MATH 2010 Chapter 1

In one-variable calculus course, you study functions $f(x)$ with both the “input" variable x and “output" value $f(x)$ are real numbers. In this course, we will look at more general functions where the input or output may consist of a tuple of numbers. For example, the function

$$f(x, y, z) = (xy - \cos z, x^2 - y + z)$$

maps the tuple $(2, 1, 0)$ to the tuple $f(2, 1, 0) = (1, 3)$. Tuples like this are called vectors. Here x, y, z are the variables. We say that f is vector-valued multi-variable function.

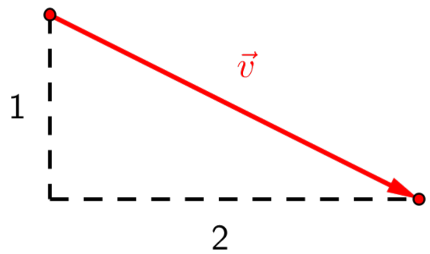
In this chapter, we will discuss vectors and some of its basic properties.

1.1 Euclidean Space

Let $\mathbb{R}$ be the set of real numbers and n be a positive integer. Consider the set

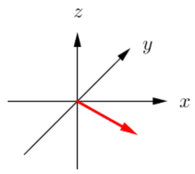
$$\begin{aligned} \mathbb{R}^n &= \mathbb{R} \times \mathbb{R} \times \ldots \times \mathbb{R} \ (n \text{ copies of } \mathbb{R}) \\ &= \{(x_1, x_2, \ldots, x_n) : x_i \in \mathbb{R} \text{ for } 1 \leq i \leq n\} \end{aligned}$$

The set $\mathbb{R}^n$ is called the n-dimensional Euclidean space. Its elements are called n-dimensional vectors or simply vectors. A vector is often written in bold ($\mathbf{v}$), or with an arrow on top ($\vec{v}$). It can be geometrically represented by an arrow. For example, the vector $\vec{v} = (2, -1) \in \mathbb{R}^2$ can be denoted by an arrow that goes to the right by 2 units and goes up by -1 unit, i.e., down by 1 unit, on the plane.



Below are two vectors in $\mathbb{R}$ and $\mathbb{R}^3$. It is more difficult to visualize n-dimensional vectors when $n \geq 4$.

(2) in 1D



$(1, 3, -4)$ in 3D

$$n \in \mathbb{N}, \quad \mathbb{R}^n = \left\{ \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \middle| \, x_i \in \mathbb{R} \right\}$$

- Each $(x_1, x_2, \ldots, x_n) \in \mathbb{R}^n$ may be viewed as a point or a **vector** in $\mathbb{R}^n$.
- A vector in $\mathbb{R}^n$ is typically denoted by a symbol of the form $\vec{v}$.
- If A and B are points in $\mathbb{R}^n$, then the vector with initial point A and terminal point B is often written as $\overrightarrow{AB}$.
- The vector whose entries are all zero is called the **zero vector**. We denote it by $\vec{0}$.

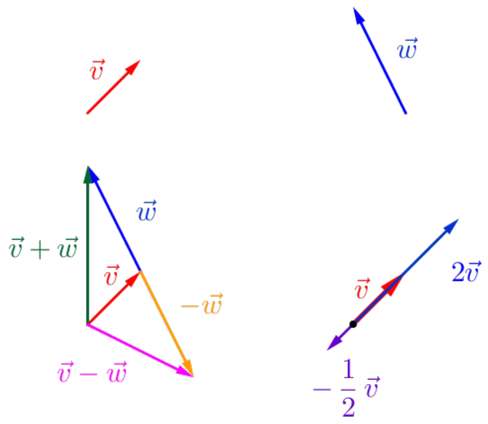
1.2 Basic operations of vectors

Let $\vec{v} = (v_1, v_2, \ldots, v_n), \vec{w} = (w_1, w_2, \ldots, w_n) \in \mathbb{R}^n$ and $r \in \mathbb{R}$. Define

- **Addition Law** $\vec{v} + \vec{w} = (v_1 + w_1, v_2 + w_2, \cdots, v_n + w_n)$
- **Scalar Multiplication** $r\vec{v} = (rv_1, rv_2 \cdots, rv_n)$
- **Subtraction** $\vec{v} - \vec{w} = v + (-1)\vec{w} = (v_1 - w_1, v_2 - w_2, \cdots, v_n - w_n)$

1.2.1 Geometric Interpretation of Vector Algebra

The algebraic operations defined on vectors can be represented graphically:



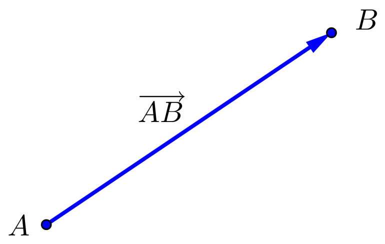
IFRAME

Similar to numbers, there is also a zero vector $\vec{0} = (0, 0, \ldots, 0) \in \mathbb{R}^n$ in each dimension n. The zero vectors and the basic operations above have many properties similar to those of numbers.

Proposition 1.1. *Let $\vec{u}, \vec{v}, \vec{w}$ be vectors, $\alpha, \beta \in \mathbb{R}$.*

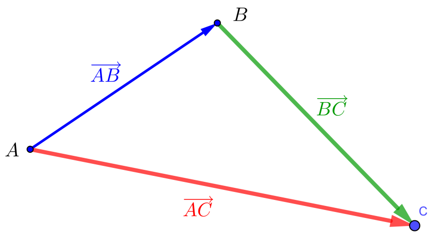
1. $0\vec{v} = \vec{0}$
2. $1\vec{v} = \vec{v}$
3. **Associativity** $(\vec{u} + \vec{v}) + \vec{w} = \vec{u} + (\vec{v} + \vec{w})$
4. **Commutativity** $\vec{v} + \vec{w} = \vec{w} + \vec{v}$
5. $\vec{v} + \vec{0} = \vec{v}$
6. **Distributivity** $(\alpha + \beta)\vec{v} = \alpha\vec{v} + \beta\vec{v}$
7. **Distributivity** $\alpha(\vec{v} + \vec{w}) = \alpha\vec{v} + \alpha\vec{w}$
8. $(\alpha\beta)\vec{v} = \alpha(\beta\vec{v})$

Given two points A and B in $\mathbb{R}^n$. An arrow can be drawn from A to B. It represents a vector which is denoted by $\overrightarrow{AB}$. The point A is called the **initial point** or the **tail** while B is called the **end point** or the **head**.



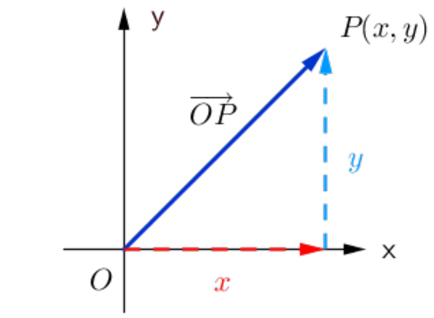
It is clear from the definitions and also the geometric properties that

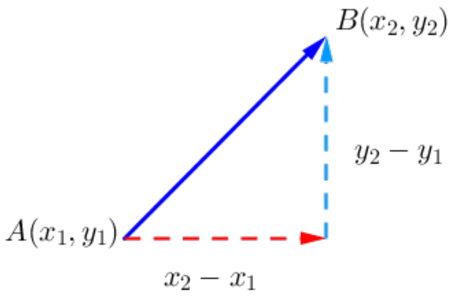
- $\overrightarrow{AB} + \overrightarrow{BC} = \overrightarrow{AC}$
- $\overrightarrow{BA} = -\overrightarrow{AB}$



A **position vector** is a vector with initial point at the origin. If P has coordinates $(x_1, x_2, \ldots, x_n)$, the position vector is also given by $\overrightarrow{OP} = (x_1, x_2, \ldots, x_n)$.

More generally, the initial point of a vector may not be the origin. For example, consider the vector from $A = (x_1, y_1)$ to $B = (x_2, y_2)$. To move from the initial point to the terminal point, the vector goes to the right by $x_2 - x_1$ and up by $y_2 - y_1$. Hence, $\overrightarrow{AB} = (x_2 - x_1, y_2 - y_1)$.





More generally, the vector from $A = (a_1, a_2, \dots, a_n)$ to $B = (b_1, b_2, \dots, b_n)$ in $\mathbb{R}^n$ is

$$\overrightarrow{AB} = (b_1 - a_1, b_2 - a_2, \dots, b_n - a_n).$$

This formula can also be obtained by considering $\overrightarrow{AB}$ as a difference of position vectors:

$$\begin{aligned} \overrightarrow{AB} &= \overrightarrow{AO} + \overrightarrow{OB} \\ &= -\overrightarrow{OA} + \overrightarrow{OB} \\ &= -(a_1, a_2, \dots, a_n) + (b_1, b_2, \dots, b_n) \\ &= (b_1 - a_1, b_2 - a_2, \dots, b_n - a_n). \end{aligned}$$

Remark. Besides vectors, an element $x \in \mathbb{R}^n$ can be viewed as a point in the Euclidean space. If we want to describe a location, it is more convenient to think about x as a point. If we want to describe a quantity with both length and direction (e.g. the displacement from one point to another), it is better to think about x as a vector. Some people use notations like $\langle x, y, z \rangle$ for vectors and (x, y, z) for points. We will not follow this convention and write (x, y, z) for both vectors and points.

Example 1.2. Let $A = (1, 0), B = (3, 3), C = (2, 4), D = (0, 1)$ be points on the plane. Show that $ABCD$ is a parallelogram.

Solution.

$$\overrightarrow{AB} = (3,3) - (1,0) = (2,3)$$
$$\overrightarrow{DC} = (2,4) - (0,1) = (2,3) = \overrightarrow{AB}$$

Hence, $ABCD$ is a parallelogram.

Remark. $\overrightarrow{AB}$ and $\overrightarrow{DC}$ are considered equal because they have the same magnitude and direction even though their initial points are different.

1.3 Length and Dot Product

Definition 1.3. The **norm** (or **length**, or **magnitude**) of a vector $\vec{v} = \langle v_1, v_2, \ldots, v_n \rangle \in \mathbb{R}^n$ is:

$$\|\vec{v}\| = \sqrt{\sum_{i=1}^{n} v_i^2} = \sqrt{v_1^2 + v_2^2 + \cdots v_n^2}.$$

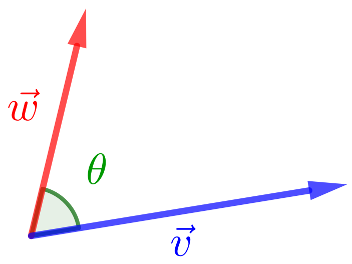
Definition 1.4. The **dot product** of two vectors $\vec{v}, \vec{w} \in \mathbb{R}^n$ is:

$$\vec{v} \cdot \vec{w} = \vec{w} \cdot \vec{v} = \sum_{i=1}^{n} v_i w_i.$$

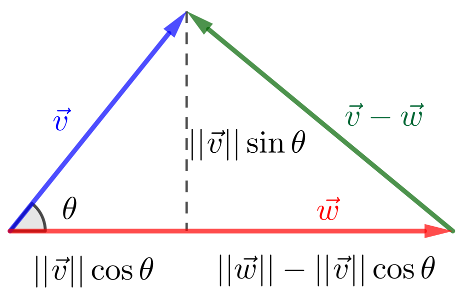
Proposition 1.5. *Let $\vec{u}, \vec{v}, \vec{w} \in \mathbb{R}^n$, $r \in \mathbb{R}$. Then:*

1. $(\vec{u} + \vec{v}) \cdot \vec{w} = \vec{u} \cdot \vec{w} + \vec{v} \cdot \vec{w}$ *and* $\vec{u} \cdot (\vec{v} + \vec{w}) = \vec{u} \cdot \vec{v} + \vec{u} \cdot \vec{w}$.
2. $(r\vec{v}) \cdot \vec{w} = \vec{v} \cdot (r\vec{w}) = r(\vec{v} \cdot \vec{w})$
3. $\vec{v} \cdot \vec{w} = \vec{w} \cdot \vec{v}$
4. $\|r\vec{v}\| = |r|\|\vec{v}\|$
5. $\vec{v} \cdot \vec{v} = \|\vec{v}\|^2$
6. $\vec{v} \cdot \vec{v} \geqslant 0$ *with equality* $\vec{v} \cdot \vec{v} = 0$ *occurs if and only if* $\vec{v} = \overrightarrow{0}$.
7. $\vec{v} \cdot \vec{w} = \|\vec{v}\|\|\vec{w}\| \cos\theta$ *where θ is the angle between $\vec{v}$ and $\vec{w}$. Hence, if $\vec{v}, \vec{w} \neq \vec{0}$, then:*

$$\vec{v} \cdot \vec{w} = 0 \iff \cos\theta = 0 \iff \vec{v} \perp \vec{w}$$



Proof of Proposition 1.5. We will prove property 7 for the case $n \leq 3$. The proof is essentially the same as that of cosine law. Assume $\theta < \dfrac{\pi}{2}$. Consider the following triangle.



Note:

$$\begin{aligned}
||\vec{v}-\vec{w}||^2 &= (||\vec{v}|| \sin\theta)^2 + (||\vec{w}|| - ||\vec{v}|| \cos\theta)^2 \\
&= ||\vec{v}||^2 \sin^2\theta + ||\vec{w}||^2 - 2||\vec{w}||\,||\vec{v}|| \cos\theta + ||\vec{v}||^2 \cos^2\theta \\
&= ||\vec{v}||^2 + ||\vec{w}||^2 - 2||\vec{w}||\,||\vec{v}|| \cos\theta \qquad (1)
\end{aligned}$$

Also,

$$\begin{aligned}
||\vec{v}-\vec{w}||^2 &= (\vec{v}-\vec{w}) \cdot (\vec{v}-\vec{w}) \\
&= \vec{v}\cdot\vec{v} - \vec{w}\cdot\vec{v} - \vec{v}\cdot\vec{w} + \vec{w}\cdot\vec{w} \\
&= ||\vec{v}||^2 + ||\vec{w}||^2 - 2\vec{v}\cdot\vec{w} \qquad (2)
\end{aligned}$$

Compare (1) and (2), we have

$$\vec{v}\cdot\vec{w} = ||\vec{v}||\,||\vec{w}|| \cos\theta.$$

The proof for the cases $\theta \geq \dfrac{\pi}{2}$ can be done similarly. □

Remark. Properties 5 and 7 are geometric properties of length and angle in $\mathbb{R}^n$ for $n \leq 3$. They are used for defining length and angle in higher dimension $n \geq 4$.

A vector of length 1 is called a **unit vector**.

For $\vec{v} \neq \vec{0}$, the vector $\dfrac{1}{\|\vec{v}\|}\vec{v}$ has length:

$$\left|\frac{1}{\|\vec{v}\|}\vec{v}\right| = \frac{1}{\|\vec{v}\|}\|\vec{v}\| = 1.$$

We call $\dfrac{1}{\|\vec{v}\|}\vec{v}$ the **unit vector associated with** $\vec{v}$. It captures the direction of $\vec{v}$.

Every nonzero vector $\vec{v}$ has the form:

$$\vec{v} = \lambda\vec{u}, \quad \lambda > 0,$$

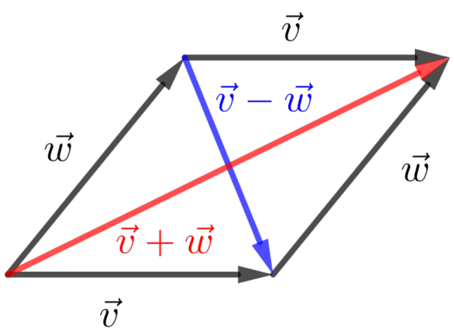
where $\vec{u} = \frac{1}{\|\vec{v}\|}\vec{v}$ is the unit vector associated with $\vec{v}$, and $\lambda = \|\vec{v}\|$ is the length of $\vec{v}$.

Example 1.6. Let $\vec{v}$, $\vec{w}$ have the same length. Show that $(\vec{v}+\vec{w})\cdot(\vec{v}-\vec{w}) = 0$.

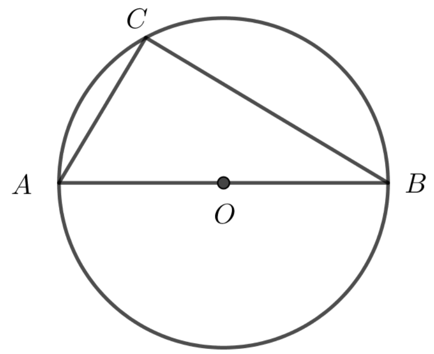
Solution.

$$\begin{aligned}(\vec{v}+\vec{w})\cdot(\vec{v}-\vec{w}) &= \vec{v}\cdot\vec{v} - \vec{v}\cdot\vec{w} + \vec{w}\cdot\vec{v} - \vec{w}\cdot\vec{w} \\ &= \|v\|^2 - \vec{v}\cdot\vec{w} + \vec{v}\cdot\vec{w} - \|w\|^2 \\ &= \|v\|^2 - \|v\|^2 \\ &= 0\end{aligned}$$

Remark. The assumption $||\vec{v}|| = ||\vec{w}||$ means that the parallelogram spanned by $\vec{v}$ and $\vec{w}$ is a rhombus. The computation above shows the fact that the diagonals of a rhombus are perpendicular.



Example 1.7. Consider a circle centered at O. AB is diameter. Show that $\angle ACB$ is a right angle.



Solution.

$$\begin{aligned}
\overrightarrow{AC} &= \overrightarrow{AO} + \overrightarrow{OC} \\
\overrightarrow{BC} &= \overrightarrow{BO} + \overrightarrow{OC} = -\overrightarrow{AO} + \overrightarrow{OC} \\
\overrightarrow{AC} \cdot \overrightarrow{BC} &= (\overrightarrow{AO} + \overrightarrow{OC}) \cdot (-\overrightarrow{AO} + \overrightarrow{OC}) \\
&= -\overrightarrow{AO} \cdot \overrightarrow{AO} + \overrightarrow{AO} \cdot \overrightarrow{OC} - \overrightarrow{OC} \cdot \overrightarrow{AO} + \overrightarrow{OC} \cdot \overrightarrow{OC} \\
&= -||\overrightarrow{AO}||^2 + ||\overrightarrow{OC}||^2 \qquad (||\overrightarrow{AO}|| = ||\overrightarrow{OC}|| \text{ are radius}) \\
&= 0
\end{aligned}$$

Therefore, $\overrightarrow{AC} \perp \overrightarrow{BC}$. Hence, $\angle ACB$ is a right angle.

Theorem 1.8 (Cauchy-Schwarz Inequality)**.** *For all $\vec{a}, \vec{b} \in \mathbb{R}^n$, the following inequality holds:*

$$\|\vec{a} \cdot \vec{b}\| \leq \|\vec{a}\|\|\vec{b}\|.$$

Remark. For lower dimensional spaces like $\mathbb{R}^2$ and $\mathbb{R}^3$, the inequality follows from the Law of Cosine, since the cosine function has absolute value at most 1.

For $n > 3$, it's not as easy to visualize the situation. We prove the inequaltiy as follows:

Proof of Cauchy-Schwarz Inequality. Observe that for all $t \in \mathbb{R}$, we have:

$$0 \leq \|\vec{a} - t\vec{b}\|^2 = (\vec{a} - t\vec{b}) \cdot (\vec{a} - t\vec{b}) = \|\vec{a}\|^2 - 2(\vec{a} \cdot \vec{b})t + t^2\|\vec{b}\|^2$$

In other words, $||\vec{a}||^2$, $-2\vec{a}\cdot\vec{b}$ and $||\vec{b}||^2$ are coefficients of a quadratic function which is always non-negative.

The discriminant of such a quadratic function must be non-positive. Hence:

$$(-2(\vec{a}\cdot\vec{b}))^2 - 4||\vec{a}||^2||\vec{b}||^2 \leq 0$$

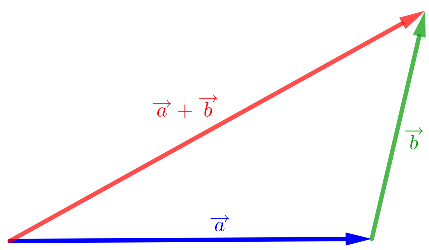
which implies that:

$$||\vec{a}\cdot\vec{b}|| \leq ||\vec{a}||\,||\vec{b}||$$

$\square$

Theorem 1.9 (Triangle Inequality)**.** *For any $\vec{a}, \vec{b} \in \mathbb{R}^n$, we have:*

$$||\vec{a}+\vec{b}|| \leq ||\vec{a}|| + ||\vec{b}||.$$



Proof of Triangle Inequality.

$$||\vec{a}+\vec{b}||^2 = \vec{a}\cdot\vec{a} + \vec{a}\cdot\vec{b} + \vec{b}\cdot\vec{a} + \vec{b}\cdot\vec{b} = ||\vec{a}||^2 + 2\vec{a}\cdot\vec{b} + ||\vec{b}||^2.$$

By the Cauchy-Schwarz inequality

$$|\vec{a}\cdot\vec{b}| \leq ||\vec{a}||\,||\vec{b}||,$$

thus

$$||\vec{a}+\vec{b}||^2 \leq ||\vec{a}||^2 + 2||\vec{a}||\,||\vec{b}|| + ||\vec{b}||^2 = (||\vec{a}|| + ||\vec{b}||)^2.$$

The result follows by taking square roots on both sides. $\square$

1.4 Cross Product

Besides dot product, there is another type of product, called cross product, for vectors in $\mathbb{R}^3$. It can be defined using determinant. Recall the following formulas for 2×2 and 3×3 determinants.

$$\begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc$$

$$\begin{vmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{vmatrix} = a_1 \begin{vmatrix} b_2 & b_3 \\ c_2 & c_3 \end{vmatrix} - a_2 \begin{vmatrix} b_1 & b_3 \\ c_1 & c_3 \end{vmatrix} + a_3 \begin{vmatrix} b_1 & b_2 \\ c_1 & c_2 \end{vmatrix}$$

Example 1.10.

$$\begin{vmatrix} 1 & 2 \\ 3 & 4 \end{vmatrix} = (1)(4) - (2)(3) = -2$$

$$\begin{aligned} \begin{vmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{vmatrix} &= (1) \begin{vmatrix} 5 & 6 \\ 8 & 9 \end{vmatrix} - (2) \begin{vmatrix} 4 & 6 \\ 7 & 9 \end{vmatrix} + (3) \begin{vmatrix} 4 & 5 \\ 7 & 8 \end{vmatrix} \\ &= (1)[(5)(9) - (6)(8)] - (2)[(4)(9) - (6)(7)] + (3)[(4)(8) - (5)(7)] \\ &= -3 + 12 - 9 \\ &= 0 \end{aligned}$$

Definition 1.11 (Cross product)**.** Let $\vec{a} = (a_1, a_2, a_3)$ and $\vec{b} = (b_1, b_2, b_3)$. Their cross product is defined to be

$$\begin{aligned} \vec{a} \times \vec{b} &= \begin{vmatrix} \hat{i} & \hat{j} & \hat{k} \\ a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \end{vmatrix} \\ &= \begin{vmatrix} a_2 & a_3 \\ b_2 & b_3 \end{vmatrix} \hat{i} - \begin{vmatrix} a_1 & a_3 \\ b_1 & b_3 \end{vmatrix} \hat{j} + \begin{vmatrix} a_1 & a_2 \\ b_1 & b_2 \end{vmatrix} \hat{k} \\ &= (a_2 b_3 - a_3 b_2, a_3 b_1 - a_1 b_3, a_1 b_2 - a_2 b_1). \end{aligned}$$

Here the vectors $\hat{i} = (1, 0, 0), \hat{j} = (0, 1, 0)$ and $\hat{k} = (0, 0, 1)$ are the standard unit vectors. A hat instead of an arrow is written on top of each of them to mean that they are unit vectors (vectors of length one).

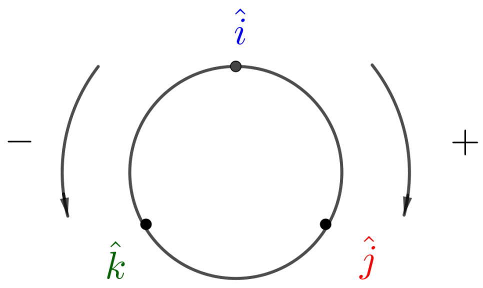
Example 1.12.

$$\begin{aligned} \hat{i} \times \hat{j} &= \begin{vmatrix} \hat{i} & \hat{j} & \hat{k} \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{vmatrix} \\ &= \begin{vmatrix} 0 & 0 \\ 1 & 0 \end{vmatrix} \hat{i} - \begin{vmatrix} 1 & 0 \\ 0 & 0 \end{vmatrix} \hat{j} + \begin{vmatrix} 1 & 0 \\ 0 & 1 \end{vmatrix} \hat{k} \\ &= 0\hat{i} - 0\hat{j} + 1\hat{k} = \hat{k} \end{aligned}$$

Similarly, we can compute the cross products of other standard unit vectors:

$$\begin{array}{lll} \hat{i} \times \hat{i} = \vec{0} & \hat{i} \times \hat{j} = \hat{k} & \hat{i} \times \hat{k} = -\hat{j} \\ \hat{j} \times \hat{i} = -\hat{k} & \hat{j} \times \hat{j} = \vec{0} & \hat{j} \times \hat{k} = \hat{i} \\ \hat{k} \times \hat{i} = \hat{j} & \hat{k} \times \hat{j} = -\hat{i} & \hat{k} \times \hat{k} = \vec{0} \end{array}$$

The diagram below helps you to remember the cross products of standard unit vectors.



