le y \le d(x)\}$$

Pick up $x_0 \in [a,b]$, then cross section under $z = f(x_0, y)$ cut by $x = x_0$ has the the area

$$A(x_0) = \int_{c(x_0)}^{d(x_0)} f(x_0, y)\,dy$$



Then more generally, for any $x \in [a, b]$, the area of the cross section would be

$$A(x) = \int_{c(x)}^{d(x)} f(x, y)\, dy$$

Then the double integral would be the volume

$$V = \int_a^b A(x)\, dx = \int_a^b \left[\int_{c(x)}^{d(x)} f(x, y)\, dy \right] dx$$

which would be

$$\iint_D f(x, y)\, dA = \int_a^b dx \int_{c(x)}^{d(x)} f(x, y)\, dy \tag{X}$$

Similarly, if we define the domain as

$$D = \{(x, y) | c \leq y \leq d, a(y) \leq x \leq b(y)\}$$

then the double integral would be

$$\iint_D f(x, y)\, dA = \int_c^d dy \int_{a(y)}^{d(y)} f(x, y)\, dx \tag{Y}$$

If $f(x, y)$ is integrable, the result evaluated by two integral techniques would be the same and we have alternative notation of the double integral

$$\iint_D f(x, y)\, dA = \iint_D f(x, y)\, dxdy = \iint_D f(x, y)\, dydx$$

Example 22. Evaluate $\iint_D x^2 e^{x^2}\, dx$ where $D = \{(x, y) | 0 \leq x \leq 2, 0 \leq y \leq x\}$. Then the double integral would be

$$\begin{aligned} I &= \int_0^2 dx \int_0^x x^2 e^{x^2}\, dy \\ &= \int_0^2 x^2 e^{x^2} (x - 0)\, dx = \int_0^2 x^3 e^{x^2}\, dx \\ &= \frac{1}{2} \int_0^2 x^2 e^{x^2}\, d(x^2) = \frac{1}{2} \int_0^4 x e^x\, dx \\ &= 3e^4 + 1 \end{aligned}$$

Notice that if we take the domain as $D = \{(x,y)|0 \le x \le 2, 0 \le y \le x\}$, the double integral would be

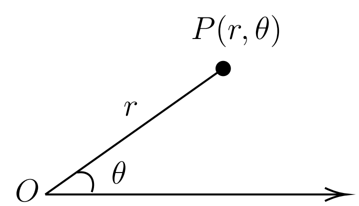
$$\int_0^2 dy \int_y^2 x^2 e^{x^2}\, dx$$

where $\displaystyle\int_y^2 x^2 e^{x^2}\, dx$ does not have analytic result.

ii. **Iteration in Polar Coordinates**

Sometimes the domain or the function we want to integrate has some features of circle, i.e. contains $x^2 + y^2$ in the equations. In this case, polar coordinates would give us a simple representation of the domain or the function, which makes the integration easier.

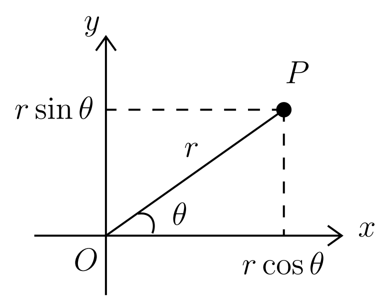
The polar coordinates is defined as (r, θ) where r is the distance between the point P and the original point and the θ is the angle between OP and the axis in positive direction.



The transformation from polar coordinate to Cartesian coordinate is

$$\begin{cases} x = r\cos\theta \\ y = r\sin\theta \end{cases}$$

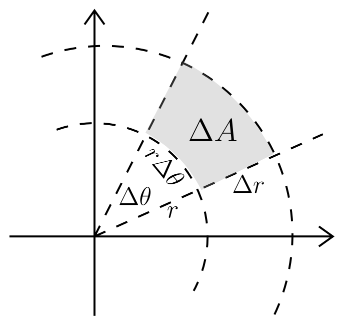
as a projection onto the x-axis and y-axis.