Example 1.13. Let $\vec{a} = 2\hat{i} + 3\hat{j} + 5\hat{k}$, $\vec{b} = \hat{i} + 2\hat{j} + 3\hat{k}$. Then

$$\begin{aligned} \vec{a} \times \vec{b} &= \begin{vmatrix} \hat{i} & \hat{j} & \hat{k} \\ 2 & 3 & 5 \\ 1 & 2 & 3 \end{vmatrix} \\ &= \begin{vmatrix} 3 & 5 \\ 2 & 3 \end{vmatrix} \hat{i} - \begin{vmatrix} 2 & 5 \\ 1 & 3 \end{vmatrix} \hat{j} + \begin{vmatrix} 2 & 3 \\ 1 & 2 \end{vmatrix} \hat{k} \\ &= -\hat{i} - \hat{j} + \hat{k} \end{aligned}$$

Find $\vec{b} \times \vec{a}$ and $\vec{b} \times \vec{b}$.

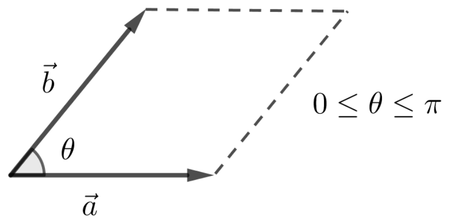
Cross product has the following properties.

Proposition 1.14. *Let $\vec{a}, \vec{b}, \vec{c} \in \mathbb{R}^3$, $\alpha, \beta \in \mathbb{R}$. Then*

1. $\vec{a} \times \vec{a} = \vec{0}$
2. $\vec{a} \times \vec{b} = -\vec{b} \times \vec{a}$
3. $(\alpha\vec{a} + \beta\vec{b}) \times \vec{c} = \alpha\vec{a} \times \vec{c} + \beta\vec{b} \times \vec{c}$
4. *Let θ be the angle between $\vec{a}$, $\vec{b}$.*

   $\|\vec{a} \times \vec{b}\| = \|\vec{a}\| \, \|\vec{b}\| \sin\theta =$ *Area of the parallelogram spanned by $\vec{a}$ and $\vec{b}$.*

5. $(\vec{a} \times \vec{b}) \cdot \vec{a} = (\vec{a} \times \vec{b}) \cdot \vec{b} = 0.$

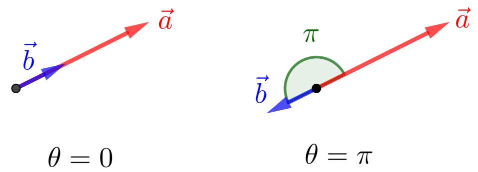


We will prove property 4 below. The other properties can be proved by straightforward computations or properties of determinant.

Remark. From property 4 above,

$$\begin{aligned}\vec{a} \times \vec{b} = \vec{0} \quad &\Leftrightarrow \quad \text{Area of parallelogram} = 0 \\ &\Leftrightarrow \quad \vec{a}, \vec{b} \text{ lie on the same line} \\ &\Leftrightarrow \quad \{\vec{a}, \vec{b}\} \text{ is linearly dependent.}\end{aligned}$$

Hence, two non-zero vectors have zero cross product if and only if they are pointing the same or opposite directions.



Moreover:

- Area of the triangle spanned by $\vec{a}$ and $\vec{b} = \dfrac{1}{2}\|\vec{a} \times \vec{b}\|$.
- If $\vec{c}, \vec{d} \in \mathbb{R}^2$, then

$$\text{Area of the parallelogram spanned by } \vec{c} \text{ and } \vec{d} = \left| \det \begin{bmatrix} c_1 & c_2 \\ d_1 & d_2 \end{bmatrix} \right|$$

Proof of Proposition 1.14. By direct expansion,

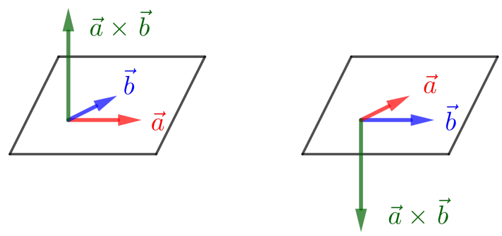
$$\begin{aligned}\|\vec{a} \times \vec{b}\|^2 &= \|\vec{a}\|^2\|\vec{b}\|^2 - (\vec{a} \cdot \vec{b})^2 \\ &= \|\vec{a}\|^2\|\vec{b}\|^2(1 - \cos^2\theta) \\ &= \|\vec{a}\|^2\|\vec{b}\|^2 \sin^2\theta\end{aligned}$$

Since $0 \leq \theta \leq \pi$, we have $\sin\theta \geq 0$ and so

$$\|\vec{a} \times \vec{b}\| = \|\vec{a}\| \, \|\vec{b}\| |\sin\theta| = \|\vec{a}\| \, \|\vec{b}\| \sin\theta$$

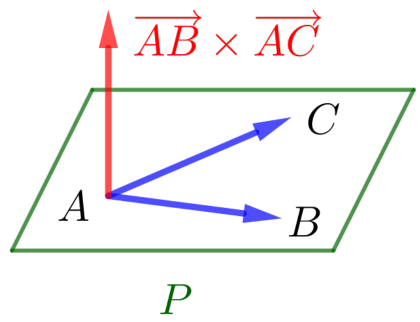
$\square$

Suppose $\vec{a} \times \vec{b}$ are non-zero. Then $\vec{a}$ and $\vec{b}$ are both non-zero. From property 5 above, $\vec{a} \times \vec{b}$ is perpendicular to both $\vec{a}$ and $\vec{b}$. It can be shown that $\vec{a}, \vec{b}, \vec{a} \times \vec{b}$ satisfy the **right hand rule**.



Example 1.15. Let $A = (1, 2, 1)$, $B = (1, -1, 0)$ and $C = (2, 3, 2)$ be points on a plane P. Find a normal vector of P, i.e. a vector perpendicular to P.

Solution. The line segments AB and AC both lie on P. Hence, the cross product $\overrightarrow{AB} \times \overrightarrow{AC}$ is perpendicular to P.



$$
\begin{aligned}
\overrightarrow{AB} &= (1,-1,0)-(1,2,1) = (0,-3,-1)\\
\overrightarrow{AC} &= (2,3,2)-(1,2,1) = (1,1,1)\\
\overrightarrow{AB}\times\overrightarrow{AC} &= \begin{vmatrix} \hat{i} & \hat{j} & \hat{k}\\ 0 & -3 & -1\\ 1 & 1 & 1 \end{vmatrix}\\
&= \begin{vmatrix} -3 & -1\\ 1 & 1 \end{vmatrix}\hat{i} - \begin{vmatrix} 0 & -1\\ 1 & 1 \end{vmatrix}\hat{j} + \begin{vmatrix} 0 & -3\\ 1 & 1 \end{vmatrix}\hat{k}\\
&= [(-3)(1)-(-1)(1)]\,\hat{i} - [(0)(1)-(-1)(1)]\,\hat{j} + [(0)(1)-(-3)(1)]\,\hat{k}\\
&= -2\hat{i} - \hat{j} + 3\hat{k}
\end{aligned}
$$

Therefore, $(-2,-1,3)\perp P$.

Another product closely related to cross product is also defined for vectors in $\mathbb{R}^3$.

Definition 1.16. The **triple product** of $\vec{a},\vec{b},\vec{c}\in\mathbb{R}^3$ is defined to be $\vec{a}\cdot(\vec{b}\times\vec{c})$.

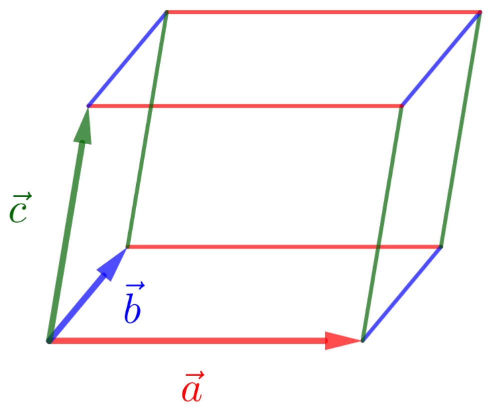
From this definition, it is easy to see that

$$
\begin{aligned}
\vec{a}\cdot(\vec{b}\times\vec{c}) &= (a_1,a_2,a_3)\cdot\begin{vmatrix} \hat{i} & \hat{j} & \hat{k}\\ b_1 & b_2 & b_3\\ c_1 & c_2 & c_3 \end{vmatrix}\\
&= (a_1,a_2,a_3)\cdot\left(\begin{vmatrix} b_2 & b_3\\ c_2 & c_3 \end{vmatrix}, -\begin{vmatrix} b_1 & b_3\\ c_1 & c_3 \end{vmatrix}, \begin{vmatrix} b_1 & b_2\\ c_1 & c_2 \end{vmatrix}\right)\\
&= a_1\begin{vmatrix} b_2 & b_3\\ c_2 & c_3 \end{vmatrix} - a_2\begin{vmatrix} b_1 & b_3\\ c_1 & c_3 \end{vmatrix} + a_3\begin{vmatrix} b_1 & b_2\\ c_1 & c_2 \end{vmatrix}\\
&= \begin{vmatrix} a_1 & a_2 & a_3\\ b_1 & b_2 & b_3\\ c_1 & c_2 & c_3 \end{vmatrix}
\end{aligned}
$$

It follows from this formula that a triple product depends on the order of its factors. From properties of determinant,

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

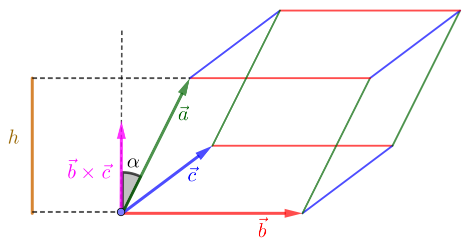
Proposition 1.17. *Three vectors $\vec{a},\vec{b}$ and $\vec{c}$ in $\mathbb{R}^3$ determine a parallelepiped as below.*



Its volume can be computed using triple product:

$$|\vec{a} \cdot (\vec{b} \times \vec{c})| = \textit{Volume of parallelepiped spanned by } \vec{a}, \vec{b}, \vec{c}.$$

Proof of Proposition 1.17. Consider the parallelogram spanned by $\vec{b}$ and $\vec{c}$ as the base of the parallelopiped. Let α be the angle between $\vec{a}$ and $\vec{b} \times c$. Suppose $\alpha \leqslant \pi/2$.



Then:

$$\begin{aligned}
\vec{a} \cdot (\vec{b} \times \vec{c}) &= \|\vec{a}\| \|\vec{b} \times \vec{c}\| \cos \alpha \\
&= \|\vec{b} \times \vec{c}\| h \\
&= \text{Base Area } \times \text{ height} \\
&= \text{Volume of parallelepiped}
\end{aligned}$$

The case for $\pi/2 < \alpha \leqslant \pi$ can be done similarly. In that case,

$$\vec{a} \cdot (\vec{b} \times \vec{c}) = - \text{ Volume of parallelepiped}$$

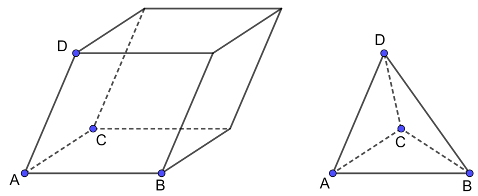
$\square$

Remark.

$$\vec{a} \cdot (\vec{b} \times \vec{c}) = \begin{vmatrix} a_1 & a_1 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{vmatrix} = 0 \iff \text{Volume of parallelepiped} = 0$$

$$\iff \{\vec{a}, \vec{b}, \vec{c}\} \text{ is linearly dependent}$$

Consider a tetrahedron with vertices $A, B, C, D \in \mathbb{R}^3$. To find a formula of its volume, we compare the tetrahedron with the parallelopiped spanned by $\overrightarrow{AB}, \overrightarrow{AC}, \overrightarrow{AD}$.



$$\begin{aligned} \text{Volume of Tetrahedron} &= \frac{1}{3} \cdot \text{Area}(\triangle ABC) \cdot \text{height} \\ &= \frac{1}{3} \cdot \frac{1}{2} \cdot (\text{Area of parallelogram spanned by } \overrightarrow{AB}, \overrightarrow{AC}) \cdot \text{height} \\ &= \frac{1}{6} \cdot \text{Volume of Parallelepiped} \\ &= \frac{1}{6} \left| \overrightarrow{AB} \cdot (\overrightarrow{AC} \times \overrightarrow{AD}) \right| \end{aligned}$$

Example 1.18. Let $A = (1, 0, 1), B = (1, 1, 2), C = (2, 1, 1), D = (2, 1, 3)$. Find the volume of the tetrahedron $ABCD$.

Solution.

$$\begin{aligned} \overrightarrow{AB} &= (1, 1, 2) - (1, 0, 1) = (0, 1, 1) \\ \overrightarrow{AC} &= (2, 1, 1) - (1, 0, 1) = (1, 1, 0) \\ \overrightarrow{AD} &= (2, 1, 3) - (1, 0, 1) = (1, 1, 2) \end{aligned}$$

Their triple product is:

$$(\overrightarrow{AB} \times \overrightarrow{AC}) \cdot \overrightarrow{AD} = \begin{vmatrix} 0 & 1 & 1 \\ 1 & 1 & 0 \\ 1 & 1 & 2 \end{vmatrix} = -2$$

and so:

$$\text{Volume of the tetrahedron } ABCD = \frac{1}{6} \cdot |2| = \frac{1}{3}$$

MATH 2010 Chapter 2

2.1 Linear Objects in $\mathbb{R}^n$

In this section, we will study linear objects in $\mathbb{R}^n$. Typical examples are 1-dimensional lines and 2-dimensional planes. We will also look at their higher dimensional analog.

2.1.1 Line

Consider the line L passing through $A = (1, 0)$ and $B = (0, 2)$ in $\mathbb{R}^2$.

Two standard ways to represent L is

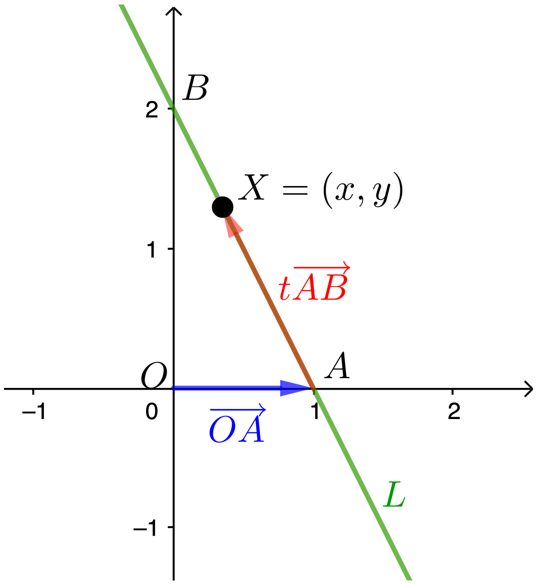
- **Equation form**

$$2x + y = 2$$

- **Parametric form**

$$\begin{aligned}(x, y) &= \overrightarrow{OA} + t\overrightarrow{AB} \\ &= (1, 0) + t(-1, 2) \\ &= (1 - t, 2t)\end{aligned}$$

  Varying $t \in \mathbb{R}$ gives all the points X on L.



- **Symmetric form** The parametric equation above implies

$$\begin{cases} x = 1 - t \\ y = 2t \end{cases} \Rightarrow \begin{cases} t = \dfrac{x-1}{-1} \\ t = \dfrac{y}{2} \end{cases}$$

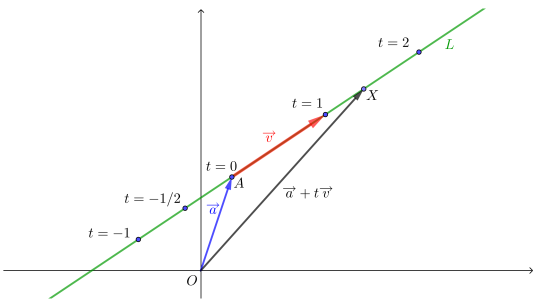
By eliminating t from the parametric form, we obtain another way to represent L. It is called the **symmetric form** :

$$\frac{x-1}{-1} = \frac{y-0}{2}$$

2.1.2 Parametric form of a line in $\mathbb{R}^n$

Let L be a line in $\mathbb{R}^n$. Let A be a point on it with $\vec{a} = \overrightarrow{OA}$ and $\vec{v}$ is a vector representing a direction of L. Then A parametric form of L is given by

$$\vec{x} = \vec{a} + t\vec{v}, \quad t \in \mathbb{R} \text{ is called a parameter}$$



L is said to be parametrized by $t \in \mathbb{R}$

Example 2.1. A line L passes through $A = (1, 2, 3)$ and $B = (-1, 3, 5)$. To find a parametric form of L, we can take

$$\vec{a} = (1, 2, 3) \quad \text{and} \quad \vec{v} = \overrightarrow{AB} = (-1 - 1, 3 - 2, 5 - 3) = (-2, 1, 2)$$

Hence, a parametric form is given by

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

Remark.

1. Parametric form is not unique. For instance,

   $$(x, y, z) = (-1, 3, 5) + t(2, -1, -2) \quad \text{and} \quad (x, y, z) = (-1, 3, 5) + t(-4, 2, 4)$$

   are two other parametrizations of L.

2. By eliminating t from the parametric equation, we get a symmetric form of L

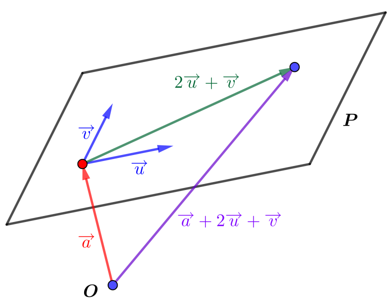
   $$\frac{x - 1}{-2} = \frac{y - 2}{1} = \frac{z - 3}{2}$$

2.2 Planes in $\mathbb{R}^3$

A plane P in $\mathbb{R}^3$ can be uniquely determined by different sets of data, for example,

- 3 non-collinear points on P ; or
- A point on P and 2 linearly independent directions (not same or opposite) ; or
- A point on P and a normal vector

We will study how to represent a plane in equation or parametric form. Suppose P is a plane in $\mathbb{R}^3$, A is a point on it, $\vec{u}$ and $\vec{v}$ are two linearly independent directions of it. Let $\vec{a} = \overrightarrow{OA}$. Then the position vector of any point on P is given by the sum of $\vec{a}$ and a linear combination of $\vec{u}$ and $\vec{v}$.

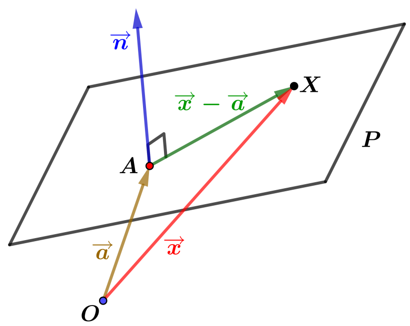


Hence, a parametric form of P can be given by

$$\vec{x} = \vec{a} + s\vec{u} + t\vec{v}$$

Here $s, t \in \mathbb{R}$ are parameters. By varying $s, t \in \mathbb{R}$, we obtain all the points on P. In another situation, suppose $\vec{a} = (a_1, a_2, a_3)$ is a point on P and $\vec{n} = (n_1, n_2, n_3)$ is a **normal vector** of P (that is, a vector perpendicular to the plane P). Let $\vec{x} = (x, y, z) \in \mathbb{R}^3$. Then:

$$\begin{aligned} \vec{x} \text{ is on } P \quad &\iff \quad \vec{x} - \vec{a} \perp \vec{n} \\ &\iff \quad (\vec{x} - \vec{a}) \cdot \vec{n} = \vec{0} \\ &\iff \quad \vec{x} \cdot \vec{n} = \vec{a} \cdot \vec{n} \end{aligned}$$



The plane P can be described by the equation

$$n_1 x + n_2 y + n_3 z = a_1 n_1 + a_2 n_2 + a_3 n_3$$

Remark. If $(a, b, c) \neq \vec{0}$, the equation

$$ax + by + cz = d$$

describes a plane in $\mathbb{R}^3$ with normal vector (a, b, c).

Normal Vector
IFRAME
Normal Vector as Cross Product
IFRAME

Example 2.2. Suppose P is a plane passing through

$$A = (0, 0, 1), B = (0, 2, 0), C = (-1, 1, 0)$$

Represent P using parametric and equation form.

Solution. For parametric form,

$$\begin{aligned}\overrightarrow{AB} &= (0,2,0)-(0,0,1) = (0,2,-1)\\ \overrightarrow{AC} &= (-1,1,0)-(0,0,1) = (-1,1,-1)\end{aligned}$$

Hence

$$(x,y,z) = (0,0,1) + s(0,2,-1) + t(-1,1,-1)$$

To represent P by an equation, we take

$$\vec{n} = \overrightarrow{AB} \times \overrightarrow{AC} = \begin{vmatrix} \hat{i} & \hat{j} & \hat{k} \\ 0 & 2 & -1 \\ -1 & 1 & -1 \end{vmatrix} = (-1,1,2) \perp P$$

Then for any point (x,y,z) on P,

$$\begin{aligned}[(x,y,z)-(0,0,1)]\cdot(-1,1,2) &= 0\\ (-1)x+(1)y+2(z-1) &= 0\\ -x+y+2z &= 2\end{aligned}$$

Example 2.3. Let two planes in $\mathbb{R}^3$ be given:

$$a_1x + b_1y + c_1z = d_1,$$

$$a_2x + b_2y + c_2z = d_2,$$

where $\vec{n}_i := \langle a_i, b_i, c_i\rangle \neq \vec{0}$ $(i = 1,2)$.

Suppose $\vec{n}_1$ and $\vec{n}_2$ are not parallel to each other. Then, the two planes are non-parallel, and the intersection of the two planes is a line parallel to the vector $\vec{v} = \vec{n}_1 \times \vec{n}_2$. Note that the vector $\vec{v}$ is nonzero, since $\vec{n}_1$ and $\vec{n}_2$ are by assumption non-parallel.

Theorem 2.4. *Given a plane in $\mathbb{R}^3$ corresponding to:*

$$a(x-x_0) + b(y-y_0) + c(z-z_0) = 0,$$

The (minimal) distance between a point $P \in \mathbb{R}^3$ and the plane is:

$$d = \left|\mathrm{Proj}_{\vec{n}}\overrightarrow{P_0P}\right| = \left|\overrightarrow{P_0P}\cdot\frac{\vec{n}}{|\vec{n}|}\right|,$$

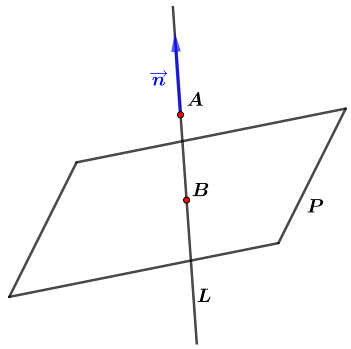
where $P_0 = (x_0, y_0, z_0)$ and $\vec{n} = \langle a,b,c\rangle$.

Example 2.5. Find the distance between $A = (2, 1, 1)$ and the plane P: $-x + 2y - z = -4$.

Solution. From the equation of P, $\vec{n} = (-1, 2, -1) \perp P$. Consider the line L defined by

$$\overrightarrow{X}(t) = \overrightarrow{A} + t\vec{n} = (2, 1, 1) + t(-1, 2, -1)$$

Let $B = L \cap P$ be the intersection of L and P.



Then B is the point of P closest to A. To find B, put:

$$\overrightarrow{X}(t) = (2 - t, 1 + 2t, 1 - t)$$

into the equation of P. Then:

$$-(2 - t) + 2(1 + 2t) - (1 - t) = -4 \Rightarrow 6t - 1 = -4 \Rightarrow 6t = -3 \Rightarrow t = -\frac{1}{2}$$

We have $B = \overrightarrow{X}(-\frac{1}{2}) = (\frac{5}{2}, 0, \frac{3}{2})$. The distance between A and P is

$$= \|\overrightarrow{AB}\| = \sqrt{\left(\frac{5}{2} - 2\right)^2 + (0 - 1)^2 + \left(\frac{3}{2} - 1\right)^2} = \frac{\sqrt{6}}{2}$$

Exercise 2.6. Find the distance between the lines

$$L_1(s) = (-4, 9, -4) + s(4, -3, 0)$$
$$L_2(t) = (5, 2, 10) + t(4, 3, 2)$$

Hint: Find A on L_1, B on L_2 such that $\overrightarrow{AB} \perp L_1, L_2$

2.2.1 Line in $\mathbb{R}^3$ by equations

Can we describe the a straight line in $\mathbb{R}^3$ by an equation? Note that each non-trivial linear equation in x, y, z can only represent a plane. At least two such equations are needed to describe a line. For instance,

Example 2.7. Consider the y-axis in $\mathbb{R}^3$. A point (x, y, z) is on the y-axis if and only if both the x and z coordinates are zero. Hence, y-axis can be described using the equations

$$\begin{cases} x = 0 \\ z = 0 \end{cases}$$

Geometrically, each of the equations $x = 0$ and $z = 0$ represents a plane in $\mathbb{R}^3$. The y-axis is the intersection of the two planes.

IFRAME

Given a linear object, for example, a line or a plane, we can describe it using either parametric form or a system of linear equations. It is easy to convert between the two using linear algebra.

Example 2.8. Let L be the line represented by the system

$$\begin{cases} x + y + 6z = 6 \\ x - y - 2z = -2 \end{cases}$$

By Gaussian elimination,

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 2 \\ 4 \\ 0 \end{bmatrix} + t \begin{bmatrix} -2 \\ -4 \\ 1 \end{bmatrix}.$$

The solution describes L in parametric form.

Conversely, from the above parametric form

$$\begin{cases} x &= 2 - 2t \\ y &= 4 - 4t \\ z &= t \end{cases}$$

The first two equations imply $2x - y = 0$ while the last two imply $y = 4 - 4z$.

We obtain another set of linear equation representing L:

$$\begin{cases} 2x - y = 0 \\ y + 4z = 4 \end{cases}$$

2.2.2 Intersection of Planes

Example 2.9. Consider a system of three non-trivial equations of the form $ax + by + cz = d$. Each of them represents a plane in $\mathbb{R}^3$. What can be their intersections?

- Case 1: Unique solution
- Case 2: Infinitely many solutions
- Case 3: No solution

- All three planes are parallel to each other.

  IFRAME

- Only two planes are parallel to each other.

  IFRAME

- The intersection of each pair of planes is a line and three such lines are parallel to each other.

  IFRAME

- Their intersection is a line.

  IFRAME

- Their intersection is a point, e.g. the xy-plane, yz-plane and zx-plane intersect at $(0, 0, 0)$.

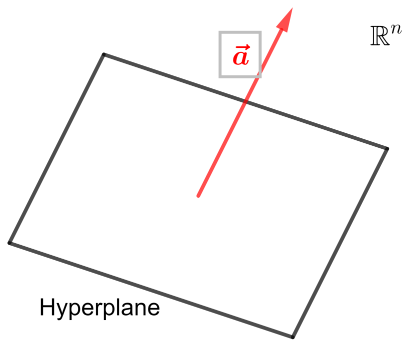
  IFRAME

2.2.3 General linear objects in $\mathbb{R}^n$

Similar to lines in $\mathbb{R}^3$, we need a system of equations to describe a 2-dimensional plane in $\mathbb{R}^n$ when $n \geq 4$. Generally in $\mathbb{R}^n$, an equation of the form

$$\vec{a} \cdot \vec{x} = a_1x_1 + a_2x_2 + \cdots + a_nx_n = c, \;\; \vec{a} \neq \vec{0}$$

describes a hyperplane (dimension $= n - 1$) with normal vector $\vec{a}$:



A k-dimensional "plane" P (called k-plane) in $\mathbb{R}^n$ can be described in parametric form or by equation(s).

1. **Parametric form**

$$\vec{x} = \vec{q} + \sum_{i=1}^{k} t_i \vec{v}_i$$

where

- $\vec{q} \in P$
- $\vec{v}_1, \cdots, \vec{v}_k$ are k linearly independent vectors parallel to P
- $t_1, \cdots, t_k$ are parameters

2. $n - k$ **non-redundant equations**

$$\sum_{j=1}^{n} a_{ij} x_j = c_i \text{ for } i = 1, 2, \cdots, n-k$$

Here non-redundant means that the $(n-k) \times n$ coefficient matrix $A = (a_{ij})$ has rank $n - k$. The solution of the system of $n - k$ equations corresponds to the intersection of the $n - k$ hyperplanes.

2.3 Curves in $\mathbb{R}^n$

Definition 2.10. Let $I \subseteq \mathbb{R}$ be an interval.
A **curve** in $\mathbb{R}^n$ is a continuous function:

$$\vec{x} : I \longrightarrow \mathbb{R}^n$$

That is, $\vec{x}$ is defined as:

$$\vec{x}(t) = (x_1(t), x_2(t), \ldots, x_n(t)), \quad t \in \mathbb{R}$$

where x_i is a continuous real-valued function on I for each i.

IFRAME

Example 2.11. Let $\vec{v} : [-1, 1) \longrightarrow \mathbb{R}^2$ be defined by $\vec{v}(t) = (t^2, t)$. Then $y^2 = t^2 = x$ and the curve lies on the parabola $x = y^2$.

IFRAME

Example 2.12. Let $\vec{p}, \vec{q} \in \mathbb{R}^3, \vec{q} \neq \vec{0}$. Define $\vec{x} : \mathbb{R} \longrightarrow \mathbb{R}^3$ by $\vec{x}(t) = \vec{p} + t\vec{q}$. Then $\vec{x}(t)$ is a straight line.

Definition 2.13. A curve $\vec{x} : [a, b] \longrightarrow \mathbb{R}^n$ is said to be:

- **closed** if $\vec{x}(a) = \vec{x}(b)$.
- **simple** if $\vec{x}(t_1) \neq \vec{x}(t_2)$ for any $a \leq t_1 < t_2 \leq b$, except possibly at $t_1 = a, t_2 = b$.

Example 2.14.

$$\vec{x} : [1, \infty) \longrightarrow \mathbb{R}^2,$$
$$\vec{x}(t) = \left(\frac{1}{t}, \frac{1}{t^2}\right), \quad t \in \mathbb{R}.$$

Definition 2.15. Let $\vec{x}(t) = (x_1(t), x_2(t), \ldots, x_n(t))$, where x_i are real-valued functions. The **derivative** of $\vec{x}$ at t is:

$$\vec{x}'(t) = \lim_{h \to 0} \frac{\vec{x}(t+h) - \vec{x}(t)}{h}.$$

For any a in the domain of $\vec{x}$, if $\vec{x}'(a)$ exists, then $\vec{x}'(a)$ is called the **tangent vector** of $\vec{x}$ at $t = a$.

IFRAME

Theorem 2.16. *Let $\vec{x}(t) = (x_1(t), x_2(t), \ldots, x_n(t))$. Then:*

- $$\lim_{t \to a} \vec{x}(t) = \left(\lim_{t \to a} x_1(t), \lim_{t \to a} x_2(t), \ldots, \lim_{t \to a} x_n(t)\right)$$
- *If $\vec{x}'(t)$ exists, then each x_i is differentiable at t, and:*

$$\vec{x}'(t) = (x_1'(t), x_2'(t), \ldots, x_n'(t)).$$

In physics, if we let $\vec{x}(t)$ be the **displacement** (position) of a moving particle at time t. Then:

- $\vec{x}'(t)$ is the **velocity** of the particle at time t.
- $\vec{x}''(t) = (\vec{x}')'(t)$ is the **acceleration** of the particle at time t.

Example 2.17.

$$\vec{x}(t) = (\cos t, \sin t), 0 \leqslant t \leqslant 2\pi$$

$$\begin{aligned}\vec{v}(t) &= \vec{x}'(t) = (-\sin t, \cos t) \perp \vec{x}(t)\\ \vec{a}(t) &= \vec{x}''(t) = (-\cos t, -\sin t) = -\vec{x}(t)\end{aligned}$$

Also speed $= \|\vec{v}(t)\| = 1$

Example 2.18. Let $\vec{x} : [1, \infty) \to \mathbb{R}^2$ be defined by

$$\vec{x}(t) = \left(\frac{1}{t}, \frac{1}{t^2}\right).$$

Then:

$$\begin{aligned}\lim_{t\to\infty} \vec{x}(t) &= \left(\lim_{t\to\infty} \frac{1}{t}, \lim_{t\to\infty} \frac{1}{t^2}\right)\\ &= (0, 0)\end{aligned}$$

Theorem 2.19. *Let $\vec{x}(t), \vec{y}(t)$ be curves in $\mathbb{R}^n$, and $c \in \mathbb{R}$ a scalar constant. Let $f : \mathbb{R} \longrightarrow \mathbb{R}$ be a real-valued function.*

1. $(\vec{x} \pm \vec{y})'(t) = \vec{x}'(t) \pm \vec{y}'(t)$.
2. $(c\vec{x}(t))' = c\vec{x}'(t)$.
3. $(f(t)\vec{x}(t))' = f'(t)\vec{x}(t) + f(t)\vec{x}'(t)$.
4. $(\vec{x}(t) \cdot \vec{y}(t))' = \vec{x}'(t) \cdot \vec{y}(t) + \vec{x}(t) \cdot \vec{y}'(t)$.
5. *If* $n = 3$,
   $(\vec{x}(t) \times \vec{y}(t))' = \vec{x}'(t) \times \vec{y}(t) + \vec{x}(t) \times \vec{y}'(t)$.

2.4 Arclength

IFRAME

Definition 2.20. Let $\vec{x} : [a, b] \longrightarrow \mathbb{R}^n$ be a curve such that $\vec{x}'$ exists and is continuous on (a, b).

The **arclength** of $\vec{x}$ on $[a, b]$ is:

$$S = \int_a^b \|\vec{x}'(t)\| dt$$

Remark. In physics, if $\vec{x}(t)$ is the displacement of a moving particle at time t, then the arclength of $\vec{x}$ on $[a, b]$ is the **distance travelled** by the particle over the time period $[a, b]$.

If $\vec{x}(t) =$ displacement at time t.
Then, $\vec{x}'(t) =$ velocity
and $\|\vec{x}'(t)\| =$ speed.
$\int_a^b \|\vec{x}'(t)\| dt =$ distance travelled.
From a mathematical point of view, approximate a curve by line segments:
Take: $a = t_0 < t_1 < t_2 < \cdots < t_n = b$. Then,

$$\begin{aligned} S &\approx \sum_{i=1}^{n} \|\vec{x}(t_i) - \vec{x}(t_{i-1})\| \qquad \left(\text{Recall } \vec{x}'(t) := \lim_{h \to 0} \frac{\vec{x}(t+h) - \vec{x}(t)}{h}\right) \\ &\approx \sum_{i=1}^{n} \|\vec{x}'(t_i)\| (t_i - t_{i-1}) \end{aligned}$$

Take Limit $\Rightarrow S = \int_a^b \|\vec{x}'(t)\| dt$

Example 2.21 (Helix)**.** $\vec{x}(t) = (\cos t, \sin t, t), t \in [0, 2\pi]$
IFRAME

1. Find the tangent line of $\vec{x}$ at $t = \pi$
2. Find arclength of the helix.

Solution.

1. $$\begin{aligned} \vec{x}(t) &= (\cos t, \sin t, t) \\ \vec{x}'(t) &= (-\sin t, \cos t, 1) \\ \vec{x}'(\pi) &= (0, -1, 1) \leftarrow \text{direction of tangent} \end{aligned}$$

   Also, $\vec{x}(\pi) = (-1, 0, \pi) \leftarrow$ a point on tangent line

   $\therefore$ Parametric form of tangent line

$$\vec{x} = (-1, 0, \pi) + t(0, -1, 1)$$

2.
$$
\begin{aligned}
\|\vec{x}'(t)\| &= \sqrt{(-\sin t)^2 + (\cos t)^2 + 1^2} \\
&= \sqrt{2} \\
\Rightarrow S &= \int_0^{2\pi} \|\vec{x}'(t)\| \, dt \\
&= \int_0^{2\pi} \sqrt{2} dt \\
&= [\sqrt{2} t]_0^{2\pi} \\
&= 2\sqrt{2}\pi
\end{aligned}
$$

Theorem 2.22. *Arclength is independent of parametrization.*

Example 2.23.

$$
\begin{aligned}
\vec{x}(t) &= (t, t) \;\; 0 \leqslant t \leqslant 4 \\
\vec{y}(t) &= (t^2, t^2) \;\; 0 \leqslant t \leqslant 2
\end{aligned}
$$

$\vec{x}, \vec{y}$ are two parametrization of the same line segment:

$\vec{x}'(t) = (1, 1)$

arclength of $\vec{x}(t)$

$$
\begin{aligned}
&= \int_0^4 \|\vec{x}'(t)\| dt \\
&= \int_0^4 \sqrt{2} dt \\
&= 4\sqrt{2}
\end{aligned}
$$

$\vec{y}'(t) = (2t, 2t)$

arclength of $\vec{y}(t)$

$$
\begin{aligned}
&= \int_0^2 \|\vec{y}'(t)\| dt \\
&= \int_0^2 \sqrt{(2t)^2 + (2t)^2} dt \\
&= \int_0^2 2\sqrt{2} dt \\
&= [\sqrt{2} t^2]_0^2 \\
&= 4\sqrt{2}
\end{aligned}
$$

MATH 2010 Chapter 3

3.1 Polar Coordinates in $\mathbb{R}^2$

A point $P = (x, y) \in \mathbb{R}^2$ can be represented by:

$$r = \sqrt{x^2 + y^2} = \text{ distance from origin.}$$

$$\theta = \text{ angel from the positive } x- \text{ axis to } \overrightarrow{OP} \text{ in counter-clockwise direction.}$$

If $x, y > 0$, then we can take $\theta = \arctan\left(\frac{y}{x}\right)$.

The angle formula above needs to be adjusted for points in other qudrants. For example, if $x < 0, y > 0$ (Quadrant II), then:

$$\theta = \pi + \arctan\left(\frac{y}{x}\right)$$

Remark.

- For $P = (0, 0)$, we have $r = 0$, but θ is not (uniquely) defined.
- Different conventions for ranges of r and θ:

$$r \in [0, \infty) \text{ or } \mathbb{R}$$

$$\theta \in [0, 2\pi) \text{ or } \mathbb{R}$$

In this course, we usually take:

$$r \in [0, \infty), \quad \theta \in \mathbb{R}.$$

3.1.1 Change of Coordinates Fomula

If the polar coordinates for a point (x, y) is (r, θ), then:

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

3.1.2 Curves in Polar Coodinates

Example 3.1 (Circle with radius r0)**. Polar equation**

$$r = r_0$$

Parametric form

$$\begin{cases} r = r_0 \\ \theta = t, \quad t \in [0, 2\pi]. \end{cases}$$

Example 3.2 (Half ray from origin)**. Polar equation**

$$\theta = \theta_0$$

Polar equation

$$\begin{cases} r = t, \quad t \in [0, \infty) \\ \theta = \theta_0. \end{cases}$$

Example 3.3 (Archimedes Spiral)**.** Let $k > 0$ be a constant

Polar equation

$$r = k\theta$$

Polar equation

$$\begin{cases} r = kt, t \in [0, \infty) \\ \theta = t, t \in [0, \infty) \end{cases}$$

Example 3.4.

$$r = 4\cos\theta$$

IFRAME

Observe that the origin, corresponding to $r = 0, \theta = \pi/2$, lies on the graph of $r = 4\cos\theta$. Hence, the solution set of $r = 4\cos\theta$ is equal to the solution set of:

$$r^2 = 4r\cos\theta,$$

which is equivalent to the Cartesian equation:

$$x^2 + y^2 = 4x$$

Completing the square, the equation above is equivalent to:

$$(x-2)^2 + y^2 = 2^2,$$

which corresponds to the circle of radius 2 centered at $(2, 0)$.

Example 3.5.

$$r\cos\left(\theta - \frac{\pi}{4}\right) = \sqrt{2}.$$

(Hint: The graph is a straight line in the Cartesian plane.)

Example 3.6. IFRAME

It is sometimes convenient to allow $r < 0$ in polar coordinates.

For instance, to describe a line through the origin which forms an angle of $\pi/6$ with the positive x-axis, we can simply describe it as the graph of:

$$\theta = \pi/6$$

with the assumption that $r \in \mathbb{R}$.

(If we only let $r \geq 0$, then we only get "half" a line.)

Example 3.7. Let $a > 1$ be constant. Consider:

$$r = 1 - acos\theta$$

If we require that $r \geq 0$, then the equatio above only possibly holds for $\theta \in [\delta, 2\pi - \delta]$, where $\delta = \arccos(1/a)$.

IFRAME

On the other hand, of we let allow r to also be negative, then for any $\theta \in [0, 2\pi]$ there is an r for which the equation holds. The resulting graph would have one extra "loop".

IFRAME

3.2 Coordinate Systems in $\mathbb{R}^3$

Definition 3.8. Given a point $P \in \mathbb{R}^3$ with Cartesian coordinates (x, y, z).

The **cylindrical coordinates** of P is:

$$(r, \theta, z),$$

where (r, θ) are the polar coordinates of (x, y).

Hence,

$$\begin{aligned} x &= r\cos\theta, \\ y &= r\sin\theta, \\ z &= z. \end{aligned}$$

IFRAME

Example 3.9. Let $a, b \in \mathbb{R}$. A vertical helix with radius a may be described with cylindrical coordinates as follows:

$$\begin{cases} r = a \\ \theta = t \\ z = bt \end{cases}, \quad t \in [0, 2\pi]$$

Definition 3.10. Given a point $P \in \mathbb{R}^3$ with Cartesian coordinates (x, y, z).

The **spherical coordinates** of P is:

$$(\rho, \theta, \phi),$$

where:

- $\rho = \sqrt{x^2 + y^2 + z^2}$ is the distance between P and the origin.
- θ is the angle coordinate of the polar coordinates of (x, y) in the xy-plane.
- ϕ is the angle between the positive z-axis and $\overrightarrow{OP}$.

Hence,

$$\begin{aligned} x &= \rho \sin\phi \cos\theta, \\ y &= \rho \sin\phi \sin\theta, \\ z &= \rho \cos\phi. \end{aligned}$$

IFRAME

Example 3.11 (Sphere)**.**

$$\rho = 2.$$

Example 3.12 (Cone)**.**

$$\phi = \pi/4.$$

Example 3.13 (Half Plane)**.**

$$\theta = \pi/3.$$

Example 3.14 (Circle)**. Equations:**

$$\begin{cases} \rho = 3, \\ \phi = \pi/2. \end{cases}$$

Parametric Form:

$$(\rho, \theta, \phi)_{sph} = (3, t, \pi/2), \quad t \in [0, 2\pi].$$

3.3 Topological Terminology

Let $\vec{x}_0 \in \mathbb{R}^n$, $\varepsilon > 0$.

Definition 3.15. The **open ball** with radius ε centered at $\vec{x}_0$ is:

$$B_\varepsilon(\vec{x}_0) = \{\vec{x} \in \mathbb{R}^n : \|\vec{x} - \vec{x}_0\| < \varepsilon.\}$$

The **closed ball** with radius ε centered at $\vec{x}_0$ is:

$$\overline{B_\varepsilon(\vec{x}_0)} = \{\vec{x} \in \mathbb{R}^n : \|\vec{x} - \vec{x}_0\| \leq \varepsilon.\}$$

Let $S \subseteq \mathbb{R}^n$.

Definition 3.16.

- The **interior** of S is the set:

  $$\operatorname{Int}(S) = \{\vec{x} \in \mathbb{R}^n : B_\varepsilon(\vec{x}) \subset S \text{ for some } \varepsilon > 0.\}$$

  Points in $\operatorname{Int}(S)$ are called **interior points** of S.

- The **exterior** of S is the set:

  $$\operatorname{Ext}(S) = \{\vec{x} \in \mathbb{R}^n : B_\varepsilon(\vec{x}) \subset \mathbb{R}^n \backslash S \text{ for some } \varepsilon > 0.\}$$

  Points in $\operatorname{Ext}(S)$ are called **exterior points** of S.

- The **boundary** of S is the set:

  $$\partial S = \{\vec{x} \in \mathbb{R}^n : B_\varepsilon(\vec{x}) \cap S \neq \varnothing \text{ and } B_\varepsilon(\vec{x}) \cap \mathbb{R}^n \backslash S \neq \varnothing, \text{ for all } \varepsilon > 0.\}$$

  Points in $\partial(S)$ are called **boundary points** of S.

IMAGE

Example 3.17.

$$S = \left\{(x, y) \in \mathbb{R}^2 : 1 < x^2 + y^2 \leq 4\right\} \subseteq \mathbb{R}^2$$

Proposition 3.18. *Let $S \subseteq \mathbb{R}^n$. Then,*

- $\mathbb{R}^n$ *is the disjoint union of* $\operatorname{Int}(S), \operatorname{Ext}(S)$ *and* ∂S.
- $\operatorname{Int}(S) \subseteq S$, $\operatorname{Ext}(S) \subseteq \mathbb{R}^n \backslash S$.

Definition 3.19. A subset $S \subseteq \mathbb{R}^n$ is said to be

- **open** if for all $x \in S$, there exists $\varepsilon > 0$ such that $B_\varepsilon(x) \subseteq S$.

- **closed** if $\mathbb{R}^n \backslash S$ is open.

Definition 3.20 (Closure)**.** The **closure** of a set $A \subseteq \mathbb{R}^n$ is:

$$\bar{A} = A \cup \partial A$$

Remark. The closure of any set is always closed.

Theorem 3.21. *A subset $S \subseteq \mathbb{R}^n$ is:*

- **open** *if and only if $S = \mathrm{Int}(S)$.*
- **closed** *if and only if $S = \mathrm{Int}(S) \cup \partial S$.*

Example 3.22.

Subset $S \subseteq \mathbb{R}^n$	$B_1(0,0) = \{(x,y) \in \mathbb{R}^2 : x^2 + y^2 < 1\}$	$\overline{B_1(0,0)} = \{(x,y) \in \mathbb{R}^2 : x^2 + y^2 \leq 1\}$
$\mathrm{Int}(S)$		
$\mathrm{Ext}(S)$		
∂S		
Open?		
Closed?		

Remark.
- There are exactly two subsets of $\mathbb{R}^n$ which are both open and closed:

$$\mathbb{R}^n, \varnothing$$

- Some subsets of $\mathbb{R}^n$ are neither open nor closed:

$$\{(x,y) \in \mathbb{R}^2 : 1 < x^2 + y^2 \leq 4\} \subseteq \mathbb{R}^n$$

$$(0,1] \subseteq \mathbb{R}$$

$$\mathbb{Q} \subseteq \mathbb{R}$$

Exercise : $\partial\mathbb{Q} = \mathbb{R}$.

Definition 3.23. A subset $S \subseteq \mathbb{R}^n$ is said to be:

- **bounded** if there exists $M > 0$ such that:

$$S \subseteq B_M(\vec{0}) = \{\vec{x} \in \mathbb{R}^n : \|\vec{x}\| < M\}$$

- **unbounded** if it is not bounded.

Definition 3.24. A subset $S \subseteq \mathbb{R}^n$ is said to be **path-connected** if any two points in S can be connected by a curve in S.

Theorem 3.25 (Jordan Curve Theorem)**.** *A simple closed curve in $\mathbb{R}^2$ divides $\mathbb{R}^2$ into two path-connected components, with one bounded and one unbounded.*

MATH 2010 Chapter 4

4.1 Vector-Valued Functions in Multiple Variables

Let

$$\vec{f} : \Omega \longrightarrow \mathbb{R}^m,$$

be a vector-valued function, where $\Omega \subseteq \mathbb{R}^n$.

Definition 4.1. The **graph** of $\vec{f}$ is:

$$\mathrm{Graph}(\vec{f}) = \left\{ \left(\vec{x}, \vec{f}(\vec{x}) \right) : \vec{x} \in \Omega \right\} \subseteq \mathbb{R}^{n+m}$$

4.1.1 Level Set

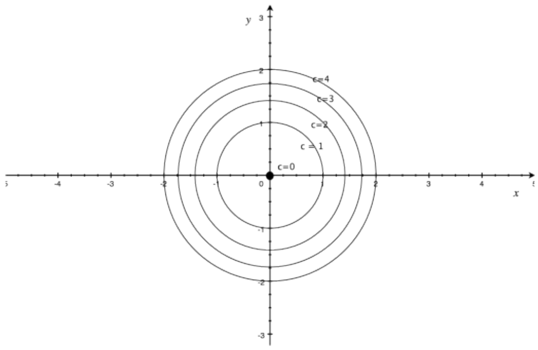
For a function $\vec{f} : \Omega \longrightarrow \mathbb{R}^m$, $\Omega \subseteq \mathbb{R}^n$, in n variables, and $\vec{c} \in \mathbb{R}^m$, the **level set** of $\vec{f}$ corresponding to $\vec{c}$ is the set of points $(x_1, x_2, \ldots, x_n) \in \Omega$ such that

$$\vec{f}(x_1, x_2, \ldots, x_n) = \vec{c}$$

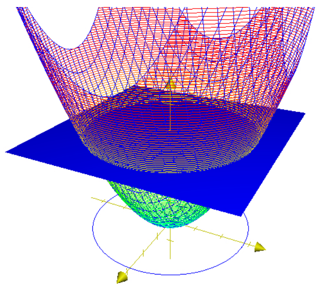
- If $n = 2$, then a level set of $\vec{f}$ is typically a curve in the xy-plane, and is often called a **level curve**.
- If $n = 3$, then a level set is typically a surface in the xyz-space, and is often called a **level surface**.

Example 4.2. $f(x, y) = x^2 + y^2$.

- For $c = -2, -1$, the level sets $f(x, y) = x^2 + y^2 = c$ are empty.
- For $c = 0$, the level set $f(x, y) = x^2 + y^2 = 0$ consists of the single point $(0, 0)$.
- For $c > 0$, the level set $f(x, y) = x^2 + y^2 = c$ is the circle in $\mathbb{R}^2$ centred at the origin with radius $\sqrt{c}$.



Each level set $f(x, y) = c$ corresponds to (the projection onto the xy-plane of) the intersection of the surface $z = f(x, y)$ and the horizontal (hence "level") plane $z = c$:



IFRAME

4.2 Limits of Multivariable Functions

First, recall Closure.

Definition 4.3 (Limit)**.** Let $\vec{f} : A \longrightarrow \mathbb{R}^m$ be a vector-valued function on $A \subseteq \mathbb{R}^n$.
For any $\vec{a} \in \bar{A}$, we say that: *The limit of $\vec{f}$ at $\vec{a}$ is $\vec{L}$*

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

if: For all $\varepsilon > 0$, there exists $\delta > 0$ such that:

$$\|\vec{f}(\vec{x}) - \vec{L}\| < \varepsilon$$

for all $\vec{x} \in A$ which satisfies $0 < \|\vec{x} - \vec{a}\| < \delta$.