Then the integral of $f(x,y)$ on D would be

$$\iint_{D'} f(r\cos\theta, r\sin\theta)\, dA$$

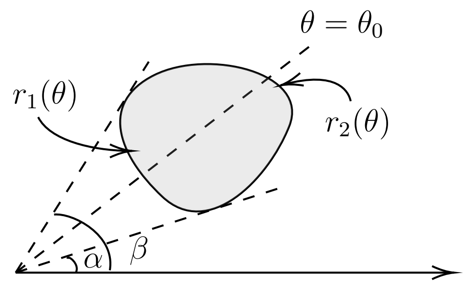
where D' is the representation of domain D in polar coordinates. Then our job is to find out what is dA. A possible guess is $dA = dr d\theta$. However we need some correction scalar since intuitively $dxdy \neq drd\theta$. Consider the finite change Δr and $\Delta\theta$ at (r, θ), we want to find out what is ΔA.



When Δr and $\Delta\theta$ are small, ΔA could be treated as a rectangle with width $r\Delta\theta$ and Δr. Then $\Delta A = r\Delta r\Delta\theta$. Then if we take the infinitely small dr and $d\theta$, we have

$$dA = rdrd\theta$$

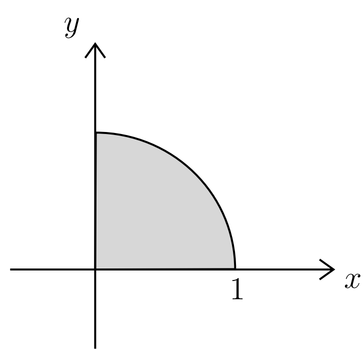
Then if the domain is $D = \{(r,\theta)|\alpha \leq \theta \leq \beta,\, r_1(\theta) \leq r \leq r_2(\theta)\}$



the integral would be

$$\iint_D f(x,y)\,dA = \int_\alpha^\beta d\theta \int_{r_1(\theta)}^{r_2(\theta)} f(r\cos\theta, r\sin\theta) r\,dr\,d\theta$$

Example 23. Evaluate $\displaystyle\iint_D \sin x^2 \cos y^2\,dA$ where $D = \{(x,y)|x^2+y^2 \leq 1, x \geq 0, y \geq 0\}$. Since the domain is symmetric along $y = x$.



We have

$$I = \iint_D \sin x^2 \cos y^2\,dA = \iint_D \sin y^2 \cos x^2\,dA$$

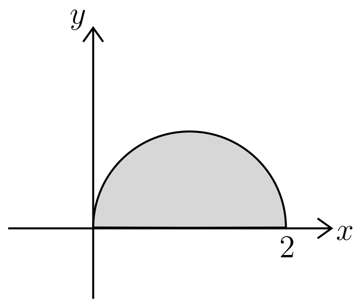
Then we have

$$2I = \iint_D (\sin x^2 \cos y^2 + \sin y^2 \cos x^2)\,dA$$

In polar coordinate, the integral would be

$$\begin{aligned}2I &= \int_0^{\frac{\pi}{2}} d\theta \int_0^1 r\sin r^3 dr = \frac{\pi}{4}\int_0^1 \sin r^2 dr^2 \\ &= -\frac{\pi}{4}\cos r^2\Big|_0^1 = -\frac{\pi}{4}(\cos 1 - 1)\end{aligned}$$

Example 24. Evaluate $\iint_D \sqrt{x^2+y^2}\,dA$ where $D = \{(x,y)|(x-1)^2+y^2 \le 1, y \ge 0\}$. First we want to find out the representation of $(x-1)^2+y^2 = 1$ in polar coordinate.