Example 4.4. Let:

$$f : \mathbb{R}^2 \longrightarrow \mathbb{R},$$
$$f(x,y) = x + y, \quad (x,y) \in \mathbb{R}^2.$$

Then,

$$\lim_{(x,y)\to(1,2)} f(x,y) = 3.$$

Proof of Example 4.4. Show that given any $\varepsilon > 0$, one can find $\delta > 0$ such that if $0 < \|(x,y) - (1,2)\| < \delta$, then $|f(x,y) - 3| < \varepsilon$.

Idea:

$$\begin{aligned} |f(x,y) - 3| &= |(x-1) + (y-2)| \\ &\leqslant |x-1| + |y-2| \\ \|(x,y) - (1,2)\| &= \sqrt{(x-1)^2 + (y-2)^2}. \end{aligned}$$

For example, for $\varepsilon = 1$, one can pick $\delta = \frac{1}{2}$:

If $\|(x,y) - (1,2)\| < \delta = \frac{1}{2}$, then:

$$|x-1| = \sqrt{(x-1)^2} \leqslant \sqrt{(x-1)^2 + (y-2)^2} < \frac{1}{2}$$
$$|y-2| = \sqrt{(y-2)^2} \leqslant \sqrt{(x-1)^2 + (y-2)^2} < \frac{1}{2}$$

This implies that:

$$|f(x,y) - 3| \leqslant |x-1| + |y-2| < \frac{1}{2} + \frac{1}{2} = 1 = \varepsilon$$

Similarly, for $\varepsilon = \frac{1}{100}$, one can pick $\delta = \frac{1}{200}$

In general, we need to do it for any $\varepsilon > 0$ For any given $\varepsilon > 0$, one can pick $\delta = \frac{\varepsilon}{2}$. Then:

$$\begin{aligned} &\|(x,y) - (1,2)\| < \delta = \frac{\varepsilon}{2} \\ \Rightarrow &|f(x,y) - 3| = |x + y - 3| \leqslant |x-1| + |y-2| \\ &\qquad < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon \end{aligned}$$

Hence, $\lim_{(x,y)\to(1,2)} f(x,y) = 3.$ □

Example 4.5. Let:

$$f : \mathbb{R}^2 \longrightarrow \mathbb{R},$$
$$f(x,y) = x^2 + y^2, \quad (x,y) \in \mathbb{R}^2.$$

Then,

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

Proof of Example 4.5. For all $\varepsilon > 0$, we need to find $\delta > 0$ such that:
if:

$$0 < \|(x,y) - (0,0)\| = \sqrt{x^2 + y^2} < \delta,$$

then:

$$|f(x,y) - 0\| = \left|x^2 + y^2\right| < \varepsilon.$$

Exercise: Complete the rest of the proof. □

Proposition 4.6. *Let $A \subseteq \mathbb{R}^n$, $a \in A$, $\vec{f} : A \longrightarrow \mathbb{R}^m$, where:*

$$\vec{f}(\vec{x}) = \begin{bmatrix} f_1(\vec{x}) \\ f_2(\vec{x}) \\ \vdots \\ f_m(\vec{x}) \end{bmatrix}, \quad f_i : A \longrightarrow \mathbb{R}.$$

Then,

$$\lim_{\vec{x} \to \vec{a}} \vec{f}(\vec{x}) = \vec{L} = \begin{bmatrix} l_1 \\ l_2 \\ \vdots \\ l_m \end{bmatrix}$$

if and only if

$$\lim_{\vec{x} \to \vec{a}} f_i(\vec{x}) = l_i$$

for $i = 1, 2, \ldots, m$.

Example 4.7. Let:

$$\vec{f} : \mathbb{R}^2 \longrightarrow \mathbb{R}^2$$

$$\vec{f}(x,y) = \begin{bmatrix} x + y \\ x^2 + y^2 + 1 \end{bmatrix}, \quad (x,y) \in \mathbb{R}^2.$$

Then,

$$\lim_{(x,y) \to (1,2)} \vec{f}(x,y) = \begin{bmatrix} \lim_{(x,y) \to (1,2)} x + y \\ \lim_{(x,y) \to (1,2)} x^2 + y^2 + 1 \end{bmatrix} = \begin{bmatrix} 3 \\ 6 \end{bmatrix}$$

Proposition 4.8. *Let $f : A \subseteq \mathbb{R}^n \longrightarrow \mathbb{R}^m$ be a function. Let $\gamma, \psi : \mathbb{R} \longrightarrow \mathbb{R}^n$ be the parameterization of two paths in $\mathbb{R}^n$, with $\gamma(0) = \psi(0) = \vec{a}$. If $\lim\limits_{t \to 0} \vec{f}(\gamma(t))$ or $\lim\limits_{t \to 0} \vec{f}(\psi(t))$ does not exist, or the two limits are not equal to each other, then the limit $\lim\limits_{\vec{x} \to \vec{a}} \vec{f}(\vec{x})$ does not exist.*

In fact:

Theorem 4.9. $\lim_{\vec{x}\to\vec{a}} \vec{f}(\vec{x}) = \vec{L}$ *if and only if the limit of* $\vec{f}(\vec{x})$ *at* $\vec{a}$ *along any path through* $\vec{a}$ *exists and is equal to* $\vec{L}$*.*

Example 4.10. Consider $\lim_{(x,y)\to(0,0)} f(x,y)$, where:

$$f(x,y) = \frac{xy}{x^2+y^2}.$$

Let:

$$\gamma(t) = (t,t), \quad t \in \mathbb{R},$$
$$\psi(t) = (t,-t), \quad t \in \mathbb{R}.$$

Then,

$$\gamma(0) = \psi(0) = (0,0),$$

and:

$$\lim_{t\to0} f(\gamma(t)) = \lim_{t\to0} \frac{t\cdot t}{t^2+t^2} = \lim_{t\to0} \frac{t^2}{2t^2} = \frac{1}{2},$$
$$\lim_{t\to0} f(\psi(t)) = \lim_{t\to0} \frac{t\cdot(-t)}{t^2+(-t)^2} = \lim_{t\to0} -\frac{t^2}{2t^2} = -\frac{1}{2},$$

Since $\lim_{t\to0} f(\gamma(t)) \neq \lim_{t\to0} f(\psi(t))$, we conclude that the limit $\lim_{(x,y)\to(0,0)} \frac{xy}{x^2+y^2}$ does not exist.

Remark. Let $\vec{a} = (x_0, y_0)$. If $\lim_{x\to x_0} f(x,y_0) = \lim_{x\to y_0} f(x_0,y) = L$, it is not necessarily true that $\lim_{(x,y)\to(x_0,y_0)} f(x,y) = L$, or that the limit even exists.

Example 4.11.

$$f : \mathbb{R}^2 \longrightarrow \mathbb{R}$$
$$f(x,y) = \begin{cases} 1 & \text{if} \quad 0 < y < x^2 \\ 0 & \text{otherwise} \end{cases}$$

Find $\lim_{(x,y)\to\vec{a}} f(x,y)$, where:

1. $\vec{a} = (0,1)$
2. $\vec{a} = (1,1)$
3. $\vec{a} = (0,0)$

4.2.1 Proporties of Limits

If all limits on the right-hand side exists, then the limit of the left-hand side exists and the formula holds:

1. $\lim_{\vec{x}\to\vec{a}}(\vec{f}(x) \pm \vec{g}(x)) = \lim_{\vec{x}\to\vec{a}} \vec{f}(\vec{x}) \pm \lim_{\vec{x}\to\vec{a}} \vec{g}(\vec{x})$.

2. $\lim_{\vec{x}\to\vec{a}} k\vec{f}(\vec{x}) = k \lim_{\vec{x}\to\vec{a}} \vec{f}(\vec{x})$ for any scalar constant k.

3. If $\vec{f}$ and $\vec{g}$ are real-valued, then $\lim_{\vec{x}\to\vec{a}} \vec{f}(\vec{x})\vec{g}(\vec{x}) = \left(\lim_{\vec{x}\to\vec{a}} \vec{f}(\vec{x})\right)\left(\lim_{\vec{x}\to\vec{a}} \vec{g}(\vec{x})\right)$.

4. $\lim_{\vec{x}\to\vec{a}} \dfrac{f(\vec{x})}{g(\vec{x})} = \dfrac{\lim_{\vec{x}\to\vec{a}} \vec{f}(\vec{x})}{\lim_{\vec{x}\to\vec{a}} \vec{g}(\vec{x})}$ provided that $\lim_{\vec{x}\to\vec{a}} \vec{g}(\vec{x}) \neq 0$.

5. $$\lim_{\vec{x}\to\vec{a}}(f(\vec{x}))^n = \left(\lim_{\vec{x}\to\vec{a}} f(\vec{x})\right)^n \qquad \text{for all } n \in \mathbb{N} = \{1, 2, 3, \ldots\},$$

6. $$\lim_{\vec{x}\to\vec{a}} (f(x))^{1/n} = \left(\lim_{\vec{x}\to\vec{a}} f(\vec{x})\right)^{1/n} \qquad \text{for all odd positive integers } n.$$

7. If $\lim_{\vec{x}\to\vec{a}} f(\vec{x}) = L > 0$, then
$$\lim_{\vec{x}\to\vec{a}} (f(\vec{x}))^{1/n} = L^{1/n}$$
for all $n \in \mathbb{N}$.

Theorem 4.12 (Squeeze Theorem)**.** *Let $f, g, h : \Omega \longrightarrow \mathbb{R}$ be real-valued functions on $\Omega \in \mathbb{R}^n$.*

If:
$$g(\vec{x}) \leq f(\vec{x}) \leq h(\vec{x})$$
for all $\vec{x}$ near $\vec{a} \in \Omega$, and
$$\lim_{\vec{x}\to\vec{a}} g(\vec{x}) = \lim_{\vec{x}\to\vec{a}} h(\vec{x}) = L,$$
then:
$$\lim_{\vec{x}\to\vec{a}} f(\vec{x}) = L.$$

Corollary 4.13. *If $|f(\vec{x})| \leq g(\vec{x})$ near $\vec{a}$ and $\lim_{\vec{x}\to\vec{a}} g(\vec{x}) = 0$, then $\lim_{\vec{x}\to\vec{a}} f(\vec{x}) = 0$.*

Example 4.14. Find:

$$\lim_{(x,y)\to(0,0)} x\cos\left(\frac{1}{x^2+y^2}\right).$$

Solution. Note:

$$\left|\cos\left(\frac{1}{x^2+y^2}\right)\right| \leqslant 1 \Rightarrow \left|x\cos\left(\frac{1}{x^2+y^2}\right)\right| \leqslant |x|$$

Also,

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

Hence,

$$\lim_{(x,y)\to(0,0)} x\cos\left(\frac{1}{x^2+y^2}\right) = 0$$

by the Squeeze Theorem.

Example 4.15. Find:

$$\lim_{(x,y)\to(1,0)} \frac{(x-1)^2\ln x}{(x-1)^2+y^2}.$$

Solution. Note:

$$\begin{aligned}\left|\frac{(x-1)^2\ln x}{(x-1)^2+y^2}\right| &= \left|\frac{(x-1)^2}{(x-1)^2+y^2}\right|\cdot|\ln x|\\ &\leqslant |\ln x|\end{aligned}$$

Also, $\lim_{(x,y)\to(1,0)} |\ln x| = |\ln(1)| = 0$
By squeeze theorem,

$$\lim_{(x,y)\to(1,0)} \frac{(x-1)^2\ln x}{(x-1)^2+y^2} = 0$$

Remark. If $a \geqslant b$, then
$ca \leqslant cb$ if $c > 0$
$ca \leqslant cb$ if $c < 0$

MATH 2010 Chapter 5

5.1 Finding Limits Using Polar Coordinates

Recall:

$$(x, y) \longleftrightarrow (r, \theta)_{pol}$$

with:

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

and:

$$(x, y) = (0, 0) \iff r = 0.$$

Example 5.1. Find:

$$\lim_{(x,y)\to(0,0)} \frac{x^3 + y^3}{x^2 + y^2}.$$

Solution.

$$\begin{aligned} &= \lim_{r\to 0} \frac{r^3\cos^3\theta + r^3\sin^3\theta}{r^2\cos^2\theta + r^2\sin^2\theta} \\ &= \lim_{r\to 0} r\left(\cos^3\theta + \sin^3\theta\right) \\ &= 0 \qquad \text{(Squeeze theorem)} \end{aligned}$$

Example 5.2. Find:

$$\lim_{(x,y)\to(0,0)} \frac{x^2 + xy}{2(x^2 + y^2)}.$$

Solution.

$$= \lim_{r\to 0} \frac{r^2\cos^2\theta + r^2\cos\theta\sin\theta}{2r^2}$$

$$= \lim_{r\to 0} \frac{\cos^2\theta + \cos\theta\sin\theta}{2}$$

$$= \begin{cases} \frac{1}{2} & \text{if} \quad \theta = 0 \\ 0 & \text{if} \quad \theta = \frac{\pi}{2} \end{cases}$$

In other words, the function approach different values as (x, y) approaches $(0, 0)$ at different angles. Hence, the limit **does not exist** .

Example 5.3. Find:

$$\lim_{(x,y)\to(0,0)} xy\ln(x^2 + y^2).$$

Solution.

$$= \lim_{r\to 0} \underbrace{r^2\cos\theta\sin\theta\ln(r^2)}$$

Observe that, as $r \to 0$,

$$|\cos\theta\sin\theta| \leqslant 1,$$
$$r^2 \to 0,$$
$$\ln\left(r^2\right) \to -\infty.$$

Moreover:

$$\left|r^2\cos\theta\sin\theta\ln\left(r^2\right)\right| \leq \left|r^2\ln\left(r^2\right)\right|$$

We have:

$$\lim_{r\to 0} r^2\ln\left(r^2\right) = \lim_{r\to 0} \frac{\ln\left(r^2\right)}{\frac{1}{r^2}} \qquad \left(\frac{-\infty}{\infty}\right)$$

$$= \lim_{r\to 0} \frac{\frac{2r}{r^2}}{-\frac{2}{r^3}} \text{ (L' Hopital's Rule)}$$

$$= \lim_{r\to 0} -r^2 = 0$$

By Squeeze theorem, it now follows that:

$$\lim_{(x,y)\to(0,0)} xy\ln\left(x^2 + y^2\right) = 0.$$

5.2 Iterated Limits

Example 5.4. Consider:

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

$$\begin{aligned}\lim_{x\to 0}\lim_{y\to 0}\frac{x+y}{x-y} &= \lim_{x\to 0}\frac{x+0}{x-0}\\ &= 1.\end{aligned}$$

On the other hand,

$$\begin{aligned}\lim_{y\to 0}\lim_{x\to 0}\frac{x+y}{x-y} &= \lim_{y\to 0}\frac{0+y}{0-y}\\ &= -1.\end{aligned}$$

Moreover, $\lim_{(x,y)\to(0,0)}\frac{x+y}{x-y}$ does not exist (**Exercise**).

Remark.

- In general, if $\lim_{x\to 0}\lim_{y\to 0} f(x,y)$ and $\lim_{y\to 0}\lim_{x\to 0} f(x,y)$ both exist and are equal to each other, it does *NOT* follow that $\lim_{(x,y)\to(0,0)} f(x,y)$ exists. Counter-example:

$$f(x,y) = \begin{cases} 1 & \text{if } x = y; \\ 0 & \text{if } x \neq y. \end{cases}$$

- Conversely, if $\lim_{(x,y)\to(0,0)} f(x,y)$ exists, it also does *NOT* follow that:

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

both exist. Counter-example:

$$f(x,y) = \begin{cases} x\cos\frac{1}{y} + y\cos\frac{1}{x} & \text{if } (x,y) \neq (0,0); \\ 0 & \text{if } (x,y) = (0,0). \end{cases}$$

- If all three limits exist, then they are equal.

5.3 Continuity

Definition 5.5. We say that a function $f : A \longrightarrow \mathbb{R}$ in n variables is **continuous** at $\vec{a} \in A$ if:

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

Definition 5.6. A function $\vec{f} : A \longrightarrow \mathbb{R}$ is **continuous** is f is continuous at every point in its domain A.

Example 5.7. Each "coordinate function" $f_i : \mathbb{R}^m \longrightarrow \mathbb{R}$, defined by:

$$f_i(x_1, x_2, \dots, x_m) = x_i,$$

is continuous.

Theorem 5.8. *Let k be a scalar constant. If $f, g : A \longrightarrow \mathbb{R}$ are continuous at $\vec{a} \in A$, then:*

- *$f + g$, kf, fg are all continuous at $\vec{a}$*
- *$\dfrac{f}{g}$ is continuous at $\vec{a}$ is $g(\vec{a}) \neq 0$.*

Proof of Theorem 5.8. This follows from the properties of limits. □

Corollary 5.9. *All polynomial and rational functions (i.e. polynomial divided by another polynomial) are continuous (on their domains).*

Theorem 5.10. *If $f : A \longrightarrow \mathbb{R}$ is continuous at $\vec{a} \in A$, and $g : I \longrightarrow \mathbb{R}$ is a single-variable real-valued function continuous at $f(\vec{a})$, then $g \circ f : A \longrightarrow \mathbb{R}$ is continuous at $\vec{a}$.*

In other words:

$$\lim_{\vec{x} \to \vec{a}} g(f(\vec{x})) = g\left(\lim_{\vec{x} \to \vec{a}} f(\vec{x})\right) = g(f(\vec{a})).$$

Corollary 5.11. *Every so-called "***elementary function***" (a function constructed from constants, power functions, trigonometric, inverse trigonometric, exponential and logarithmic functions, via addition, subtraction, multiplication, division and composition) is continuous at all points in its domain.*

Example 5.12.

- Every polynomial in n variables (e.g. $f(x, y, z) = x^2yz + 5yz^2 + 16y^3 - 8$) is continuous everywhere.
- Every rational function in n variables is continuous at all points where the function is defined.
- $f(x, y) = e^{\cos(x^2+y^2)}$ is continuous at all $(x, y) \in \mathbb{R}^2$.
- $f(x, y) = \dfrac{1}{\sqrt{x^2 + y}}$ is continuous at all $(x, y) \in \mathbb{R}^2$ such that $x^2 + y > 0$.

Example 5.13. • Consider:

$$g(x,y) = \frac{x^4 - y^4}{x^2 + y^2}.$$

Since $x^2 + y^2 = 0 \Leftrightarrow (x,y) = (0,0)$, the domain of g is $\mathbb{R}^2 \backslash \{(0,0)\}$.

$$\begin{aligned}
\lim_{(x,y \to (0,0)} g(x,y) &= \lim_{(x,y) \to (0,0)} \frac{x^4 - y^4}{x^2 + y^2} \\
&= \lim_{r \to 0} \frac{r^4 \cos^4 \theta - r^4 \sin^4 \theta}{r^2 \cos^2 \theta + r^2 \sin^2 \theta} \\
&= \lim_{r \to 0} r^2 \left(\cos^2 \theta - \sin^2 \theta\right) \\
&= 0 \quad \text{(Sandwich theorem)}
\end{aligned}$$

Hence, g *can* be extended to a continuous function on the whole $\mathbb{R}^2$ as follows:

$$g(x,y) = \begin{cases} \frac{x^4 - y^4}{x^2 + y^2} & \text{if } (x,y) \neq (0,0) \\ 0 & \text{if } (x,y) = (0,0) \end{cases}$$

• On the other hand, consider:

$$f(x,y) = \frac{xy + y^3}{x^2 + y^2}$$

$$\begin{aligned}
\lim_{\substack{(x,y) \to (0,0) \\ y = mx}} f(x,y) &= \lim_{\substack{(x,y) \to (0,0) \\ y = mx}} \frac{xy + y^3}{x^2 + y^2} \\
&= \lim_{x \to 0} \frac{mx^2 + m^3 x^3}{x^2 + m^2 x^2} \\
&= \lim_{x \to 0} \frac{m + m^3 x}{1 + m^2} \\
&= \frac{m}{1 + m^2} = \begin{cases} 0, & \text{if } m = 0 \\ \frac{1}{2} & \text{if } m = 1 \end{cases}
\end{aligned}$$

Since the limit varies with slope, $\lim_{(x,y) \to (0,0)} f(x,y)$ does not exist.

The function f cannot be extended to a function defined on $\mathbb{R}^2$.

5.4 Partial Derivatives

Definition 5.14. Let $f : A \longrightarrow \mathbb{R}$ be a function on an open region $A \in \mathbb{R}^n$, $\vec{a} = (a_1, a_2, \ldots, a_n) \in A$. For $i = 1, 2, \ldots, n$, we define the **partial derivative**

with respect to x_i of f at $\vec{a}$ to be:

$$\frac{\partial f}{\partial x_i}(\vec{a}) = \left(\frac{d}{dx_i} f(a_1, a_2, \ldots, a_{i-1}, \underbrace{x_i}_{i\text{-th coordinate}}, a_{i+1}, \ldots, a_n) \right) \Bigg|_{x_i = a_i}$$

$$= \lim_{h \to 0} \frac{f(a_1, a_2, \ldots, a_{i-1}, a_i + h, a_{i+1}, \ldots, a_n) - f(\vec{a})}{h}$$

Observe that as $\vec{a}$ varies, the correspondence:

$$\vec{a} \mapsto \frac{\partial f}{\partial x_i}(\vec{a})$$

defines a real-valued function on a subset A' of A consisting of those points $\vec{a} \in A$ where $\frac{\partial f}{\partial x_i}(\vec{a})$ is defined.

We have therefore a multivariable function defined as follows:

Definition 5.15.

$$\frac{\partial f}{\partial x_i} : A' \longrightarrow \mathbb{R},$$

$$\frac{\partial f}{\partial x_i}(x_1, x_2, \ldots, x_n)$$
$$= \lim_{h \to 0} \frac{f(x_1, x_2, \ldots, x_{i-1}, x_i + h, x_{i+1}, \ldots, x_n) - f(x_1, x_2, \ldots, x_n)}{h}.$$

Notation. Other notations for $\dfrac{\partial f}{\partial x_i}$ are:

$$f_{x_i},\ \partial_i f,\ D_i f,\ \nabla_i f$$

IFRAME

Example 5.16.

$$f(x, y) = x^2 + y^2$$
$$\frac{\partial f}{\partial x} = 2x + 0 = 2x \quad \text{(Regard } y \text{ as a constant)}$$
$$\frac{\partial f}{\partial y} = 0 + 2y = 2y \quad \text{(Regard } x \text{ as a constant)}$$

In particular:

$$\frac{\partial f}{\partial x}(1, -1) = 2(1) = 2 > 0$$
$$\frac{\partial f}{\partial y}(1, -1) = 2(-1) = -2 < 0$$

This means that $f(x, y)$ increases as x increases at $(1, -1)$, and it decreases as y increases at $(1, -1)$.

Example 5.17.

$$f(x, y, z) = xy^2 - \cos(xz)$$

Find f_x, f_y, f_z.

Solution.

$$f_x = y^2 + z\sin(xz)$$

$$f_y = 2xy + 0 = 2xy$$

$$f_z = 0 + x\sin(xz) = x\sin(xz)$$

Example 5.18.

$$f(x, y) = \begin{cases} 1 & \text{if } xy \geq 0; \\ 0 & \text{if } xy < 0. \end{cases}$$

Find $\frac{\partial f}{\partial x}(1, 1)$, $\frac{\partial f}{\partial x}(0, 1)$, $\frac{\partial f}{\partial x}(0, 0)$.

Solution. $\frac{\partial f}{\partial x}$: Fix y, differentiate $f(x, y)$ with respect to x.

Along $y = 1$

$$f(x, 1) = \begin{cases} 1 & \text{if } x \geqslant 0 \\ 0 & \text{if } x < 0 \end{cases}$$

Hence:

$$\frac{\partial f}{\partial x}(1, 1) = 0,$$

and:

$$\frac{\partial f}{\partial x}(0, 1) \ \text{ DNE}$$

Along $y = 0$ We have $f(x, 0) = 1$ for all $x \in \mathbb{R}$. This implies that:

$$\frac{\partial f}{\partial x}(0, 0) = 0.$$

Remark. In the previous example, we can similarly conclude that: $\frac{\partial f}{\partial y}(0, 0) = 0$.

Also, it may be shown that f is not continuous at $(0, 0)$ (**exercise**).

Hence, in general, the existence of $\frac{\partial f}{\partial x}$ and $\frac{\partial f}{\partial y}$ at a point P does **not** imply that f is continuous at P.

5.5 Higher Order Partial Derivatives

Since, $\dfrac{\partial f}{\partial x_i}$ is itself a function in n variables, we can consider its partial derivative with respect to any of the variables x_j. We can likewise further consider partial derivatives of *that* partial derivative, and so on. The notation is as follows:

$$\frac{\partial^2 f}{\partial x_i^2} = f_{x_i x_i} := \frac{\partial}{\partial x_i}\left(\frac{\partial f}{\partial x_i}\right).$$

For $j \neq i$,

$$\frac{\partial^2 f}{\partial x_j \partial x_i} = f_{x_i x_j} := \frac{\partial}{\partial x_j}\left(\frac{\partial f}{\partial x_i}\right).$$

For $m \in \mathbb{N}$,

$$\frac{\partial^m f}{\partial x_i^m} = f\underbrace{x_i x_i \cdots x_i}_{m \text{ times}} := \frac{\partial}{\partial x_i}\left(\frac{\partial^{m-1} f}{\partial x_i^{m-1}}\right)$$

For $i_1, i_2, \ldots, i_m \in \{1, 2, 3, \ldots, n\}$,

$$\frac{\partial^m f}{\partial x_{i_m} \partial x_{i_{m-1}} \partial x_{i_{m-2}} \cdots \partial x_{i_1}} = f_{x_{i_i} x_{i_2} \cdots x_{i_m}} := \frac{\partial}{\partial x_{i_m}}\left(\frac{\partial^{m-1} f}{\partial x_{i_{m-1}} \partial x_{i_{m-2}} \cdots \partial x_{i_1}}\right).$$

Example 5.19. Find all first and second order partial derivatives of:

$$f(x, y) = x \sin y + y^2 e^{2x}$$

Solution.

$$f_x = \sin y + 2y^2 e^{2x}$$

$$f_y = x \cos y + 2ye^{2x}$$

$$f_{xx} = (f_x)_x = 4y^2 e^{2x}$$

$$f_{xy} = (f_x)_y = \cos y + 4ye^{2x}$$

$$f_{yx} = (f_y)_x = \cos y + 4ye^{2x}$$

$$f_{yy} = (f_y)_y = -x \sin y + 2e^{2x}$$

Is $f_{xy} = f_{yx}$ a coincidence?

Example 5.20. Compute $f_{xy}(0,0), f_{yx}(0,0)$, where:

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

Solution. By definition, $f_{xy} = \left(f_x\right)_y$.

So, $f_{xy}(0,0) = \lim_{k\to 0} \dfrac{f_x(0,k) - f_x(0,0)}{k}$

Need to find: $f_x(0,k)$ for $k \neq 0$ and $f_x(0,0)$ for $k \neq 0$,

$f = \dfrac{xy\left(x^2-y^2\right)}{x^2+y^2}$ near $(0,k)$.

$$f_x = \frac{\left(x^2+y^2\right)\left(3x^2y - y^3\right) - xy\left(x^2-y^2\right)(2x)}{\left(x^2+y^2\right)^2}$$

near $(0,k)$

Hence:

$$\begin{aligned} f_x(0,k) &= \frac{k^2\left(-k^3\right) - 0}{k^4} = -k \\ f_x(0,0) &= \lim_{h\to 0} \frac{f(h,0) - f(0,0)}{h} \\ &= \lim_{h\to 0} \frac{0-0}{h} = 0 \\ f_{xy}(0,0) &= \lim_{k\to 0} \frac{f_x(0,k) - f_x(0,0)}{k} \\ &= \lim_{k\to 0} \frac{-k-0}{k} = -1 \end{aligned}$$

Similar calculation gives: $f_{yx}(0,0) = 1$.

(Alternatively, note that $f(x,y) = -f(y,x)$. Hence $f_{yx}(0,0) = -f_{xy}(0,0) = 1$.)

Hence, in this example, $f_{xy}(0,0) \neq f_{yx}(0,0)$

Theorem 5.21 (Mixed Derivative Theorem)**.** *Let x and y be two of the variables of a real-valued function f in multiple variables. If f_{xy} and f_{yx} exist and are continuous on an open region containing a point $\vec{a}$, then:*

$$f_{xy}(\vec{a}) = f_{yx}(\vec{a}).$$

Proof of Mixed Derivative TheoremClairaut's Theorem. We prove the theorem for the special case where $f : A \longrightarrow \mathbb{R}$ has two variables (i.e. $A \subseteq \mathbb{R}^2$).

Without loss of generality, we may assume that $\vec{a} = (0,0) \in A$. We want to show that:

$$f_{xy}(0,0) = f_{yx}(0,0)$$

Let h, k be any positive real numbers such that $[0, h] \times [0, k] \subseteq A$. Let:

$$\alpha = (f(h, k) - f(h, 0)) - (f(0, k) - f(0, 0))$$

Let:

$$g(x) = f(x, k) - f(x, 0), \quad 0 \leq x \leq h.$$

Then:

$$\alpha = g(h) - g(0),$$

and:

$$g'(x) = f_x(x, k) - f_x(x, 0).$$

By the Mean Value Theorem , there exists $h_1 \in (0, h)$ such that:

$$\frac{\alpha}{h} = \frac{g(h) - g(0)}{h} = g'(h_1) = f_x(h_1, k) - f_x(h_1, 0).$$

By MVT again, there exists $k_1 \in (0, k)$ such that:

$$\frac{f_x(h_1, k) - f_x(h_1, 0)}{k} = f_{xy}(h_1, k_1).$$

Hence:

$$\alpha = h\left[f_x(h, k) - f_x(h, 0)\right] = hkf_{xy}(h_1, k_1).$$

Similarly, there exists $(h_2, k_2) \in (0, h) \times (0, k)$ such that:

$$\alpha = hkf_{yx}(h_2, k_2)$$

Hence, for any positive real numbers h, k sufficiently small, we have:

$$f_{xy}(h_1, k_1) = f_{yx}(h_2, k_2) \tag{5.1}$$

for some $(h_1, k_1), (h_2, k_2)$ lying the rectangle $[0, h] \times [0, k]$.

If we let $(h, k) \to (0, 0)$, then $(h_1, k_1), (h_2, k_2) \to (0, 0)$. So, from an intuitive perspective, it follows from (5.1), and the continuity of f_{xy} and f_{yx} at $(0, 0)$, that:

$$f_{xy}(0, 0) = f_{yx}(0, 0).$$

More rigorously:

Suppose $f_{xy}(0, 0) \neq f_{yx}(0, 0)$. Then, $d := |f_{xy}(0, 0) - f_{yx}(0, 0)| > 0$.

The continuity of f_{xy} and f_{yx} at $(0,0)$ implies that there exists $\delta > 0$ such that for all $(x, y) \in B_\delta(0,0)$, we have:

$$|f_{xy}(x,y) - f_{xy}(0,0)| < d/2$$

and

$$|f_{yx}(x,y) - f_{yx}(0,0)| < d/2.$$

Hence, if we take (h, k) such that $0 < \|(h,k)\| < \delta$, then, (5.1) implies that the intervals:

$$(f_{xy}(0,0) - d/2, f_{xy}(0,0) + d/2), \quad (f_{yx}(0,0) - d/2, f_{yx}(0,0) + d/2),$$

have nonempty intersection (i.e. the common value $f_{xy}(h_1, k_1) = f_{yx}(h_2, k_2)$ lies in both intervals). This contradicts the assumption that the distance d between $f_{xy}(0,0)$ and $f_{yx}(0,0)$ is nonzero. $\square$

MATH 2010 Chapter 6

6.1 Differentiability

Definition 6.1. Let $\Omega \subseteq \mathbb{R}^n$ be an open set, and $f : \Omega \longrightarrow \mathbb{R}$ a real-valued function on Ω. Let r be a non-negative integer.

The function f is said to be a C^r **function** if all partial derivatives of f up to order r exist and are continuous on Ω.

The function f is said to be a C^∞ **function** if it is C^r for any $r \geq 0$.

Example 6.2.

- f is C^0 is if it continuous.
- $f : \mathbb{R}^2 \longrightarrow \mathbb{R}$ is C^2 if:

$$f,\ f_x,\ f_y,\ f_{xx},\ f_{xy},\ f_{yx},\ f_{yy}$$

  are all defined and continuous everywhere.

Polynomials, rational functions, exponentials, logarithms, trigonometric functions, and their sum/difference/product quotient/compositions are all C^∞ functions on any open set where all partial derivatives of all orders are defined.

For example:

$$f(x,y) = e^{x^2-y} \sin \frac{x}{y}$$

is C^∞ on:

$$\Omega = \big\{(x,y) \in \mathbb{R}^2 : y \neq 0\big\}$$

Theorem 6.3 (Generalization of Clairaut's Theorem)**.** *Let r be a non-negative integer. If a function f is C^r on an open set $\Omega \subseteq \mathbb{R}^n$, then the order of differentiation does not matter for all partial derivatives of order up to r.*

Example 6.4. If $f(x,y,z)$ is C^3, then:

$$f_{xz} = f_{zx}, \quad f_{xyz} = f_{yzx} = f_{zyx}, \quad f_{xxy} = f_{xyx} = f_{yxx}$$

6.1.1 Differentiability for Functions in One Variable

Recall the following definition of differentiability in one-variable calculus: A function $f : \mathbb{R} \longrightarrow \mathbb{R}$ is differentiable at a if:

$$f'(a) := \lim_{x \to a} \frac{f(x) - f(a)}{x - a}$$

exists. In geometric terms, this means that the graph $y = f(x)$ of f "resembles" the line $y = L(x)$ with slope $f'(a)$ which passes through $(x, f(a))$:

$$f(x) \approx L(x) := f(a) + f'(a)(x - a)$$

for x "near" a.

The degree 1 polynomial $L(x)$ is called the **linear approximation** (or **linearization**) of f at $x = a$. The error of the approximation is simply the difference:

$$\varepsilon(x) = f(x) - L(x) = f(x) - f(a) - f'(a)\underbrace{(x - a)}_{\Delta x}.$$

Observe that:

$$\frac{\varepsilon(x)}{x - a} = \frac{f(x) - f(a)}{x - a} - f'(a).$$

Hence,

$$\lim_{x \to a} \frac{\varepsilon(x)}{x - a} = f'(a) - f'(a) = 0,$$

which is equivalent to:

$$\lim_{x \to a} \frac{\varepsilon(x)}{|x - a|} = 0.$$

This motivates an equivalent formulation of differentiability for functions in one variable, namely:

A real-valued function f is differentiable if there exists a line $y = L(x)$ such that the "error of approximation" $\varepsilon(x) := f(x) - L(x)$ satisfies the condition:

$$\lim_{x \to a} \frac{\varepsilon(x)}{|x - a|} = 0.$$

The benefit of such a formulation is that it readily extends to a definition of differentiability for functions in multiple variables.

Consider a function $f : \mathbb{R}^2 \longrightarrow \mathbb{R}$ is two variables. A possible formulation for the differentiability of f at (a, b) is as follows:

There exists a plane $L(x, y) = f(a, b) + C(x - a) + D(y - b)$ which well approximates $f(x, y)$ near $(x, y) = (a, b)$, in the sense that:

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

Notice that if the limit above exists, then the limit along every path towards (a, b) must be the same, in particular, fixing $y = b$ and letting $x \to a^+$:

$$\begin{aligned}
0 &= \lim_{(x,b)\to(a^+,b)} \frac{f(x,b) - L(x,y)}{\|(x-a, b-b)\|} \\
&= \lim_{x\to a^+} \frac{f(x,b) - L(x,b)}{|x-a|} \\
&= \lim_{x\to a^+} \frac{f(x,b) - L(x,b)}{x-a} \\
&= \lim_{x\to a^+} \frac{f(x,b) - f(a,b) - C(x-a)}{x-a} \\
&= \lim_{x\to a^+} \frac{f(x,b) - f(a,b)}{x-a} - C
\end{aligned}$$

This implies that:

$$\lim_{x\to a^+} \frac{f(x,b) - f(a,b)}{x-a} = C$$

and likewise:

$$\lim_{x\to a^-} \frac{f(x,b) - f(a,b)}{x-a} = C$$

Hence, for the plane L to even have a *chance* to well approximate f near (a, b), the partial derivative:

$$f_x(a,b) = \lim_{x\to a} \frac{f(x,b) - f(a,b)}{x-a}$$

must exist, and the coefficient C must be equal to $f_x(a, b)$. Similarly, $f_y(a, b)$ must exist and be equal to D.

The only candidate for a plane which well approximates f near (a, b) is therefore:

$$L(x,y) = f(a,b) + f_x(a,b)(x-a) + f_y(a,b)(y-b),$$

provided that $f_x(a, b)$ and$f_y(a, b)$ both exist.

Note that this is a *necessary* but *not sufficient* condition for f to be well approximated by a plane near (a, b).

6.1.2 Differentiability for Function in Multiple Variables

Definition 6.5. Let $\Omega \subseteq \mathbb{R}^n$ be an open set. Let $\vec{a} = (a_1, a_2, \ldots, a_n) \in \Omega$. A function $f : \Omega \longrightarrow \mathbb{R}$ is said to be **differentiable** at $\vec{a}$ if:

- Each first order partial derivative $f_{x_i}(\vec{a})$ exists, for $i = 1, 2, \ldots, n$.

- For:
$$L(\vec{x}) = f(\vec{a}) + \sum_{i=1}^{n} f_{x_i}(\vec{a})(x_i - a_i)$$
(i.e. L is the linear approximation of f at $\vec{a}$), and:
$$\varepsilon(\vec{x}) = f(\vec{x}) - L(\vec{x})$$
(i.e. "error" of the approximation), we have:
$$\lim_{\vec{x}\to\vec{a}} \frac{\varepsilon(\vec{x})}{\|\vec{x}-\vec{a}\|} = 0.$$

Remark. Observe that:

- $L(\vec{x})$ is a polynomial of degree ≥ 1.
- $L(\vec{a}) = f(\vec{a})$.
- $\dfrac{\partial L}{\partial x_i}(\vec{a}) = \dfrac{\partial f}{\partial x_i}(\vec{a})$, for $i = 1, 2, \ldots, n$.
- The graph $y = L(\vec{x})$ is the n-dimensional "tangent plane" to $y = f(\vec{x})$ at $(\vec{a}, f(\vec{a}))$, in exact analogy to the fact that $y = f(a) + f'(a)(x-a)$ is the tangent line to the graph $y = f(x)$ of a differentiable function f at $(a, f(a))$ in one-variable calculus.

Example 6.6. Let $f(x,y) = x^2 y$.

1. Show that f is differentiable at $(1,2)$.
2. Approximate $f(1.1, 1.9)$ using linearization
3. Find tangent plane of $z = f(x,y)$ at $(1, 2, f(1,2)) = (1,2,2)$.

Solution. 1. Since:
$$\frac{\partial f}{\partial x} = 2xy, \qquad \frac{\partial f}{\partial y} = x^2,$$
$$\frac{\partial f}{\partial x}(1,2) = 4, \qquad \frac{\partial f}{\partial y}(1,2) = 1,$$
the linearization of f at $(1,2)$ is:
$$\begin{aligned} L(x,y) &= f(1,2) + \frac{\partial f}{\partial x}(1,2)(x-1) + \frac{\partial f}{\partial y}(1,2)(y-2) \\ &= 2 + 4(x-1) + (y-2) \end{aligned}$$

with error term:

$$\begin{aligned}\varepsilon(x,y) &= f(x,y) - L(x,y) \\ &= x^2y - 2 - 4(x-1) - (y-2)\end{aligned}$$

To show that f is differentiable at $(1,2)$, we compute the limit:

$$\lim_{(x,y)\to(1,2)} \frac{\varepsilon(x,y)}{\|(x,y)-(1,2)\|}$$

$$= \lim_{(x,y)\to(1,2)} \frac{x^2y - 2 - 4(x-1) - (y-2)}{\sqrt{(x-1)^2 + (y-2)^2}}$$

Let $h = x - 1, k = y - 2.$

$$= \lim_{(h,k)\to(0,0)} \frac{(1+h)^2(2+k) - 2 - 4h - k}{\sqrt{h^2 + k^2}}$$

$$= \lim_{(h,k)\to(0,0)} \frac{h^2k + 2hk + 2h^2}{\sqrt{h^2 + k^2}}$$

Let $h = r\cos\theta, k = r\sin\theta$

$$= \lim_{r\to 0} \frac{r^3\cos^2\theta\sin\theta + 2r^2\cos\theta\sin\theta + 2r^2\cos^2\theta}{r}$$

$$= \lim_{r\to 0} r^2\cos^2\theta\sin\theta + 2r\cos\theta\sin\theta + 2r\cos^2\theta$$

$$= 0 \qquad \text{by Sandwich theorem}$$

Hence, f is differentiable at $(1,2)$.

2. Using the linearization L of f at $(1,2)$ we have:

$$f(1.1, 1.9) \approx L(1.1, 1.9)$$

$$= 2 + 4(1.1 - 1) + (1.9 - 2)$$

$$= 2 + 0.4 + (-0.1)$$

$$= 2.3$$

Compare: $f(1, 1, 1.9) = 2.299$.

3. The tangent plane to $z = f(x, y)$ at $(1, 1, 2)$ is:

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

$$= 2 + 4(x - 1) + (y - 2)$$

$$z = -4 + 4x + y$$

Exercise 6.7. Is $f(x, y) = \sqrt{|xy|}$ differentiable at $(0, 0)$?

Solution.

$$\frac{\partial f}{\partial x}(0,0) = \lim_{h \to 0} \frac{f(h,0) - f(0,0)}{h} = \lim_{h \to 0} \frac{0 - 0}{h} = 0$$

Similarly:

$$\frac{\partial f}{\partial y}(0,0) = 0$$

Hence,

$$L(x, y) = f(0,0) + \frac{\partial f}{\partial x}(0,0)(x - 0) + \frac{\partial f}{\partial y}(0,0)(y - 0)$$

$$= 0 + 0 + 0 = 0.$$

So, $L(x, y)$ is the zero function.

The error term is:

$$\varepsilon(x, y) = f(x, y) - L(x, y) = \sqrt{|xy|}$$

So,

$$\lim_{(x,y)\to(0,0)} \frac{\varepsilon(x,y)}{\|(x,y)-(0,0)\|} = \lim_{(x,y\to 0,0)} \frac{\sqrt{|xy|}}{\sqrt{x^2+y^2}}$$

$$= \lim_{r\to 0} \frac{\sqrt{|r^2\cos\theta\sin\theta|}}{r}$$

$$= \lim_{r\to 0} \sqrt{|\cos\theta\sin\theta|},$$

which varies with θ. Hence, the limit does not exist.

We conclude that f is not differentiable at $(0,0)$.

Theorem 6.8. *If a real-valued function f in multiple variables is differentiable at $\vec{a}$, then f is continuous at $\vec{a}$.*

Proof of Theorem 6.8. Let L be the linear approximation of f at $\vec{a}$, that is:

$$f(\vec{x}) = L(\vec{x}) + \varepsilon(\vec{x})$$

By definition of differentiability, we have:

$$\lim_{\vec{x}\to\vec{a}} \frac{\varepsilon(\vec{x})}{\|\vec{x}-\vec{a}\|} = 0$$

Hence,

$$\begin{aligned}\lim_{\vec{x}\to\vec{a}} \varepsilon(\vec{x}) &= \lim_{\vec{x}\to\vec{a}} \frac{\varepsilon(\vec{x})}{\|\vec{x}-\vec{a}\|}\cdot\|\vec{x}-\vec{a}\| \\ &= \lim_{\vec{x}\to\vec{a}} \frac{\varepsilon(\vec{x})}{\|\vec{x}-\vec{a}\|}\cdot\lim_{\vec{x}\to\vec{a}}\|\vec{x}-\vec{a}\| \\ &= 0\cdot 0 = 0.\end{aligned}$$

It now follows that:

$$\begin{aligned}\lim_{\vec{x}\to\vec{a}} f(\vec{x}) &= \lim_{\vec{x}\to\vec{a}} L(\vec{x}) + \lim_{\vec{x}\to\vec{a}} \varepsilon(\vec{x}) \\ &= \lim_{\vec{x}\to\vec{a}} f(\vec{a} + \underbrace{\lim_{\vec{x}\to\vec{a}}\left(\sum_{i=1}^{n} f_{x_i}(\vec{a})(x_i-a_i)\right)}_{=0} +0 \\ &= f(\vec{a}).\end{aligned}$$

□

MATH 2010 Chapter 7

7.1 Differentiability, Gradient

Theorem 7.1. *If $f, g : \Omega \subseteq \mathbb{R}^n \to \mathbb{R}$ are differentiable at $\vec{a} \in \Omega$. Then:*

1. *$f(\vec{x}) \pm g(\vec{x}), cf(\vec{x}), f(\vec{x})g(\vec{x})$ are differentiable at $\vec{a}$.*

2. *$\dfrac{f(\vec{x})}{g(\vec{x})}$ is differentiable at $\vec{a}$ if $g(\vec{a}) \neq 0$.*

3. *(Special case of Chain Rule) Let $h(x)$ be a one-variable function which is differentiable at $f(\vec{a})$. Then, $h \circ f$ is differentiable at $\vec{a}$.*

$$\vec{a} \longmapsto f(\vec{a}) \longmapsto (h \circ f)(\vec{a})$$

4. *Any constant function $f(\vec{x}) = c$ is differentiable.*

5. *Coordinate functions $f(\vec{x}) = x_i$ are differentiable.*

Remark. We will discuss general case of chain rule later

Proof of 1,2,3 are similar to those for one variable. (MATH 2050)
The results above give many examples of differentiable functions:

- Polynomials (Sum of products of x_i)
  e.g. $4x^3y^2 + xy^2 - xyz + z^2 \quad (\deg 5)$

- Rational functions (Quotient of polynomials) e.g. $\dfrac{x^3y + z}{x^2 + y^2 + z^2 + 1}$

- If $f(\vec{x})$ is differentiable, then the followings are differentiable:

$$e^{f(\vec{x})}, \quad \sin(f(\vec{x})), \quad \cos(f(\vec{x}))$$

$$\ln(f(\vec{x})) \quad \text{where } f(\vec{x}) > 0$$

$$\sqrt{f(\vec{x})} \quad \text{where } f(\vec{x}) > 0$$
$$\ln|f(\vec{x})| \quad \text{where } f(\vec{x}) \neq 0$$

e.g. $\dfrac{e^{\sqrt{4+\sin(x^2+xy)}}}{\ln(1+\cos(x^2y))}$ is differentiable on its domain.

Theorem 7.2. *If a function f is C^1 on an open set $\Omega \subseteq \mathbb{R}^n$, then f is differentiable on Ω.*

Remark. The theorem provides a simple way to verify differentiability if all $\dfrac{\partial f}{\partial x_i}$ can be easily shown to be continuous.

e.g. $f(x, y, z) = xe^{x+y} - \log(x + z)$. The domain of f:

$$\{(x, y, z) \in \mathbb{R}^3 : x + z > 0\}$$

is open.

$$\frac{\partial f}{\partial x} = e^{x+y} + xe^{x+y} - \frac{1}{x+z}$$
$$\frac{\partial f}{\partial y} = xe^{x+y}$$
$$\frac{\partial f}{\partial z} = -\frac{1}{x+z}$$

Hence, $\dfrac{\partial f}{\partial x}, \dfrac{\partial f}{\partial y}, \dfrac{\partial f}{\partial z}$ are all continuous on the open domain of f.

So, f is C^1, and by the theorem is it is differentiable.

Proof of Theorem 7.2. We prove the theorem for the special case where f has two variables.

Let B be an open ball centered at (a, b) such that f_x, f_y are defined on B.

For each fixed $x \in B$, viewing $f(x, y)$ as a one-variable function in y, by the MVT there exists k between b and y such that:

$$f(x, y) - f(x, b) = f_y(x, k)(y - b).$$

Likewise, for fixed $y = b$, there exists h between a and x such that:

$$f(x, b) - f(a, b) = f_x(h, b)(x - a).$$

Hence,

$$f(x, y) - f(a, b) = \underbrace{f(x, y) - f(x, b)}_{f_y(x,k)(y-b)} + \underbrace{f(x, b) - f(a, b)}_{f_x(h,b)(x-a)}$$

We have:

$$\begin{aligned}
\left|\frac{\varepsilon(x,y)}{\|(x-a,y-b)\|}\right| &= \left|\frac{f(x,y)-f(a,b)-f_x(a,b)(x-a)-f_y(a,b)(y-b)}{\|(x-a,y-b)\|}\right| \\
&= \left|\frac{[f_y(x,k)-f_y(a,b)](y-b)+[f_x(h,b)-f_x(a,b)](x-a)}{\|(x-a,y-b)\|}\right| \\
&\leq \left|\frac{[f_y(x,k)-f_y(a,b)](y-b)}{\|(x-a,y-b)\|}\right| + \left|\frac{[f_x(h,b)-f_x(a,b)](x-a)}{\|(x-a,y-b)\|}\right| \\
&\leq |f_y(x,k)-f_y(a,b)| + |f_x(h,b)-f_x(a,b)|
\end{aligned}$$

Take the limit of both sides of the above inequality as $(x,y) \to (a,b)$. Then, $(x,k),(h,b) \to (a,b)$, and by the continuity of f_x and f_y at (a,b) the right-hand side of the inequality tends to zero.

It follows that:

$$\lim_{(x,y)\to(a,b)} \frac{\varepsilon(x,y)}{\|(x-a,y-b)\|} = 0.$$

So, f is differentiable at (a,b). $\square$

7.2 Gradient and Directional derivative

Definition 7.3. Let $\Omega \subseteq \mathbb{R}^n$ be open , $\vec{a} \in \Omega, f : \Omega \to \mathbb{R}$. The **gradient**, or **gradient vector**, of f at $\vec{a}$ is:

$$\nabla f(\vec{a}) = \left(\frac{\partial f}{\partial x_1}(\vec{a}), \cdots, \frac{\partial f}{\partial x_n}(\vec{a})\right) \in \mathbb{R}^n$$

Example 7.4.

$$\begin{aligned}
f(x,y) &= x^2 + 2xy \\
\nabla f(x,y) &= (f_x, f_y) = (2x+2y, 2x) \\
\nabla f(1,2) &= (6,2)
\end{aligned}$$

Remark. Using ∇f, the linearization of f at $\vec{a}$ can be expressed as:

$$\begin{aligned}
L(\vec{x}) &= f(\vec{a}) + \sum \frac{\partial f}{\partial x_i}(\vec{a})\,(x_i - a_i) \\
&= f(\vec{a}) + \nabla f(\vec{a}) \cdot (\vec{x} - \vec{a})
\end{aligned}$$

Definition 7.5. Let $\Omega \subseteq \mathbb{R}^n$ be open, $\vec{a} \in \Omega$, $f : \Omega \to \mathbb{R}$.