With expansion $x^2+y^2 = 2x$ and substitution $r^2 = 2r\cos\theta$, we have $r = 2\cos\theta$. Then the integral would be

$$I = \int_0^{\frac{\pi}{2}} d\theta \int_0^{2\cos\theta} r^2\,dr = \frac{8}{3}\int_0^{\frac{\pi}{2}} \cos^3\theta\,d\theta = \frac{16}{9}$$

iii. **General Case**

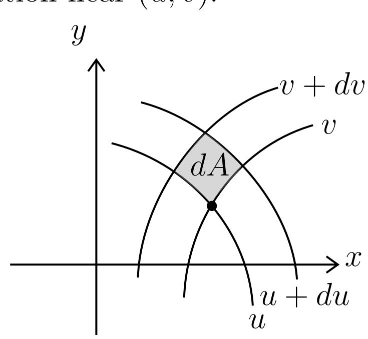
Then we want to generalize the idea of integral in polar coordinate. Consider the coordinate (u,v) such that the transformation

$$\begin{cases} x = x(u,v) \\ y = y(u,v)\end{cases}$$

is a one-to-one relationship between (u,v) and (x,y). With the substitution, we could have

$$\iint_D f(x,y)\,dA = \iint_S f(x(u,v),y(u,v))\,dA$$

where S is the the representation of D in the new coordinate (u,v). We want to apply the linear approximation near (u,v).



The linearization is

$$\begin{pmatrix} dx \\ dy \end{pmatrix} = \begin{pmatrix} x_u & x_v \\ y_u & y_v \end{pmatrix} \begin{pmatrix} du \\ dv \end{pmatrix}$$

By the property of determinant, we have

$$\frac{dxdy}{dudv} = \left\| \begin{matrix} x_u & x_v \\ y_u & y_v \end{matrix} \right\| = \left| \frac{\partial(x,y)}{\partial(u,v)} \right|$$

Therefore the area in the new coordinate (u, v) is

$$dA = \left| \frac{\partial(x,y)}{\partial(u,v)} \right| dudv$$

Up to now we obtain a very important theorem as the change of variables formula for double integrals.

Theorem 13. (Change of Variables Formula for Double Integrals) Let $x = x(u,v)$ and $y = y(u,v)$ be a one-to-one transformation from a domain S in the uv-plane onto a domain D in the xy-plane. Suppose that the functions x and y, and their first partial derivatives with respect to u and v, are continuous in S. If $f(x,y)$ is integrable on D, and if $g(u,v) = f(x(u,v), y(u,v))$, then g is integrable on S and

$$\iint_D f(x,y)dA = \iint_S f(x(u,v), y(u,v)) \left| \frac{\partial(x,y)}{\partial(u,v)} \right| dudv$$

This theorem could be generalized to more variables. With the one-to-one transformation

$$\begin{cases} y_1 = y_1(x_1, \ldots, x_n) \\ \qquad \vdots \\ y_n = y_n(x_1, \ldots, x_n) \end{cases}$$

from domain S in $x_1 \cdots x_n$-plane onto domain D in $y_1 \cdots y_n$-plane, and with the same continuity condition, we have

$$\int_D f(y_1, \ldots, y_n)\, dA = \int_S f(y_1(x_1, \ldots, x_n), \ldots, y_n(x_1, \ldots, x_n)) \left| \frac{\partial(y_1, \cdots, y_n)}{\partial(x_1, \cdots, x_n)} \right| dx_1 \cdots dx_n$$

Example 25. Evaluate the area of the domain $D = \{x \le y \le 4x, 1 \le xy \le 2\}$. We construct a new coordinate such that $\begin{cases} u = xy \\ v = \dfrac{y}{x} \end{cases}$. Then the domain turns out to be

$D = \{(u,v)|1 \le u \le 2, 1 \le v \le 4\}$. And the area is

$$A = \iint_D 1\, dA = \int_1^2 \int_1^4 \left\| \begin{matrix} u_x & u_y \\ v_x & v_y \end{matrix} \right\| dudv$$

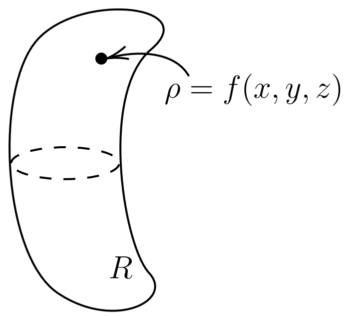
$$= \int_1^2 \int_1^4 \left\| \begin{matrix} y & x \\ -\frac{y}{x^2} & \frac{1}{x} \end{matrix} \right\| dudv = \int_1^2 \int_1^4 \frac{2y}{x}\, dudv$$

$$= \int_1^2 \int_1^4 2v\, dudv = 3 \int_1^2 2v\, dv = 3v^2\Big|_1^2 = 9$$

2. **Triple Integral**

(a) **Definition**

Given an object on the domain $R \subset \mathbb{R}^3$ and the density function $\rho = f(x,y,z)$. We want to find out the mass of the object.



First we divide the R into many pieces with volume dV the mass $dm = \rho dV = f(x,y,z)dV$. Then we add them up as the mass

$$M = \iiint_R f(x,y,z)\, dV$$

We want to define it as the triple integral

Definition 15. Let R to be closed and bounded and $f(x,y,z)$ is defined and bounded on R. Then we define the *triple integral* in next four steps:

1° Partition the domain R into n subdomains $\Delta V_1, \dots, \Delta V_n$ such that $\Delta V_1 \cup \cdots \cup \Delta V_n = \emptyset$ and $\Delta V_1 \cap \cdots \cap \Delta V_n = R$

2° Pick up all the sample (representative) points $P_i^*(x_i^*, y_i^*, z_i^*) \in \Delta V_i^*$

3° Add up all the small volumes $\sum_{i=1}^{n} f(x_i^*, y_i^*, z_i^*)\Delta V_i$ which is called *Riemann sum*

4° Let Δx_i to be the largest difference in x of ΔV_i, Δy_i to be the largest difference in y of ΔV_i, Δz_i to be the largest difference in z of ΔV_i. Define $\lambda = \max_i \sqrt{\Delta_i^2 + \Delta y_i^2 + \Delta z_i^2}$. Then take the limit $\lim_{\lambda \to 0} \sum_{i=1}^{n} f(x_i^*, y_i^*, z_i^*)\Delta V_i$.

Then we say f is integrable over the domain R and has *triple integral*

$$I = \iiint_R f(x,y,z)\,dV$$

if the limit $\lim_{\lambda \to 0} \sum_{i=1}^{n} f(x_i^*, y_i^*, z_i^*)\Delta V_i = I$ and does not depend on the choice of ΔV_i and P_i^*.

Similar to the double integral, if $f(x,y,z)$ is continuous on R then $f(x,y,z)$ is integrable on R.

(b) **Properties**

The the properties of the triple are very similar to the double integral and we are going to highlight three of them.

i. $\iiint_R 1\,dV = V$ where V is the volume of the domain R.

ii. Let $R \subset \mathbb{R}^3$ bounded, closed and connected and $f(x,y,z)$ is continuous on R. Then there exists (x_0, y_0, z_0) such that $\iiint_R f(x,y,z)\,dV = f(x_0,y_0,z_0)V$ where V is the volume of the domain R.

iii. Let R symmetric with respect to plane xOy and the the upper half noted as R_1. Then if $f(x,y,z) = -f(x,y,-z)$, $\iiint_R f(x,y,z)\,dV = 0$; if $f(x,y,z) = f(x,y,-z)$, $\iiint_R f(x,y,z)\,dV = 2\iiint_{R_1} f(x,y,z)\,dV$.