Let $\vec{u} \in \mathbb{R}^n$ be an unit vector (i.e. $\|\vec{u}\| = 1$) The **directional derivative** of f in the direction of $\vec{u}$ at $\vec{a}$ is:

$$\begin{aligned}
D_{\vec{u}} f(\vec{a}) &= \lim_{t\to 0} \frac{f(\vec{a}+t\vec{u}) - f(\vec{a})}{t} \\
&= \text{ the rate of change of } f \text{ in the direction of } \vec{u} \text{ at the point } \vec{a}
\end{aligned}$$

Example 7.6. $e_2 = (0,1) \in \mathbb{R}^2$

$$\begin{aligned} D_{e_2} f(a,b) &= \lim_{t\to 0} \frac{f((a,b)+te_2) - f(a,b)}{t} \\ &= \lim_{t\to 0} \frac{f(a,b+t) - f(a,b)}{t} \\ &= \frac{\partial f}{\partial y}(a,b) \end{aligned}$$

Remark. In general, if $e_i = (0, \cdots, 0, 1, 0, \cdots 0) \in \mathbb{R}^n$ (with the i-th entry equal 1), then:

$$D_{e_i} f(\vec{a}) = \frac{\partial f}{\partial x_i}(\vec{a}).$$

Theorem 7.7. *Suppose f is differentiable at $\vec{a}$. Let $\vec{u} \in \mathbb{R}^n$ be a unit vector . Then:*

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

Remark. Recall that if $\vec{v} \neq \vec{0} \in \mathbb{R}^n$, then the unit vector $\dfrac{\vec{v}}{\|\vec{v}\|}$ is essentially the direction of $\vec{v}$.

Example 7.8. Let $f(x,y) = \arcsin\left(\dfrac{x}{y}\right)$.

Find the rate of change of f at $(1, \sqrt{2})$ in the direction of $\vec{v} = (1, -1)$.

Solution. Let $\vec{u} = \dfrac{\vec{v}}{\|\vec{v}\|} = \left(\dfrac{1}{\sqrt{2}}, -\dfrac{1}{\sqrt{2}}\right)$.

Recall: $\dfrac{d}{dz}(\arcsin z) = \dfrac{1}{\sqrt{1-z^2}}$.

Hence.

$$\frac{\partial f}{\partial x} = \frac{1}{\sqrt{1 - \left(\frac{x}{y}\right)^2}} \cdot \frac{1}{y},$$

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

Note that: $f, \dfrac{\partial f}{\partial x}, \dfrac{\partial f}{\partial y}$ are continuous near $(1, \sqrt{2})$.

Hence, f is C^1 near $(1, \sqrt{2})$

So, f is differentiable at $(1, \sqrt{2})$.

By the theorem above, it follows that:

$$\begin{aligned}
D_{\vec{u}}f(1,\sqrt{2}) &= \nabla f(1,\sqrt{2})\cdot\vec{u} \\
&= \left(\frac{\partial f}{\partial x}(1,\sqrt{2}), \frac{\partial f}{\partial y}(1,\sqrt{2})\right)\cdot\left(\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}\right) \\
&= \left(1, -\frac{1}{\sqrt{2}}\right)\cdot\left(\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}\right) \\
&= \frac{1}{\sqrt{2}} + \frac{1}{2}
\end{aligned}$$

Proof of Theorem 7.7. Suppose f is differentiable at $\vec{a}$.

Let $L(\vec{x})$ be the linearization of $f(\vec{x})$ at $\vec{a}$.

Then,

$$\begin{aligned}
f(\vec{x}) &= L(\vec{x}) + \varepsilon(\vec{x}) \\
&= f(\vec{a}) + \nabla f(\vec{a})\cdot(\vec{x}-\vec{a}) + \varepsilon(\vec{x})
\end{aligned}$$

Put $\vec{x} = \vec{a} + t\vec{u}$:

$$f(\vec{a}+t\vec{u}) = f(\vec{a}) + \nabla f(\vec{a})\cdot(t\vec{u}) + \varepsilon(\vec{a}+t\vec{u})$$

.

Then,

$$\begin{aligned}
D_{\vec{u}}f(\vec{a}) &= \lim_{t\to 0}\frac{f(\vec{a}+t\vec{u}) - f(\vec{a})}{t} \\
&= \lim_{t\to 0}\frac{\nabla f(\vec{a})\cdot(t\vec{u}) + \varepsilon(\vec{a}+t\vec{u})}{t} \\
&= \nabla f(\vec{a})\cdot\vec{u} + \lim_{t\to 0}\frac{\varepsilon(\vec{a}+t\vec{u})}{t}
\end{aligned}$$

Differentiability of f at $\vec{a}$ implies that:

$$\lim_{t\to 0}\left|\frac{\varepsilon(a+t\vec{u})}{t}\right| = \lim_{t\to 0}\frac{|\varepsilon(a+t\vec{u})|}{\|(\vec{a}+t\vec{u})-\vec{a}\|} = 0,$$

It now follows that:

$$D_{\vec{u}}f(\vec{a}) = \nabla f(\vec{a})\cdot\vec{u} + 0 = \nabla f(\vec{a})\cdot\vec{u}.$$

□

7.3 Geometric Meanings of $\nabla f(\vec{a})$

Suppose f is differentiable at $\vec{a}$ and $\|\vec{u}\| = 1$. Then:

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

By the Cauchy-Schwarz Theorem, we have:

$$|\nabla f(\vec{a}) \cdot \vec{u}| \leqslant \|\nabla f(\vec{a})\| \|\vec{u}\| = \|\nabla f(\vec{a})\|$$

Also, if $\nabla f \neq \vec{0}$, then:

$$-\|\nabla f(\vec{a})\| \leq \nabla f(\vec{a}) \cdot \vec{u} \leq \|\nabla f(\vec{a})\|,$$

where each inequality is equality if and only if $\nabla f(\vec{a})$ is parallel to $\vec{u}$.

This means that: f increases (resp. decreases) *most rapidly* in the direction of $\nabla f(\vec{a})$ (resp. $(-\nabla f(\vec{a}))$, at the rate of $\|\nabla f(\vec{a})\|$.

7.4 Properties of the Gradient

Theorem 7.9. *Let $\Omega \subseteq \mathbb{R}^n$ be open. Suppose $f, g : \Omega \subseteq \mathbb{R}^n \longrightarrow \mathbb{R}$ are differentiable, and $c \in \mathbb{R}$ is a constant. Then:*

1. $\nabla(f + g) = \nabla f + \nabla g$
2. $\nabla(cf) = c\nabla f$
3. $\nabla(fg) = g\nabla f + f\nabla g$
4. $\nabla\left(\dfrac{f}{g}\right) = \dfrac{g\nabla f - f\nabla g}{g^2}$, *provided that:* $\quad g \neq 0$

Proof of Theorem 7.9. **Exercise**. □

Remark. In the definition of $D_{\vec{u}} f(\vec{a})$, the vector $\vec{u}$ is assumed to be a unit vector. It can also be generalized to $D_{\vec{v}} f(\vec{a})$ for any vector $\vec{v}$ of any length.

In that case,

$$D_{\vec{v}} f(\vec{a}) = \lim_{t \to 0} \frac{f(\vec{a} + t\vec{v}) - f(\vec{a})}{t}$$

and $D_{\vec{v}} f(\vec{a}) = \nabla f(\vec{a}) \cdot \vec{v}$.

Note that:

$$D_{\vec{v}} f = \begin{cases} \|\vec{v}\| D_{\vec{u}} f & \text{if } \vec{v} \neq \vec{0},\ \vec{u} = \dfrac{\vec{v}}{\|\vec{v}\|} \\ 0 & \text{if } \vec{v} = \vec{0} \end{cases}$$

7.5 Total Differential

(of a real-valued function)

Let $f : \Omega \subseteq \mathbb{R}^n \to \mathbb{R}$ be differentiable at $\vec{a} \in \Omega$.

Consider linearization at $\vec{a}$:

$$f(\vec{x}) = f(\vec{a}) + \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(\vec{a})\,(x_i - a_i) + \varepsilon(\vec{x})$$

Denote:

$$\Delta f = f(\vec{x}) - f(\vec{a}), \Delta x_i = x_i - a_i$$

Then,

$$\Delta f \approx \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(\vec{a})\Delta x_i.$$

The approximation is good up to first order, since:

$$\lim_{\vec{x}\to\vec{a}} \frac{\varepsilon(\vec{x})}{\|\vec{x} - \vec{a}\|} = 0.$$

Classically, this first order approximated change is denoted by

$$df = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(\vec{a})dx_i$$

and is called the **total differential** of f at $\vec{a}$.

Example 7.10. Let $V(r,h) = \pi r^2 h$, the volume of a cylinder of radius r and height h.

Observe that V is C^1 on $\mathbb{R}^2$, hence it is differentiable everywhere.

We have:

$$\begin{aligned} dV &= \frac{\partial V}{\partial r}dr + \frac{\partial V}{\partial h}dh \\ &= 2\pi r h dr + \pi r^2 dh \end{aligned}$$

For application:

Suppose we want to approximate change of V when (r,h) changes from $(r,h) = (3,12)$ to $(3+0.08, 12-0.3)$

Let

$$dr = \Delta r = 0.08,$$

$$dh = \Delta h = -0.3$$

Then:

$$\begin{aligned}\Delta V &\approx dV \leftarrow \text{ approximated change}\\ &= 2\pi rhdr + \pi r^2 dh\\ &= 2\pi(3)(12)(0.08) + \pi(3)^2(-0.3)\\ &= 3.06\pi \approx 9.61\end{aligned}$$

7.5.1 Properties of the Total Differential

Theorem 7.11. *Suppose $f, g : \Omega \subseteq \mathbb{R}^n \to \mathbb{R}$ are differentiable and $c \in \mathbb{R}$ is a constant. Then:*

1. $d(f+g) = df + dg$
2. $d(cf) = c\,df$
3. $d(fg) = g\,df + f\,dg$
4. $d\left(\dfrac{f}{g}\right) = \dfrac{g\,df - f\,dg}{g^2}$

Proof of Theorem 7.11. **Exercise.** $\square$

7.6 Summary: Differentiating a real-valued function $f(\vec{x}) = f(x_1, \cdots, x_n)$ at $\vec{a} \in \mathbb{R}^n$

7.6.1 Different types of derivatives

- Directional derivative: $D_{\vec{u}} f(\vec{a}) = \lim_{t\to 0} \dfrac{f(\vec{a}+t\vec{u}) - f(\vec{a})}{t}$ for $\|\vec{u}\| = 1$
- Partial derivative: $\dfrac{\partial f}{\partial x_i}(\vec{a}) = D_{e_i} f(\vec{a}) \quad e_i = (0, \cdots, 0, 1, 0, \cdots, 0)$
- Gradient: $\nabla f(\vec{a}) = \left(\dfrac{\partial f}{\partial x_1}(\vec{a}), \cdots, \dfrac{\partial f}{\partial x_n}(\vec{a})\right)$
- Total differential: $df = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(\vec{a}) dx_i$
- Higher derivatives: eg $\dfrac{\partial^2 f}{\partial x_1 \partial x_2} = f_{x_2 x_1}$

f is C^k means f and all its partial derivatives up to order k exist and are continuous

7.6.2 Linear Approximation of $f(\vec{x})$ near $\vec{a}$

- $L(\vec{x}) = f(\vec{a}) + \nabla f(\vec{a}) \cdot (\vec{x} - \vec{a})$
- $f(\vec{x}) = L(\vec{x}) + \varepsilon(\vec{x})$
- f is differentiable at $\vec{a}$ if $\lim_{\vec{x}\to\vec{a}} \dfrac{\varepsilon(\vec{x})}{\|\vec{x}-\vec{a}\|} = 0 \Rightarrow df \approx \Delta f$

7.6.3 Relations among derivatives

1. $C^\infty \Rightarrow \cdots \Rightarrow C^{k+1} \Rightarrow C^k \Rightarrow \cdots \Rightarrow C^1 \Rightarrow C^0$

2.
$$\begin{gathered} f \text{ is } C^1 \text{ on an open set containing } \vec{a} \\ \Downarrow \\ f \text{ is differentiable at } \vec{a}. \\ \Downarrow \\ D_{\vec{u}} f(\vec{a}) = \nabla f(\vec{a}) \cdot \vec{u} \\ \Downarrow \\ D_{\vec{u}} f(\vec{a}) \text{ exists for any unit vector } \vec{u} \in \mathbb{R}^n \\ \Downarrow \\ \frac{\partial f}{\partial x_i}(\vec{a}) \text{ exists for } i = 1, \cdots, n \end{gathered}$$

3. All the $\Rightarrow$ in the reverse direction are false. See next slide for counter examples

Verify the following (counter-) example:

Example 7.12. $f : \mathbb{R} \longrightarrow \mathbb{R}$
$f(x) = \begin{cases} x^2 \sin \frac{1}{x} & \text{if } x \neq 0; \\ 0 & \text{if } x = 0. \end{cases}$
f is differentiable on $\mathbb{R}$ but $f'(x)$ is not continuous at $x = 0$.
Simiarly,
$g(x) = x^{2k-2} f(x)$ is k-time differentiable but $g^{(k)}(x)$ is not continuous at $x = 0$.
Hence, k-time differentiable $\nRightarrow C^k$
In particular, $C^{k-1} \nRightarrow C^k$
For a multivariable example, let: $h(\vec{x}) = g\,(x_1)$.

Example 7.13. Let:

$$f(x,y) = \begin{cases} \dfrac{xy^2}{x^2+y^4} & \text{if } (x,y) \neq (0,0). \\ 0 & \text{if } (x,y) = (0,0). \end{cases}$$

Then, $D_{\vec{u}}f(0,0)$ exists for any unit vector $\vec{u} \in \mathbb{R}^2$ but f is not continuous at $(0,0)$.

Example 7.14. Let $f(x,y) = |x+y|$.
Then, f is continuous on $\mathbb{R}^2$ but $f_x(0,0)$, $f_y(0,0)$ do not exist.

Example 7.15. Let $f(x,y) = \sqrt{|xy|}$.
Then, $f_x(0,0)$, $f_y(0,0)$ exist, but $D_{\vec{u}}f(0,0)$ does not exist for any $\vec{u} \neq \pm\vec{e}_1, \pm\vec{e}_2$

MATH 2010 Chapter 8

8.1 Matrix Multiplication

A be an $m \times n$ (m rows, n columns) matrix. Let $b = \begin{bmatrix} | \\ \vec{b} \\ | \end{bmatrix}$ be a (column) vector in $\mathbb{R}^n$.

If we view A as a collection of row vectors:

$$A = \begin{bmatrix} a_{11} & \dots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \dots & a_{mn} \end{bmatrix} = \begin{bmatrix} -\vec{a}_1- \\ \vdots \\ -\vec{a}_m- \end{bmatrix},$$

then by definition of matrix multiplication we have:

$$A\vec{b} = \begin{bmatrix} -\vec{a}_1- \\ \vdots \\ -\vec{a}_m- \end{bmatrix} \begin{bmatrix} | \\ \vec{b} \\ | \end{bmatrix} = \begin{bmatrix} \vec{a}_1 \cdot \vec{b} \\ \vdots \\ \vec{a}_m \cdot \vec{b} \end{bmatrix} \in \mathbb{R}^m$$

Now let, B be an $n \times k$ matrix. Then, view B as a collection of column vectors:

$$B = \begin{bmatrix} | & & | \\ \vec{b}_1 & \cdots & \vec{b}_k \\ | & & | \end{bmatrix},$$

we have:

$$AB = A \begin{bmatrix} | & & | \\ \vec{b}_1 & \cdots & \vec{b}_k \\ | & & | \end{bmatrix} = \begin{bmatrix} | & & | \\ A\vec{b}_1 & \cdots & A\vec{b}_k \\ | & & | \end{bmatrix}$$

Alternatively, we also have:

$$AB = \begin{bmatrix} -\vec{a}_1- \\ \vdots \\ -\vec{a}_m- \end{bmatrix} B = \begin{bmatrix} -\vec{a}_1 B- \\ \vdots \\ -\vec{a}_m B- \end{bmatrix},$$

where:

$$\vec{a}_i B = (\vec{a}_i \cdot \vec{b}_1,\ \vec{a}_i \cdot \vec{b}_2,\ \ldots,\ \vec{a}_i \cdot \vec{b}_k)$$

Example 8.1.

$$\underset{\substack{A\\ \shortparallel}}{\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}} \underset{\substack{B\\ \shortparallel}}{\begin{bmatrix} 5 & 6 & 7 \\ 8 & 9 & 10 \end{bmatrix}} = \begin{bmatrix} 21 & 24 & 27 \\ 47 & 54 & 61 \end{bmatrix}$$

$$A\begin{bmatrix} 5 \\ 8 \end{bmatrix} = \begin{bmatrix} 21 \\ 47 \end{bmatrix} \quad A\begin{bmatrix} 6 \\ 9 \end{bmatrix} = \begin{bmatrix} 24 \\ 54 \end{bmatrix} \quad A\begin{bmatrix} 7 \\ 10 \end{bmatrix} = \begin{bmatrix} 27 \\ 61 \end{bmatrix}$$

$$\begin{bmatrix} 1 & 2 \end{bmatrix} B = \begin{bmatrix} 21 & 24 & 27 \end{bmatrix} \qquad \begin{bmatrix} 3 & 4 \end{bmatrix} B = \begin{bmatrix} 47 & 54 & 61 \end{bmatrix}$$

8.2 Vector-valued Functions

Let $\vec{f} : \Omega \subseteq \mathbb{R}^n \to \mathbb{R}^m$

$$\vec{f}(\vec{x}) = \underbrace{\big(f_1(\vec{x}), \cdots, f_m(\vec{x})\big)}_{\text{vector-valued}} = \begin{bmatrix} f_1(\vec{x}) \\ \vdots \\ f_m(\vec{x}) \end{bmatrix}$$

Suppose $\frac{\partial f_i}{\partial x_j}(\vec{a})$ exists for each i, j . For each $1 \leqslant i \leqslant m$,

$$\underset{1\times 1}{f_i(\vec{x})} = \underset{1\times 1}{f_i(\vec{a})} + \underset{1\times n}{\nabla f_i(\vec{a})} \cdot \underset{n\times 1}{(\vec{x} - \vec{a})} + \underset{1\times 1}{\varepsilon_i(\vec{x})} \quad \circledast$$

Here, regard $\nabla f_i(\vec{a})$ as a row vector and $\vec{x} - \vec{a}$ as a column vector,
in order to use multiplication Writing $\circledast$ for $1 \leqslant i \leqslant m$ in a matrix:

$$\begin{bmatrix} f_1(\vec{x}) \\ \vdots \\ f_m(\vec{x}) \end{bmatrix} = \begin{bmatrix} f_1(\vec{a}) \\ \vdots \\ f_m(\vec{a}) \end{bmatrix} + \underbrace{\begin{bmatrix} -\nabla f_1(\vec{a})- \\ \vdots \\ -\nabla f_m(\vec{a})- \end{bmatrix}}_{m\times n \text{ matrix of } \frac{\partial f_i}{\partial x_j}} \begin{bmatrix} x_1 - a_1 \\ \vdots \\ x_n - a_n \end{bmatrix} + \underbrace{\begin{bmatrix} \varepsilon_1(\vec{x}) \\ \vdots \\ \varepsilon_m(\vec{x}) \end{bmatrix}}_{\text{Errors}}$$

Definition 8.2. The **Jacobian matrix** of $\vec{f}$ at $\vec{a}$ is:

$$D\vec{f}(\vec{a}) = \begin{bmatrix} -\nabla f_1(\vec{a})- \\ \vdots \\ -\nabla f_m(\vec{a})- \end{bmatrix} \qquad (m \times n \text{ matrix })$$

The **linearization** of $\vec{f}$ at $\vec{a}$ is:

$$\vec{L}(\vec{x}) = \vec{f}(\vec{a}) + D\vec{f}(\vec{a})(\vec{x} - \vec{a})$$

The function $\vec{f}$ is said to be **differentiable** at $\vec{a}$ if the **error term**:

$$\vec{\varepsilon}(\vec{x}) := \vec{f}(\vec{x}) - \vec{L}(\vec{x})$$

of the linearization $\vec{L}$ of $\vec{f}$ satisfies:

$$\lim_{\vec{x} \to \vec{a}} \frac{\|\vec{\varepsilon}(\vec{x})\|}{\|\vec{x} - \vec{a}\|} = 0.$$

Remark. 1.

$$[D\vec{f}(\vec{a})]_{ij} = \frac{\partial f_i}{\partial x_j}(\vec{a})$$

2.

$$\underset{m\times 1}{\vec{f}(\vec{x})} = \underset{m\times 1}{\vec{f}(\vec{a})} + \underset{m\times n}{D\vec{f}(\vec{a})}\underset{n\times 1}{(\vec{x} - \vec{a})} + \underset{m\times 1}{\vec{\varepsilon}(\vec{x})}$$

3. If f is real-valued $(m = 1)$, then

$$Df(\vec{a}) = \nabla f(\vec{a})$$

4. $\|\vec{\varepsilon}(\vec{x})\|$, $\|\vec{x} - \vec{a}\|$ are lengths in $\mathbb{R}^m, \mathbb{R}^n$, respectively.

5.

$$\lim_{\vec{x} \to \vec{a}} \frac{\|\vec{\varepsilon}(\vec{x})\|}{\|\vec{x} - \vec{a}\|} = 0 \Leftrightarrow \lim_{\vec{x} - \vec{a}} \frac{\varepsilon_i(\vec{x})}{\|\vec{x} - \vec{a}\|} = 0,$$

for all $i = 1, \cdots, m$.

Hence,

$$\vec{f} \text{ is differentiable at } \vec{a} \Leftrightarrow f_i \text{ is differentiable at } \vec{a},$$

for all $i = 1, \cdots, m$.

8.2.1 Approximation:

$$\vec{f}(\vec{x}) \approx L(\vec{x}) = \vec{f}(\vec{a}) + D\vec{f}(\vec{a})(\vec{x} - \vec{a})$$

$$\Rightarrow \underbrace{\vec{f}(\vec{x}) - \vec{f}(\vec{a})}_{\Delta\vec{f} = \text{ change in } f} \approx \underbrace{D\vec{f}(\vec{a})}_{\text{Jacobian Matrix}} \times \underbrace{(\vec{x} - \vec{a})}_{\Delta\vec{x} = \text{ change in } \vec{x}}$$

Can consider $D\vec{f}(\vec{a})$ as a linear map:

$$D\vec{f}(\vec{a}) : \mathbb{R}^n \longrightarrow \mathbb{R}^m$$

$$\Delta\vec{x} \longmapsto \underset{\text{approximated change in } f}{D\vec{f}(\vec{a})\Delta\vec{x} = d\vec{f}}$$

$$\underset{\text{(vector)}}{\Delta\vec{f}} \approx d\vec{f} = \underset{\text{(matrix)}}{D\vec{f}(\vec{a})} \times \underset{\text{(vector)}}{d\vec{x}}$$

Remark. Compare with $f : \mathbb{R} \to \mathbb{R}$

$$\underset{\text{(number)}}{\Delta y} \approx df = \underset{\text{(number)}}{f'(a)} \times \underset{\text{(number)}}{\Delta x}$$

Example 8.3.

$$\begin{aligned} \vec{f}(x, y) &= [(y + \overset{\overset{f_1}{\shortparallel}}{1}) \ln x, x^2 - \overset{\overset{f_2}{\shortparallel}}{\sin} y + 1] \\ &= \begin{bmatrix} (y+1)\ln x \\ x^2 - \sin y + 1 \end{bmatrix} \text{(rewrite as column vector)} \end{aligned}$$

1. Find $D\vec{f}(1, 0)$
2. Approximate $\vec{f}(0.9, 0.1)$

Solution.

$$f_1(x, y) = (y + 1) \ln x$$

$$f_2(x, y) = x^2 - \sin y + 1$$

$$\nabla f_1 = \begin{bmatrix} \dfrac{y+1}{x} & \ln x \end{bmatrix}$$

$$\nabla f_2 = \begin{bmatrix} 2x & -\cos y \end{bmatrix}$$

$$\Rightarrow D\vec{f}(x,y) = \begin{bmatrix} \frac{y+1}{x} & \ln x \\ 2x & -\cos y \end{bmatrix}$$

$$\therefore D\vec{f}(1,0) = \begin{bmatrix} 1 & 0 \\ 2 & -1 \end{bmatrix}$$

Linearization of $\vec{f}$ at (1,0):

$$\begin{aligned} \vec{L}(x,y) &= \vec{f}(1,0) + D\vec{f}(1,0)(\begin{bmatrix} x \\ y \end{bmatrix} - \begin{bmatrix} 1 \\ 0 \end{bmatrix}) \\ &= \begin{bmatrix} 0 \\ 2 \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ 2 & -1 \end{bmatrix} \begin{bmatrix} x-1 \\ y \end{bmatrix} \end{aligned}$$

$$\begin{aligned} \vec{f}(0.9,0.1) &\approx \vec{L}(0.9,0.1) \\ &= \begin{bmatrix} 0 \\ 2 \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ 2 & -1 \end{bmatrix} \underbrace{\begin{bmatrix} 0.9-1 \\ 0.1 \end{bmatrix}}_{\Delta\vec{x}=d\vec{x}=\text{change in } \vec{x}} \\ &= \begin{bmatrix} 0 \\ 2 \end{bmatrix} + \underbrace{\begin{bmatrix} -0.1 \\ -0.3 \end{bmatrix}}\, d\vec{f}=\text{approximated change of } \vec{f} \\ &= \begin{bmatrix} -0.1 \\ 1.7 \end{bmatrix} \end{aligned}$$

Remark. Actual change in $\vec{f}$:

$$\Delta\vec{f} = \vec{f}(0.9,0.1) - \vec{f}(1,0) = \begin{bmatrix} -0.1159\cdots \\ -0.2898\cdots \end{bmatrix}$$

Remark. Total differential can also be written in matrix form:

$$f : \Omega \subseteq \mathbb{R}^n \to \mathbb{R}^m$$

$$\vec{f} = \begin{bmatrix} f_1 \\ \vdots \\ f_m \end{bmatrix}$$

$$d\vec{f} = \begin{bmatrix} df_1 \\ \vdots \\ df_m \end{bmatrix} = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1} & \cdots & \frac{\partial f_m}{\partial x_n} \end{bmatrix} \begin{bmatrix} dx_1 \\ \vdots \\ dx_n \end{bmatrix} = Df(\vec{a})d\vec{x}$$

8.3 Chain Rule

Recall the **chain rule** for functions in one variable:

$$\begin{aligned} w &= g(u) = 2u + 1 \\ u &= f(x) = x^2 \end{aligned}$$

$$\begin{aligned} (g \circ f)'(x) &= g'(f(x))f'(x) \text{ or} \\ \frac{dw}{dx} &= \frac{dw}{du} \cdot \frac{du}{dx} \\ &= 2 \cdot 2x = 4x \end{aligned}$$

For multivariable functions,

Theorem 8.4 (Chain Rule)**.** *Let:*

$$\vec{f} : \Omega_1 \subseteq \mathbb{R}^k \longrightarrow \mathbb{R}^n$$

$$\vec{g} : \Omega_2 \subseteq \mathbb{R}^n \longrightarrow \mathbb{R}^m$$

Suppose that $\vec{f}$ is differentiable at $\vec{a}$, and $\vec{g}$ is differentiable at $\vec{b} = \vec{f}(\vec{a})$.