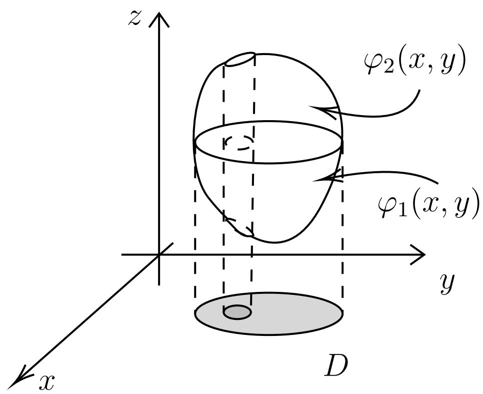
Note: This property is also true for the symmetry with respect to plane xOz and plane yOz.

(c) **Integral Techniques**

i. **Iteration in Cartesian Coordinates**

As what we apply on the double integral, we want to reduce R into lower dimension and then integrate them.

One way is to divide the domain R into many thin rods. We first integrate along those rods and then add then up. We represent the domain R alternatively as $R' = \{(x,y,z) | (x,y) \in D,\ \varphi_1(x,y) \le z \le \varphi_2(x,y)\}$.



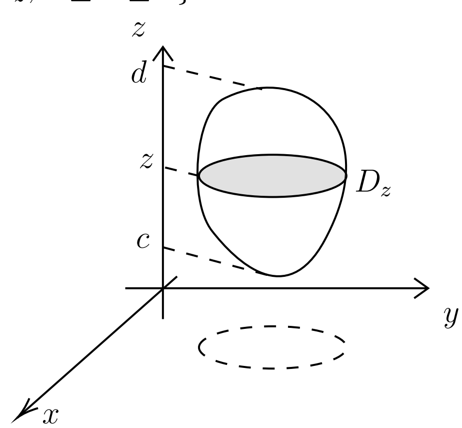
Then we have the triple integral

$$\iiint_R f(x,y,z)\,dV = \iint_{D_{xy}} dxdy \int_{\varphi_1(x,y)}^{\varphi_2(x,y)} f(x,y,z)\,dz \tag{A}$$

If the domain D or the function $f(x,y,z)$ has x^2+y^2 term, we could use polar coordinate on D to get the double integral. Let $R = \{(r,\theta,z)|\alpha \le \theta \le \beta,\ r_1(\theta) \le r \le r_2(\theta),\ \varphi_1(r\cos\theta, r\sin\theta) \le z \le \varphi_2(r\cos\theta, r\sin\theta)\}$. Then the double integral would be

$$\iiint_R f(x,y,z)\,dV = \int_\alpha^\beta d\theta \int_{r_1(\theta)}^{r_2(\theta)} r\,dr \int_{\varphi_1(r\cos\theta,r\sin\theta)}^{\varphi_2(r\cos\theta,r\sin\theta)} f(r\cos\theta, r\sin\theta, z)\,dz \tag{A'}$$

Another way is to divide the domain into many thin slices. We first integrate along the slices and then add them up. Therefore, we want to represent the domain as $R = \{(x,y,z)|(x,y) \in D_z,\ c \le z \le d\}$.



Then the triple integral is

$$\iiint_R f(x,y,z)\,dV = \int_c^d dz \iint_{D_z} f(x,y,z)\,dxdy \tag{B}$$

If the domain D_z or the function $f(x, y, z)$ has $x^2 + y^2$ term, we could use polar coordinate on D_z to get the double integral. Let R defined alternatively as $R' = \{(r, \theta, z)|c \le z \le d, \alpha_z \le \theta \le \beta_z,\, r_{1z}(\theta) \le r \le r_{2z}(\theta)\}$. The the triple integral is

$$\iiint_R f(x, y, z)\, dz = \int_c^d dz \int_{\alpha_z}^{\beta_z} d\theta \int_{r_{1z}(\theta)}^{r_{2z}(\theta)} f(r\cos\theta, r\sin\theta, z) r\, dr \qquad \text{(B')}$$

Example 26. Evaluate $\iiint_R z\, dV$ where R is the region above $z = x^2 + y^2$ and below $z = 1$.

Method A'. The domain is represented as $\{(x, y, z)|(x, y) \in D, x^2 + y^2 \le z \le 1\}$ where $D = \{(x, y)|x^2 + y^2 \le 1\}$. Then the integral is

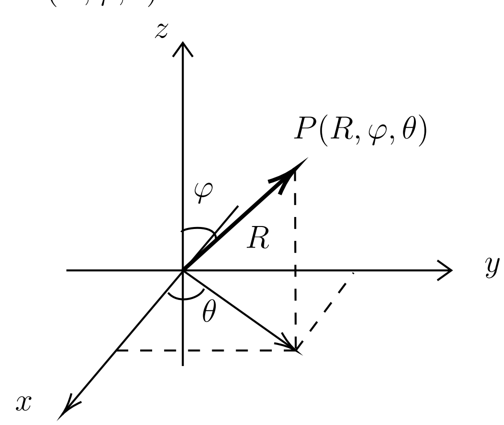
$$\begin{aligned} I &= \iint_D dxdy \int_{x^2+y^2}^{1} z\, dz = \frac{1}{2}\iint_D [1 - (x^2 + y^2)^2] dxdy \\ &= \frac{1}{2}\int_0^{2\pi} d\theta \int_0^1 (1 - r^4) r\, dr = \pi \int_0^1 (r - r^5)\, dr = \frac{1}{3}\pi \end{aligned}$$

Method B. We represent the domain $R = \{(x, y, z)|(x, y) \in D_z,\, 0 \le z \le 1\}$ where $D_z = \{(x, y)|0 \le x^2 + y^2 \le z\}$. Then the triple integral is

$$I = \int_0^1 z\, dz \iint_{D_z} dxdy = \pi \int_0^1 z^2\, dz = \frac{\pi}{3}$$

ii. **Iteration in Spherical Coordinates**
When the domain or the function has the $x^2 + y^2 + z^2$ term, spherical coordinates would be useful to simplify the representation and to make the integral easier.

The spherical coordinate (R, φ, θ) is defined as below.



And the transformation from (R, φ, θ) to (x, y, z) is

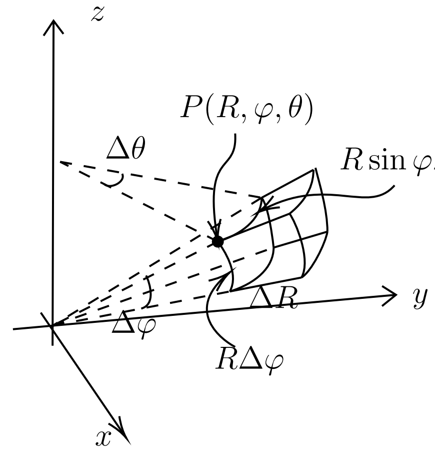
$$\begin{cases} x = R\sin\varphi\cos\theta \\ y = R\sin\varphi\sin\theta \\ z = R\cos\varphi \end{cases}$$

With the substitution, we have

$$\iiint_R f(x,y,z)\,dV = \iiint_{R'} f(R\sin\varphi\cos\theta, R\sin\varphi\sin\theta, R\cos\varphi)\,dV$$

where R' is the domain R in spherical coordinates. Our job is to find out dV.

One way is to find out the small volume element ΔV under the small changes ΔR, $\Delta\varphi$ and $\Delta\theta$.