Then, $\vec{g} \circ \vec{f}$ is differentiable at $\vec{a}$, with:

$$\underbrace{D(\vec{g} \circ \vec{f})}_{m \times k}(\vec{a}) = \underbrace{(D\vec{g})}_{m \times n}(f(\vec{a}))\underbrace{(D\vec{f})}_{n \times k}(\vec{a})$$

Remark. For simplicity, we might omit $\rightarrow$ for vectors

From now on: $\vec{f} = f$, $\vec{x} = x$

Example 8.5. Let:

$$f : \mathbb{R} \to \mathbb{R}^2,$$

$$g : \mathbb{R}^2 \to \mathbb{R}^2,$$

where:

$$f(\theta) = (\cos\theta, \sin\theta) = \begin{bmatrix} \cos\theta \\ \sin\theta \end{bmatrix}$$

$$g(u, v) = (2uv, u^2 - v^2) = \begin{bmatrix} 2uv \\ u^2 - v^2 \end{bmatrix}$$

Find $D(g \circ f)(\theta)$.

Solution. **Method 1** Find composition explicitly.

$$\begin{aligned}(g\circ f)(\theta) &= g(\cos\theta,\sin\theta)\\ &= \begin{bmatrix} 2\cos\theta\sin\theta \\ \cos^2\theta-\sin^2\theta\end{bmatrix}\\ &= \begin{bmatrix}\sin 2\theta\\ \cos 2\theta\end{bmatrix}\end{aligned}$$

$$\therefore D(g\circ f)(\theta) = \begin{bmatrix}\frac{d\sin 2\theta}{d\theta}\\ \frac{d\cos 2\theta}{d\theta}\end{bmatrix} = \begin{bmatrix}2\cos 2\theta\\ -2\sin 2\theta\end{bmatrix}$$

Method 2 Chain Rule

$$Df(\theta) = \begin{bmatrix} {f_1}' \\ {f_2}'\end{bmatrix} = \begin{bmatrix}-\sin\theta\\ \cos\theta\end{bmatrix}$$

$$Dg(u,v) = \begin{bmatrix}\nabla g_1\\ \nabla g_2\end{bmatrix} = \begin{bmatrix}2v & 2u\\ 2u & -2v\end{bmatrix}$$

$$Dg(f(\theta)) = Dg(\cos\theta,\sin\theta) = \begin{bmatrix}2\sin\theta & 2\cos\theta\\ 2\cos\theta & -2\sin\theta\end{bmatrix}$$

By Chain Rule,

$$\begin{aligned}D(g\circ f)(\theta) &= Dg(f(\theta))Df(\theta)\\ &= \begin{bmatrix}2\sin\theta & 2\cos\theta\\ 2\cos\theta & -2\sin\theta\end{bmatrix}\begin{bmatrix}-\sin\theta\\ \cos\theta\end{bmatrix}\\ &= \begin{bmatrix}-2\sin^2\theta+2\cos^2\theta\\ -4\cos\theta\sin\theta\end{bmatrix}\\ &= \begin{bmatrix}2\cos 2\theta\\ -2\sin 2\theta\end{bmatrix} \text{(same answer)}\end{aligned}$$

Example 8.6.

$$f(x,y) = (x^2, 3xy, x+y^2)$$

$$g(u,v,w) = \frac{uw}{v}$$

Consider $g \circ f$:

$$\begin{matrix} x \\ y \end{matrix} \xmapsto{f} \begin{matrix} f_1 = u \\ f_2 = v \\ f_3 = w \end{matrix} \xmapsto{g} g$$

Find $\frac{\partial g}{\partial x}(1,1)$.

Solution.

$$Dg = \nabla g = \begin{bmatrix} \frac{w}{v} & -\frac{uw}{v^2} & \frac{u}{v} \end{bmatrix}$$

$$\begin{aligned} Dg(f(1,1)) &= Dg(1,3,2) \\ &= \begin{bmatrix} \frac{2}{3} & -\frac{2}{9} & \frac{1}{3} \end{bmatrix} \end{aligned}$$

$$Df = \begin{bmatrix} \nabla f_1 \\ \nabla f_2 \\ \nabla f_3 \end{bmatrix} = \begin{bmatrix} 2x & 0 \\ 3y & 3x \\ 1 & 2y \end{bmatrix}$$

$$Df(1,1) = \begin{bmatrix} 2 & 0 \\ 3 & 3 \\ 1 & 2 \end{bmatrix}$$

Hence,

$$\begin{aligned} D(g \circ f)(1,1) &= Dg(f(1,1))Df(1,1) \\ &= \begin{bmatrix} \frac{2}{3} - \frac{2}{9} & \frac{1}{3} \end{bmatrix} \begin{bmatrix} 2 & 0 \\ 3 & 3 \\ 1 & 2 \end{bmatrix} \\ &= \begin{bmatrix} 1 & 0 \end{bmatrix} \end{aligned}$$

Note $D(g \circ f) = \begin{bmatrix} \frac{\partial g}{\partial x} & \frac{\partial g}{\partial y} \end{bmatrix}$
$\therefore \frac{\partial g}{\partial x}(1,1) = 1$

In the previous example, we have:

$$D(g \circ f) = Dg\, Df$$

$$\begin{bmatrix} 1 & 0 \end{bmatrix} = \begin{bmatrix} \frac{2}{3} & -\frac{2}{9} & \frac{1}{3} \end{bmatrix} \begin{bmatrix} 2 & 0 \\ 3 & 3 \\ 1 & 2 \end{bmatrix} \quad \begin{matrix} f_1 = u \\ f_2 = v \\ f_3 = w \end{matrix}$$

From matrix multiplication, we get another form of chain rule (in classical notation)

$$\frac{\partial g}{\partial x} = \frac{\partial g}{\partial u} \cdot \frac{\partial u}{\partial x} + \frac{\partial g}{\partial v} \cdot \frac{\partial v}{\partial x} + \frac{\partial g}{\partial w} \cdot \frac{\partial w}{\partial x}$$

$$\frac{\partial g}{\partial y} = \frac{\partial g}{\partial u} \cdot \frac{\partial u}{\partial y} + \frac{\partial g}{\partial v} \cdot \frac{\partial v}{\partial y} + \frac{\partial g}{\partial w} \cdot \frac{\partial w}{\partial y}$$

Example 8.7.

$$w(x, y, z) = \sqrt{x^2 + y^2 + z^2},$$

where:

$$\begin{cases} x = 3e^t \sin s \\ y = 3e^t \cos s \\ z = 4e^t \end{cases}$$

Find $\frac{\partial w}{\partial s}$ at $(s, t) = (0, 0)$.

Solution.

$$\begin{aligned} \frac{\partial w}{\partial s} &= \frac{\partial w}{\partial x} \cdot \frac{\partial x}{\partial s} + \frac{\partial w}{\partial y} \cdot \frac{\partial y}{\partial s} + \frac{\partial w}{\partial z} \cdot \frac{\partial z}{\partial s} \\ &= \frac{x}{\sqrt{x^2 + y^2 + z^2}} \cdot 3e^t \cos s - \frac{y}{\sqrt{x^2 + y^2 + z^2}} \cdot 3e^t \sin s + \frac{z}{\sqrt{x^2 + y^2 + z^2}}(0) \end{aligned}$$

$s = t = 0 \Rightarrow (x, y, z) = (0, 3, 4)$. Hence,

$$\left.\frac{\partial w}{\partial s}\right|_{(s,t)=(0,0)} = 0 - \frac{3}{5}(0) + 0 = 0.$$

Example 8.8. John is hiking with position at time t given by:

$$\begin{cases} x(t) = t^3 + 1 \\ y(t) = 2t^2 \end{cases}$$

His altitude is given by: $H(x, y) = x^2 - y^2 + 100$

1. Is John going up/down at $t = 1$?

2. Which direction should he go instead at $t = 1$ to go down most quickly?

Solution. 1. Find $\frac{dH}{dt}\big|_{t=1}$:

$$\begin{aligned}\frac{dH}{dt} &= \frac{\partial H}{\partial x}\cdot\frac{dx}{dt}+\frac{\partial H}{\partial y}\cdot\frac{dy}{dt}\\ &= (2x)(3t^2)+(-2y)(4t)\\ &= 2(t^3+1)(3t^2)-2(2t^2)(4t)\\ &= 6t^5-16t^3+6t^2\end{aligned}$$

$\therefore \frac{dH}{dt}\big|_{t=1} = 6-16+6=-4<0$

$\therefore$ John is going downhill at $t=1$.

2. At $t=1, (x,y)=(2,2)$

$$\nabla H = (2x,-2y)$$

$$\nabla H = (4,-4)$$

$\therefore$ H decreases most rapidly in the direction of $-\nabla H(2,2)=(-4,4)$

$\therefore$ John should go NW .

Remark.

$$\frac{dH}{dt} = \overset{\text{slope in } x\text{- and } y\text{- direction}}{\overbrace{\frac{\partial H}{\partial x}\cdot\underbrace{\frac{dx}{dt}}_{\text{velocity in } x\text{- and } y\text{- direction}}+\frac{\partial H}{\partial y}\cdot\frac{dy}{dt}}} = \overset{\text{gradient}}{\nabla H}\cdot\underbrace{\begin{bmatrix}\frac{dx}{dt} & \frac{dy}{dt}\end{bmatrix}}_{\text{velocity}}$$

8.3.1 Idea of Proof of Chain Rule

Suppose

$$\begin{aligned}&f:\Omega_1\subseteq\mathbb{R}^k\longmapsto\mathbb{R}^n,\ \text{differentiable at } a\\ &g:\Omega_2\subseteq\mathbb{R}^n\longmapsto\mathbb{R}^m,\ \text{differentiable at } b=f(a)\in\Omega_2\end{aligned}$$

For

$$x\in\Omega_1,\qquad f(x)-f(a)=Df(a)(x-a)+\varepsilon_f(x) \tag{8.2}$$
$$y\in\Omega_2,\qquad g(y)-g(b)=Dg(b)(y-b)+\varepsilon_g(y) \tag{8.3}$$

Put $y=f(x), b=f(a)$ and (1) into (2):

$$\begin{aligned}g(f(x))-g(f(a)) &= Dg(f(a))[Df(a)(x-a)+\varepsilon_f(x)]+\varepsilon_g(f(x))\\ &= \underbrace{Dg(f(a))Df(a)(x-a)}_{\text{linear in } x-a}+\underbrace{Dg(f(a))\varepsilon_f(x)+\varepsilon_g(f(x))}_{\text{Denote this by } \varepsilon_{g\circ f}(x)}\end{aligned}$$

Then, show that:

$$\lim_{x \to a} \frac{\|\varepsilon_{g \circ f}(x)\|}{\|x - a\|} = 0.$$

Sketch of the argument: For x close to a, the continuity of f at a implies that $\|f(x) - f(a)\|$ is small. The differentiability of g at $f(a)$ then implies that $\varepsilon_g(f(x)))$ is small.

Similarly, the differentiability of f at a implies that $\varepsilon_f(x)$ is small. Hence, $Dg(f(a))\varepsilon_f(x)$ is small.

Hence, $g \circ f$ is differentiable at a, with:

$$D(g \circ f)(a) = Dg(f(a))\, Df(a).$$

8.3.2 Summary

Jacobian Matrix

1. $\underset{x \longmapsto f(x)}{f : \Omega \subseteq \mathbb{R} \to \mathbb{R}}$ (1-variable, real-valued)

   $Df(x) = \frac{df}{dx}$ (scalar, 1×1 matrix)

2. $\underset{x=(x_1,\cdots,x_n) \longmapsto f(x)=f(x_1,\cdots,x_n)}{f : \Omega \subseteq \mathbb{R}^n \to \mathbb{R}}$ (multivariable, real-valued)

$$\begin{aligned} Df(x) &= \nabla f(x) \\ &= \left(\frac{\partial f}{\partial x_1}(x), \cdots, \frac{\partial f}{\partial x_n}(x) \right) \quad \begin{pmatrix} \text{vectors in } \mathbb{R}^n \\ 1 \times n \text{ matrix} \end{pmatrix} \end{aligned}$$

3. $f : \Omega \subseteq \mathbb{R}^n \to \mathbb{R}^m$(multivariable, vector-valued)

$$x=(x_1,\cdots,x_n) \longmapsto \begin{bmatrix} f_1(x) \\ \vdots \\ f_m(x) \end{bmatrix} \quad f_i(x)=f_i(x_1,\cdots,x_n)$$

$$Df(x) = \begin{bmatrix} -\nabla f_1- \\ \vdots \\ -\nabla f_m- \end{bmatrix} = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1} & \cdots & \frac{\partial f_m}{\partial x_n} \end{bmatrix} (m \times n \text{ matrix})$$

Chain Rule

$$(x_1, \cdots, x_k) \overset{f}{\longmapsto} (y_1, \cdots, y_n) \overset{g}{\longmapsto} (g_1, \cdots, g_m)$$

$g_i = g_i(y_1, \cdots, y_n)$ are functions of $y_1, \cdots, y_n$

$y_j = f_j = f_j(x_1, \cdots, x_k)$ are functions of $x_1, \cdots, x_k$

$\therefore$ We can regard $g_i = g_i(x_1, \cdots, x_k)$ as functions of $x_1, \cdots, x_k$

Chain Rule in Matrix Notation

$$\underbrace{\begin{bmatrix} \frac{\partial g_1}{\partial x_1} & \cdots & \frac{\partial g_1}{\partial x_k} \\ \vdots & \cdots & \vdots \\ \frac{\partial g_m}{\partial x_1} & \cdots & \frac{\partial g_m}{\partial x_k} \end{bmatrix}}_{m\times k} = \underbrace{\begin{bmatrix} \frac{\partial g_1}{\partial y_1} & \cdots & \frac{\partial g_1}{\partial y_n} \\ \vdots & \cdots & \vdots \\ \frac{\partial g_m}{\partial y_1} & \cdots & \frac{\partial g_m}{\partial y_n} \end{bmatrix}}_{m\times n} \underbrace{\begin{bmatrix} \frac{\partial y_1}{\partial x_1} & \cdots & \frac{\partial y_1}{\partial x_k} \\ \vdots & \cdots & \vdots \\ \frac{\partial y_n}{\partial x_1} & \cdots & \frac{\partial y_n}{\partial x_k} \end{bmatrix}}_{n\times k}$$

By definition of matrix multiplication:

$$\frac{\partial g_i}{\partial x_j} = \frac{\partial g_i}{\partial y_1}\frac{\partial y_1}{\partial x_j} + \cdots + \frac{\partial g_i}{\partial y_n}\frac{\partial y_n}{\partial x_j}$$

MATH 2010 Chapter 9

9.1 Application of Chain Rule

9.1.1 Level Sets

Let

$$f : \Omega \subseteq \mathbb{R}^n \to \mathbb{R}, c \in \mathbb{R}$$

Recall that the level set of f corresponding to $c \in \mathbb{R}$ is:

$$L_c = f^{-1}(c) = \{x \in \Omega : f(x) = c\}$$

Example 9.1. Some level sets of $f(x, y) = x^2 + y^2$:

$$f^{-1}(1) = \{x^2 + y^2 = 1\}$$

$$f^{-1}(4) = \{x^2 + y^2 = 4\}$$

IFRAME
IFRAME

Theorem 9.2. *Let $f : \Omega \subseteq \mathbb{R}^n \to \mathbb{R}, \Omega$ is open,*
Let $c \in \mathbb{R}, S = f^{-1}(c)$ and $a \in S$.
Suppose f is differentiable at a, and $\nabla f(a) \neq 0$. Then, $\nabla f(a) \perp S$ at a.

Example 9.3.

$$f(x, y) = x^2 + y^2 \quad \nabla f = (2x, 2y)$$

Let $S = f^{-1}(25)$, then $(4, 3) \in S$
$\nabla f(4, 3) = (8, 6)$
IFRAME

Example 9.4. $f(x, y) = x^2 - y^2 \quad \nabla f(x, y) = (2x, -2y)$
IFRAME

Example 9.5.
$$S : x^2 + 4y^2 + 9z^2 = 22 \quad \text{(Ellipsoid)}$$
Find equation of tangent plane of S at $(3, 1, 1)$

IFRAME

Solution. Let $f(x, y, z) = x^2 + 4y^2 + 9z^2, S = f^{-1}(22)$

Also $f(3, 1, 1) = 22$, so $(3, 1, 1) \in S$

$\nabla f = (2x, 8y, 18z)$

$\nabla f(3, 1, 1) = (6, 8, 18) \perp S$ at $(3, 1, 1)$

$\therefore$ $(6, 8, 18)$ is a normal vector for the tangent plane. Equation of the tangent plane:

$$\begin{aligned}[(x, y, z) - (3, 1, 1)] \cdot (6, 8, 18) &= 0\\ 6(x - 3) + 8(y - 1) + 18(z - 1) &= 0\\ 3x + 4y + 9z &= 22\end{aligned}$$

Proof of Example 9.5. Let $r(t)$ be a curve on S , $r(0) = a$.

Then $r(t)$ on $S = f^{-1}(c)$

$$\Rightarrow f(r(t)) = c \text{ is a constant}$$

By the chain rule,

$$\nabla f(r(t)) \cdot r'(t) = \frac{df}{dt} = 0$$

Put $t = 0$, then $\nabla f(a) \cdot r'(0) = 0$

$\therefore$ $\nabla f(a) \perp$ any curve on S at a .

$\therefore$ $\nabla f(a) \perp S$ at a . $\square$

9.1.2 Implicit Differentiation

Consider the curve:
$$C : \quad x^2 + y^2 = 1$$
Find $\frac{dy}{dx}$ at $(\frac{3}{5}, -\frac{4}{5})$. Locally near $(\frac{3}{5}, -\frac{4}{5})$, we have:

$$y^2 = 1 - x^2, y < 0 \Rightarrow y = -\sqrt{1 - x^2}$$

$\therefore$ y is a function of x near $(\frac{3}{5}, -\frac{4}{5})$.

To find $\frac{dy}{dx}$ at $(\frac{3}{5}, -\frac{4}{5})$,

Method 1: Compute:
$$\frac{d}{dx}\left(-\sqrt{1 - x^2}\right)$$

Method 2: Implicit Differentiation (chain rule)

$$x^2 + y^2 = 1 \begin{pmatrix} \text{Regard } x \text{ as a variable} \\ y \text{ as a function of } x \end{pmatrix}$$

Differentiating both sides:

$$\frac{d}{dx} : 2x + 2y\frac{dy}{dx} = 0$$

Evaluating at $(x, y) = (\frac{3}{5}, -\frac{4}{5})$, we have:

$$2(\frac{3}{5}) + 2(-\frac{4}{5})\frac{dy}{dx} = 0$$

$\therefore \frac{dy}{dx}\Big|_{(\frac{3}{5},-\frac{4}{5})} = \frac{3}{4}$

Example 9.6. Consider

$$S : x^3 + z^2 + ye^{xz} + z\cos y = 0 \quad \circledast$$

Given that z can be regarded as a function $z = z(x, y)$ of independent variables x, y locally near $(0, 0, 0)$.

Find $\frac{\partial z}{\partial x}, \frac{\partial z}{\partial y}$ at $(0, 0, 0)$.

Remark. It is not easy to express z in terms of x, y .

Solution. Take $\frac{\partial}{\partial x}$ to $\circledast$,

$$3x^2 + 2z\frac{\partial z}{\partial x} + ye^{xz}(z + x\frac{\partial z}{\partial x}) + \frac{\partial z}{\partial x}\cos y = 0$$

Substitute $(x, y, z) = (0, 0, 0)$,

$$0 + 0 + 0 + \frac{\partial z}{\partial x}(1) = 0 \Rightarrow \boxed{\frac{\partial z}{\partial x}(0, 0) = 0}$$

Similarly, take $\frac{\partial}{\partial y}$ to $\circledast$

$$0 + 2z\frac{\partial z}{\partial y} + e^{xz} + ye^{xz}(x\frac{\partial z}{\partial y}) + \frac{\partial z}{\partial y}\cos y - z\sin y = 0$$

Substitute $(x, y, z) = (0, 0, 0)$, then

$$0 + 0 + 1 + 0 + \frac{\partial z}{\partial y}(1) - 0 = 0 \Rightarrow \boxed{\frac{\partial z}{\partial y}(0, 0) = -1}$$

Remark. From computations above, we have:

$$\frac{\partial z}{\partial x} = -\frac{3x^2 + yze^{xz}}{2z + xye^{xz} + \cos y}$$
$$\frac{\partial z}{\partial y} = \frac{z\sin y - e^{xz}}{2z + xye^{xz} + \cos y}$$

whenever the denominator is non-zero.

9.2 Finding Extrema (Maximum or Minimum)

Definition 9.7. Let:

$$f : A \subseteq \mathbb{R}^n \to \mathbb{R}, \quad a \in A.$$

1. The function f is said to have **global (absolute) maximum** at a if:

$$f(x) \leq f(a)$$

   for all $x \in A$.

2. The function f is said to have **local (relative) maximum** at a if:

$$f(x) \leq f(a)$$

   for all $x \in A$ near a, (i.e. There exists $\epsilon > 0$ such that $f(x) \leq f(a)$ for all $x \in A \cap B_\varepsilon(a)$.)

3. **Global (absolute) minimum** and **local (relative) minimum** are defined similarly.

Remark. Any global extremum (max/min) is also a local extremum.

A function does not necessarily have a global maximum/minimum.

Example 9.8. Let $f(x) = e^x$: on $\mathbb{R}$

$$\lim_{x \to -\infty} f(x) = 0.$$

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

But $f(x) > 0$ for all $x \in \mathbb{R}$. Hence, f has neither global maximum nor global minimum.

Example 9.9. Let $f(x) = x$ on $(-1, 1]$ (Domain is not closed).
Then f attains its global maximum at $x = 1$, but it has no global minimum.

Example 9.10. Let:

$$f : [-1, 1] \to \mathbb{R}$$

$$f(x) = \begin{cases} 1 - x & \text{if} \quad x \in (0, 1] \\ 0 & \text{if} \quad x = 0 \\ -1 - x & \text{if} \quad x \in [-1, 0) \end{cases}$$

The function f has neither global maximum nor global minimum.
(f is not continuous)

Question: When must a function have global extremum?

Theorem 9.11 (Extreme Value Theorem EVT)**.** *Let A be closed and bounded subset of $\mathbb{R}^n$. Let $f : A \to \mathbb{R}$ be a continuous function.*

Then f has a global maximum and a global minimum.

Remark.

1. A closed and bounded subset of $\mathbb{R}^n$ is said to be **compact**.
2. The theorem provides a sufficient, but not necessary, condition for the existence of global extrema.

Example 9.12. Let:

$$f : A = [0, 4] \to \mathbb{R},$$

IFRAME

Observe that A is closed and bounded, and f is continuous.

Hence, by Extreme Value Theorem (EVT), the function f has a global maximum and a global minimum on A.

Recall that in one-variable calculus: local extrema can only occur at:

1. Critical points (i.e. points a in the interior of the domain where $f'(a) = 0$ or DNE.)
2. The boundary points of the domain.

Definition 9.13. Let:

$$f : A \to \mathbb{R}, \quad a \in \mathrm{Int}(A).$$

Then, a is called a **critical point** of f if either of the following conditions holds:

1. $\nabla f(a)$ DNE (i.e. $\frac{\partial f}{\partial x_i}(a)$ DNE for some i)
2. $\nabla f(a) = \vec{0}$ (i.e. $\frac{\partial f}{\partial x_i}(a) = 0$ for all i)

Theorem 9.14. *Suppose $f : A \subseteq \mathbb{R}^n \to \mathbb{R}$ attains a local extremum at $a \in \mathrm{Int}(A)$, then a is a critical point of f.*

Proof of Theorem 9.14. Suppose f has a local extremum at $a \in \mathrm{Int}(A)$.

If $\nabla f(a)$ DNE, then a is a critical point.

If $\nabla f(a)$ exists, then all $\frac{\partial f}{\partial x_i}(a)$ exist.

For any $i = 1, \cdots, n$, let:

$$g_i(t) = f(a + te_i)$$

Note that $a \in \mathrm{Int}(A)$ implies that $g_i(t)$ is defined for t near 0.