The volume element could be treated as a small cuboid and therefore $\Delta V = (R\sin\varphi\Delta\theta)(\Delta R)(R\Delta\varphi) = R^2\sin\varphi\Delta R\Delta\theta\Delta\varphi$. Take the limit, we could have

$$dV = R^2\sin\varphi\,dRd\varphi d\theta$$

Alternatively, we could use the change variable formula for general case

$$\left|\frac{\partial(x,y,z)}{\partial(R,\theta,\varphi)}\right| = \left\|\begin{matrix} \sin\varphi\cos\theta & R\cos\varphi\cos\theta & -R\sin\varphi\cos\theta \\ \sin\varphi\sin\theta & R\cos\varphi\sin\theta & R\sin\varphi\sin\theta \\ \cos\varphi & -R\sin\varphi & 0 \end{matrix}\right\| = R^2\sin\varphi$$

Therefore, we have

$$dV = \left|\frac{\partial(x,y,z)}{\partial(R,\theta,\varphi)}\right| dRd\varphi d\theta = R^2\sin\varphi\,dRd\varphi d\theta$$

Then the triple integral finally becomes

$$\iiint_R f(x,y,z)dV = \iiint_{R'} f(R\sin\varphi\cos\theta, R\sin\varphi\sin\theta, R\cos\varphi)R^2\sin\varphi\,dRd\varphi d\theta$$

where R' is the representation of R in spherical coordinates.

Example 27. Evaluate $\displaystyle\iiint_R \sqrt{x^2+y^2}\,dV$ where R is the top half of the ball centered at the original with radius $r=2$. The R is defined as $R = \{(R,\varphi,\theta)|0 \le$

$\theta \leq 2\pi,\ 0 \leq \varphi \leq \frac{\pi}{2}, 0 \leq r \leq 2\}$. Then the triple integral is

$$\begin{aligned} I &= \int_0^{2\pi} d\theta \int_0^{\frac{\pi}{2}} d\varphi \int_0^2 R^2 \sin^2\varphi \, dR \\ &= 2\pi \left(\int_0^{\frac{\pi}{2}} \sin^2\varphi \, d\varphi \right) \left(\int_0^2 R^3 \, dR \right) \\ &= 2\pi \times \frac{1}{2} \times \frac{\pi}{2} \times \frac{1}{4} \times 2^4 = 2\pi^2 \end{aligned}$$

3. **Application of Multiple Integrals**
   Multiple integrals have a wide application. We are going to highlight two of them in physics: center of mass and moment of inertia.

   The center of mass is defined as

$$\bar{x} = \frac{\iint_D x\rho(x,y)\, dA}{\iint_D \rho(x,y)\, dA},\ \bar{y} = \frac{\iint_D y\rho(x,y)\, dA}{\iint_D \rho(x,y)\, dA}$$

   in two dimension and as

$$\bar{x} = \frac{\iiint_R x\rho(x,y,z)\, dV}{\iiint_R \rho(x,y,z)\, dV},\ \bar{y} = \frac{\iiint_R y\rho(x,y,z)\, dV}{\iiint_R \rho(x,y,z)\, dV},\ \bar{z} = \frac{\iiint_R z\rho(x,y,z)\, dV}{\iiint_R \rho(x,y,z)\, dV}$$

   in three dimension. Notice the denominator is just the mass.

   The moment of inertia in three dimension is defined as

$$I_\ell = \iiint_R d^2 \rho(x,y,z)\, dV$$

   where ℓ is the rotation axis. In particular, we have

$$I_x = \iiint_D (y^2+z^2)\rho(x,y,z)\, dV,\ I_y = \iiint_D (x^2+z^2)\rho(x,y,z)\, dV,\ I_z = \iiint_D (x^2+y^2)\rho(x,y,z)\, dV$$

4. **Integrals of higher multiplicity**
   Let $f(x_1, \cdots, x_n)$ be a function on $D \subset \mathbb{R}^n$. We could still have integral of $f(x_1, \cdots, x_n)$ on D and we could denote it as

$$\int_D f(x_1, \cdots, x_n) dV = \int_{\varphi_n}^{\phi_n} \cdots \int_{\varphi_1}^{\phi_1} f(x_1, \cdots, x_n)\, dx_1 \cdots dx_n$$