By assumption, $g_i'(0) = \frac{\partial f}{\partial x_i}(a)$ exists.
Hence, f has a local extremum at a.
This implies that g_i has a local extremum at 0.
This in turn implies that $g_i'(0) = 0$ since by assumption $g_i'(0)$ exists.
We conclude that:

$$\frac{\partial f}{\partial x_i}(a) = 0 \quad (\text{for all } i = 1, 2, \cdots, n).$$

Hence, $\nabla f(a) = \vec{0}$. So, a is a critical point. □

9.2.1 Finding Extrema on a Bounded Region

Strategy for finding extrema:
Given:

$$f : A \subseteq \mathbb{R}^n \to \mathbb{R}.$$

To find the extrema of f:

1. Find critical points of f in $\mathrm{Int}(A)$.
2. Consider the restriction of f to the boundary ∂A of A.
   Find maximum/minimum of f on ∂A
3. Comparing values of f at points found in 1. and 2.

Example 9.15. Find global maximum/minimum of

$$f(x, y) = x^2 + 2y^2 - x + 3$$

on the region:

$$A = \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 \leq 1\}$$

Remark. The region A is is closed and bounded.
Moreover, since f is polynomial, it is continuous.
So, by Extreme Value Theorem (EVT), the function f has global maximum and minimum on A.

Solution. We follow the strategy above:
Step 1 Consider the critical points of f in $\mathrm{Int}(A)$:
First, notice that $\nabla f = (2x - 1, 4y)$ exists everywhere. Moreover:

$$\begin{aligned} \nabla f = \vec{0} \;&\Leftrightarrow\; \begin{cases} 2x - 1 = 0 \\ 4y = 0 \end{cases} \\ &\Leftrightarrow\; (x, y) = \left(\frac{1}{2}, 0\right) \end{aligned}$$

Also, $(\frac{1}{2}, 0)$ lies in $\text{Int}(A) = \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 < 1\}$.

So, we conclude that f has only one critical point $\left(\frac{1}{2}, 0\right)$ in $\text{Int}(A)$, with:

$$f\left(\frac{1}{2}, 0\right) = \left(\frac{1}{2}\right)^2 + 0 - \frac{1}{2} + 3 = \frac{11}{4}$$

Step 2 Consider f on $\partial A = \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 = 1\}$.

Parametrize ∂A as follows:

$$(x, y) = (\cos\theta, \sin\theta), \theta \in [0, 2\pi]$$

$$\begin{aligned}
f(\cos\theta, \sin\theta) &= \cos^2\theta + 2\sin^2\theta - \cos\theta + 3 \\
&= \cos^2\theta + 2(1 - \cos^2\theta) - \cos\theta + 3 \\
&= -\cos^2\theta - \cos\theta + 5 \\
&= -(\cos\theta + \frac{1}{2})^2 + \frac{1}{4} + 5 \\
&= \frac{21}{4} - (\cos\theta + \frac{1}{2})^2
\end{aligned}$$

Maximum value of f on ∂A is $\frac{21}{4}$. It is attained when:

$$x = \cos\theta = -\frac{1}{2} \quad \text{i.e. } (x, y) = (-\frac{1}{2}, \pm\frac{\sqrt{3}}{2})$$

Minimum value of f on ∂A is 3. It is attained when:

$$x = \cos\theta = 1, \quad \text{i.e. } (x, y) = (1, 0)$$

Step 3 Reviewing the values of f at the points obtained in Steps 1 and 2, we have:

$$\begin{aligned}
f\left(\frac{1}{2}, 0\right) &= \frac{11}{4} \quad \text{(minimum)} \\
f\left(-\frac{1}{2}, \pm\frac{\sqrt{3}}{2}\right) &= \frac{21}{4} \quad \text{(maximum)} \\
f(1, 0) &= 3
\end{aligned}$$

Hence, the maximum value of f is $\frac{21}{4}$. It is attained at $\left(-\frac{1}{2}, \pm\frac{\sqrt{3}}{2}\right)$

The minimum value of f is $\frac{11}{4}$. It is attained at $\left(\frac{1}{2}, 0\right)$.

Example 9.16. Find the global extrema of

$$f(x, y) = \sqrt{x^2 + y^4} - y$$

on $R = \{(x, y) \in \mathbb{R}^2 : -1 \le x, y \le 1\}$

IFRAME

Solution. R is the square $[-1,1] \times [-1,1] \subseteq \mathbb{R}^2$

R is closed and bounded.

Also, f is continuous.

By Extreme Value Theorem (EVT), the function f has a global maximum and a global minimum.

First, observe that: $\mathrm{Int}(R) = \{(x, y) \in \mathbb{R}^2, -1 < x, y < 1\}$

Exercise : Show that $\frac{\partial f}{\partial x}(0,0)$ DNE. (Hint: $(f(x,0)) = |x|$).)

For$(x, y) \neq (0,0)$, the gradient ∇f exists, with:

$$\nabla f = \left(\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}\right) = \left(\frac{x}{\sqrt{x^2+y^4}}, \frac{2y^3}{\sqrt{x^2+y^4}} - 1\right)$$

Hence:

$$\nabla f = (0,0) \Leftrightarrow \begin{cases} \frac{x}{\sqrt{x^2+y^4}} = 0 \\ \frac{2y^3}{\sqrt{x^2+y^4}} - 1 = 0 \end{cases}$$

Hence, $\nabla f(x, y) = \vec{0}$ if and only if: $x = 0$, and

$$\frac{2y^3}{y^2} - 1 = 0,$$

which holds if and only if $y = \frac{1}{2}$.

Therefore, f has two critical points in $\mathrm{Int}(R)$:

$$\underbrace{(0,0)}_{\nabla f \text{ DNE}}, \quad \underbrace{\left(0, \frac{1}{2}\right)}_{\nabla f = \vec{0}}$$

Note that:

$$f(0,0) = 0, \quad f\left(0, \frac{1}{2}\right) = -\frac{1}{4}$$

Consider f on ∂R:

$$f(x, y) = \sqrt{x^2+y^4} - y$$

$$\partial R = \{(x, y) : |x| = 1, -1 \leq y \leq 1\} \cup \{(x, y) : |y| = 1, -1 \leq x \leq 1\}$$

Consider different parts of ∂R:

1. $y = 1, -1 \leq x \leq 1$

$$f(x,1) = \sqrt{x^2+1} - 1 \Rightarrow 0 \leq f \leq \sqrt{2} - 1$$

2. $y = -1, -1 \leq x \leq 1$

$$f(x,-1) = \sqrt{x^2+1} + 1 \Rightarrow 2 \leq f \leq \sqrt{2} + 1$$

3. $|x| = 1, -1 \leq y \leq 1$

$$f(x,y) = \sqrt{1+y^4} - y.$$

   If $-1 \leq y \leq 1$, then $1 \leq \sqrt{1+y^4} \leq \sqrt{2}$, and $-1 \leq -y \leq 1$. Hence:

$$0 = 1 - 1 \leq \sqrt{1+y^4} - y \leq \sqrt{2} + 1.$$

   Restricted to $C = \{(x,y) | \, |x| = 1, -1 \leq y \leq 1\}$, the maximum value of $f(x,y)$ is therefore $f(\pm 1, -1) = \sqrt{2} + 1$.

   Since we already know that $f(0,1) = 0$, which is less than all possible values of f restricted to C. The exact minimum of f on C is of little interest to us.

Hence, on ∂R, the function f has a minimum value of 0 at $(0,1)$, and a maximum value of $\sqrt{2}+1$ at $(\pm 1, -1)$.

Comparing values of f at points obtained in Steps 1 and 2:

$$\begin{aligned} f(0,0) &= 0 \\ f(0,\frac{1}{2}) &= -\frac{1}{4} \quad \text{(minimum)} \\ f(0,1) &= 0 \\ f(\pm 1, -1) &= \sqrt{2} + 1 \quad \text{(maximum)} \\ f(\pm 1, 1) &= \sqrt{2} - 1, \end{aligned}$$

we conclude that the maximum value of f is $\sqrt{2}+1$, attained at $(\pm 1, -1)$, and the minimum value is $-\frac{1}{4}$, attained at $(0, \frac{1}{2})$.

9.2.2 Finding Extrema on an Unbounded Region

Example 9.17. Find the global extrema of

$$f(x,y) = x^2 + y^2 - 4x + 6y + 7$$

on $\mathbb{R}^2$.

Remark. $\mathbb{R}^2$ is not bounded. So f might not have global extrema. Observe that:

$$\underbrace{\lim_{(x,y)\to\infty}}\ f(x,y) = +\infty$$

"(x, y) are far away from origin."

Hence,

1. f has no global maximum on $\mathbb{R}^2$
2. Strategy for finding global minimum

   Find a closed and bounded region A such that f is "large enough" outside R. Then, the minimum of f on A is equal to the minimum of f of $\mathbb{R}$.

$$\text{min on } R = \text{ min on } \mathbb{R}^2$$

Solution. Find the critical points of f.

$$\nabla f = (\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}) = (2x - 4, 2y + 6)$$

is defined everywhere on $\mathbb{R}^2$.

$$\begin{aligned}\nabla f = (0,0) &\Leftrightarrow \begin{cases} 2x - 4 = 0 \\ 2y + 6 = 0 \end{cases} \\ &\Leftrightarrow (x, y) = (2, -3)\end{aligned}$$

Hence, the function f has only one critical point $(2, -3)$, with $f(2, -3) = -6$.
We want to show that f has a global minimum at $(2, -3)$:
For $(x, y) \in \mathbb{R}^2$, let $r = \sqrt{x^2 + y^2}$

$$\begin{aligned}\text{Then } f(x,y) &= x^2 + y^2 - 4x + 6y + 7 \\ &\geq r^2 - 4r - 6r + 7 \\ &= r(r - 10) + 7.\end{aligned}$$

This is because:

$$r = \sqrt{x^2 + y^2} \geq |x|, |y|,$$

which implies:

$$\begin{cases} x \leq r \Rightarrow -4x \geq -4r \\ -y \leq r \Rightarrow 6y \geq -6r \end{cases}$$

Hence, if $\sqrt{x^2+y^2} = r \geq 10$, then $f(x,y) \geq 7 > f(2,-3)$.

Let $A = \overline{B_{10}(0,0)}$. Let $f|_R$ denote the restriction of f on R.

By Extreme Value Theorem (EVT) , the function $f|_R$ has global a minimum.

In $\operatorname{Int}(R)$, the point $(2,-3)$ is the only critical point, with:

$$f(2,-3) = -6$$

On ∂A, we have $f(x,y) \geq 7 > f(2,-3)$. Hence, $f|_R$ has a global minimum at $(2,-3)$.

For $(x,y) \notin R$, we have $f(x,y) \geq 7 > f(2,-3)$. Hence, f has no global maximum, but it has a global minimum value of -6 at $(2,-3)$.

Remark.

1. It is in fact easier to solve this problem using elementary algebra:

   Since

   $$\begin{aligned} f(x,y) &= x^2+y^2-4x+6y+7 \\ &= (x-2)^2+(y+3)^2-6, \end{aligned}$$

   it is quite clear what the global minimum of f is.

2. A function can have neither global maximum nor global minimum.

   For example, let $g(x,y) = x^2 - y^2 - 4x + 6y + 7$

   Along the line $x = 0$, we have $g(0,y) = -y^2 + 6y + 7$. So:

   $$\lim_{y\to\pm\infty} g(0,y) = -\infty \Rightarrow \text{ no global minimum.}$$

   Along $y = 0$, we have $g(x,0) = x^2 - 4x + 7$. So:

   $$\lim_{x\to\pm\infty} g(x,0) = \infty \Rightarrow \text{ no global maximum.}$$

Another example of extrema on unbounded region.

Example 9.18. Make a box (without top) with volume $= 16$

Cost:

Base \$2/unit area

Side \$0.5/unit area

Question :

How to minimize cost?

Solution. Want to minimize

$$\begin{aligned} C(x,y) &= 2xy + (\frac{16}{xy}x + \frac{16}{xy}y)(2)(0.5) \\ &= 2xy + \frac{16}{x} + \frac{16}{y} \end{aligned}$$

on the domain $\Omega = \{(x,y) \in \mathbb{R}^2 : x, y > 0\}$

- Ω is neither closed nor bounded. Hence, Extreme Value Theorem (EVT) cannot be applied directly.
- C is large if x or y is small or large.

Strategy: Find a rectangle R such that the values of $C|_{\mathbb{R}^2\setminus R}$ are all greater than the minimum of $C|_R$.

Step 1

Find critical points

$$\nabla C = (2y - \frac{16}{x^2}, 2x - \frac{16}{y^2}) \text{ exists everywhere}$$

$$\nabla C = \vec{0} \Leftrightarrow \begin{cases} 2y - \frac{16}{x^2} = 0 \\ 2x - \frac{16}{y^2} = 0 \end{cases}$$

Hence, $y = \frac{8}{x^2}, x = \frac{8}{y^2} = \frac{8}{\frac{64}{x^4}} = \frac{x^4}{8}, x > 0 \Rightarrow x^3 = 8, x = 2, y = 2$ Hence, Only one critical point $(2,2), C(2,2) = 24$.

Step 2

Choose R s.t. $C > 24$ on ∂R and outside R.

$$C(x,y) = 2xy + \frac{16}{x} + \frac{16}{y}$$

One possible choice: $R = [0.1, 1000] \times [0.1, 1000]$

- If $x \leq 0.1$ or $y \leq 0.1$,

$$\text{then } C > \frac{16}{x} + \frac{16}{y} > \frac{16}{0.1} = 160 > 24$$

- If $(x \geq 0.1, y \geq 1000)$ or $(y \geq 0.1, x \geq 1000)$,

$$\text{then } C > 2(0.1)(1000) = 200 > 24$$

Step 3

Analysis

- R is closed and bounded, C is continuous.

  By Extreme Value Theorem (EVT), $C|_R$ has minimum.
- C has only one critical point $(2,2) \in \Omega$ $(2,2) \in \text{Int}(R), C(2,2) = 24$ $C > 24$ on $\partial R \Rightarrow C|_R$ has minimum value 24 at $(2,2)$
- $C > 24$ outside R

Hence, C has the minimum value of 24 at $(2,2)$ on Ω.

MATH 2010 Chapter 10

10.1 Taylor Series Expansion

Recall

Taylor expansion for 1-variable function $g(t)$ at $t = 0$ up to order k.

$$\boxed{g(t) = g(0) + g'(0)t + \frac{1}{2!}g''(0)t^2 + \cdots + \frac{1}{k!}g^{(k)}(0)t^k + \text{remainder} \quad \circledast}$$

We want a similar formula for a multi-variable function $f(x)$ defined near a, where $x = (x_1, \cdots, x_n), \quad a = (a_1, \cdots, a_n)$.
Let $g(t) = f(a + t(x - a))$
If $\|x - a\|$ is small, then for $|t| \leq 1$,

$$\|t(x - a)\| = |t|\|x - a\| \leq \|x - a\| \text{ is small}$$

and $g(t)$ is defined.

By $\circledast$,

$$f(a + t(x - a)) = g(0) + g'(0)t + \frac{1}{2!}g''(0)t^2 + \cdots + \frac{1}{k!}g^{(k)}(0)t^k + \text{remainder}$$

Put $t = 1$,

$$f(x) = g(0) + g'(0) + \frac{1}{2!}g''(0) + \cdots + \frac{1}{k!}g^{(k)}(0) + \text{remainder}$$

Next, express $g^{(k)}(0)$ in terms of f :

$$g(0) = f(a + t(x - a)) = f(a)$$

$$\begin{aligned} g'(t) &= \nabla f(a + t(x - a)) \cdot \frac{d}{dt}(a + t(x - a)) \\ &= \nabla f(a + t(x - a)) \cdot (x - a) \\ &= \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(a + t(x - a))(x_i - a_i) \end{aligned}$$

$$
\begin{aligned}
\Rightarrow g'(0) &= \nabla f(a)\cdot(x-a)\\
&= \sum_{i=1}^{n}\frac{\partial f}{\partial x_i}(a)(x_i-a_i)
\end{aligned}
$$

$$
\begin{aligned}
g''(t) &= \frac{d}{dt}g'(t)\\
&= \sum_{i=1}^{n}\frac{d}{dt}[\frac{\partial f}{\partial x_i}(a+t(x-a))(x_i-a_i)]\\
&= \sum_{i=1}^{n}\sum_{j=1}^{n}\frac{\partial^2 f}{\partial x_j\partial x_i}(a+t(x-a))(x_j-a_j)(x_i-a_i)\\
g''(0) &= \sum_{i=1}^{n}\sum_{j=1}^{n}\frac{\partial^2 f}{\partial x_j\partial x_i}(a)(x_j-a_j)(x_i-a_i)
\end{aligned}
$$

Hence, Taylor Expansion at a up to order 2 is

$$
\boxed{f(x)=f(a)+\sum_{i=1}^{n}\frac{\partial f}{\partial x_i}(a)(x_i-a_i)+\frac{1}{2!}\sum_{i,j=1}^{n}\frac{\partial^2 f}{\partial x_i\partial x_j}(a)(x_i-a_i)(x_j-a_j)+\text{remainder}}
$$

Similarly, the general term is

$$
g^{(k)}(0)=\sum_{i_1,\cdots,i_k=1}^{n}\frac{\partial^k f}{\partial x_{i_1}\cdots\partial x_{i_k}}(a)(x_{i_1}-a_{i_1})\cdots(x_{i_k}-a_{i_k})
$$

Example 10.1. If $n=2$, i.e. $f=f(x,y), a=(x_0,y_0)$ f is C^2 (so $f_{xy}=f_{yx}$), then

$$
\begin{aligned}
f(x,y)=&f(x_0,y_0)+f_x(x_0,y_0)(x-x_0)+f_y(x_0,y_0)(y-y_0)\\
&+\frac{1}{2}[f_{xx}(x_0,y_0)(x-x_0)^2+2f_{xy}(x_0,y_0)(x-x_0)(y-y_0)+f_{yy}(x_0,y_0)(y-y_0)^2]\\
&+\text{remainder}
\end{aligned}
$$

Theorem 10.2 (Taylor's Theorem). *Let $\Omega\subseteq\mathbb{R}^n$ be open, $f:\Omega\to\mathbb{R}$ be C^k.*

Then for any $x,a\in\Omega$,

$$
\begin{aligned}
f(x)=&f(a)+\sum_{i=1}^{n}\frac{\partial f}{\partial x_i}(a)(x_i-a_i)+\frac{1}{2!}\sum_{i,j=1}^{n}\frac{\partial^2 f}{\partial x_i\partial x_j}(a)(x_i-a_i)(x_j-a_j)+\cdots\\
&+\frac{1}{k!}\sum_{i_1,\cdots,i_k=1}^{n}\frac{\partial^k f}{\partial x_{i_1}\cdots\partial x_{i_k}}(a)(x_{i_1}-a_{i_1})\cdots(x_{i_k}-a_{i_k})+\varepsilon_k(x,a)\qquad,
\end{aligned}
$$

with:

$$\lim_{x\to a}\frac{\varepsilon_k(x,a)}{\|x-a\|^k}=0$$

Definition 10.3.

$$p_k(x)=f(a)+\sum_{i=1}^{n}\frac{\partial f}{\partial x_i}(a)(x_i-a_i)+\cdots$$
$$+\frac{1}{k!}\sum_{i_1,\cdots,i_k=1}^{n}\frac{\partial^k f}{\partial x_{i_1}\cdots\partial x_{i_k}}(a)(x_{i_1}-a_{i_1})\cdots(x_{i_k}-a_{i_k})$$

is called the k**-th order Taylor polynomial** of f at a.

Remark.

- $p_1(x)=L(x)=$ Linearization of f at a
- p_k and f have equal partial derivatives up to order k at a.

IFRAME
open in new window

Example 10.4. $f(x,y)=e^x\cos y$ Find the 2^{nd} order Taylor polynomial at $a=(0,0)$

Solution.

$$\begin{aligned} f_x &= e^x\cos y & f_y &= -e^x\sin y\\ f_{xx} &= e^x\cos y & f_{yx} &= -e^x\sin y\\ f_{xy} &= -e^x\sin y & f_{yy} &= -e^x\cos y \end{aligned}$$

$\Rightarrow \quad f(0,0)=1$,

$$\begin{aligned} f_x(0,0) &= 1 & f_y(0,0) &= 0\\ f_{xx}(0,0) &= 1 & f_{yy}(0,0) &= -1\\ f_{xy}(0,0) &= f_{yx}(0,0)=0 \end{aligned}$$

$$\begin{aligned} p_2(x,y) &= f(0,0)+f_x(0,0)x+f_y(0,0)y+\frac{1}{2!}(f_{xx}(0,0)x^2+2f_{xy}(0,0)xy+f_{yy}(0,0)y^2)\\ &= 1+x+\frac{1}{2}x^2-\frac{1}{2}y^2 \end{aligned}$$

How about $p_3(x,y)$ at $(0,0)$?

$$p_3(x,y) = p_2(x,y) + \underbrace{\frac{1}{3!}g^{(3)}(0)}_{3^{rd}\text{ order terms}}$$

$$f_{xxx} = e^x \cos y \qquad f_{xxy} = f_{xyx} = f_{yxx} = -e^x \sin y$$
$$f_{yyy} = e^x \sin y \qquad f_{xyy} = f_{yxy} = f_{yyx} = -e^x \cos y$$

$$\Rightarrow f_{xxx}(0,0) = 1 \qquad f_{xxy}(0,0) = 0$$
$$f_{xyy}(0,0) = -1 \qquad f_{yyy}(0,0) = 0$$

$$\begin{aligned} g^{(3)}(0) &= f_{xxx}(0,0)x^3 + 3f_{xxy}(0,0)x^2y + 3f_{xyy}(0,0)xy^2 + f_{yyy}(0,0)y^3 \\ &= x^3 - 3xy^2 \end{aligned}$$

$$\begin{aligned} p_3(x,y) &= p_2(x,y) + \frac{1}{3!}(x^3 - 3xy^2) \\ &= 1 + x + \frac{1}{2}x^2 - \frac{1}{2}y^2 + \frac{1}{6}x^3 - \frac{1}{2}xy^2 \end{aligned}$$

Question If $f = f(x,y,z)$ is C^6 , then coefficient of xy^2z^3 in $p_6(x,y,z)$ at $(0,0,0)$ is $\alpha f_{xyyzzz}(0,0,0), \alpha =$?

10.1.1 Matrix form for 2nd order Taylor Polynomial

Definition 10.5. Let $\Omega \subseteq \mathbb{R}^n$ be open, $f : \Omega \to \mathbb{R}$ be C^2.
Then the **Hessian matrix** of f at $a \in \Omega$ is:

$$Hf(a) = \begin{bmatrix} f_{x_1x_1}(a) & \cdots & f_{x_1x_n}(a) \\ \vdots & \ddots & \vdots \\ f_{x_nx_1}(a) & \cdots & f_{x_nx_n}(a) \end{bmatrix}$$

Remark.

- $Hf(a)$ is a symmetric $n \times n$ matrix by the mixed derivatives theorem.
- In Thomas' Calculus, Hessian of f is defined to be the determinant of our Hessian matrix.

With the Hessian matrix, the 2nd order Taylor polynomial of f at a can be written as:

$$\boxed{\underset{1\times 1}{p_2(x)} = \underset{1\times 1}{f(a)} + \underset{1\times n}{\nabla f(a)}\underset{n\times 1}{(x-a)} + \frac{1}{2}\underset{1\times n}{(x-a)^\top}\underset{n\times n}{Hf(a)}\underset{n\times 1}{(x-a)}}$$

where $x, a \in \mathbb{R}^n$ are written as column vectors:

$$\begin{aligned}(x-a)^\top &= \text{Transpose of } x-a \\ &= [x_1 - a_1, \cdots, x_n - a_n]\end{aligned}$$

Remark.

$$\begin{aligned}
&(x-a)^\top Hf(a)(x-a) \\
=&[x_1 - a_1, \cdots, x_n - a_n]\begin{bmatrix} f_{x_1x_1}(a) & \cdots & f_{x_1x_n}(a) \\ \vdots & \ddots & \vdots \\ f_{x_nx_1}(a) & \cdots & f_{x_nx_n}(a)\end{bmatrix}\begin{bmatrix} x_1 - a_1 \\ \vdots \\ x_n - a_n\end{bmatrix} \\
=&[x_1 - a_1, \cdots, x_n - a_n]\begin{bmatrix} f_{x_1x_1}(a)(x_1-a_1) + \cdots + f_{x_1x_n}(a)(x_n - a_n) \\ \vdots \\ f_{x_nx_1}(a)(x_1-a_1) + \cdots + f_{x_nx_n}(a)(x_n - a_n)\end{bmatrix} \\
=&f_{x_1x_1}(a)(x_1-a_1)(x_1-a_1) + \cdots + f_{x_1x_n}(a)(x_1-a_1)(x_n-a_n) \\
&+\cdots \\
&\ \vdots \\
&+ f_{x_nx_1}(a)(x_1-a_1)(x_n-a_n) + \cdots + f_{x_nx_n}(a)(x_n-a_n)(x_n-a_n) \\
=&\sum_{i,j=1}^{n} f_{x_ix_j}(a)(x_i-a_i)(x_j-a_j) \\
=&g^{(2)}(0)
\end{aligned}$$

Example 10.6.

$$f(x,y) = e^x \cos y$$

Find $p_2(x,y)$ at $a = (0,0)$ using matrix form.

Solution.

$$f(0,0) = 1$$

$$\nabla f(0,0) = (1,0)$$

$$Hf(0,0) = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$$

$$\begin{aligned} p_2(x,y) &= f(0,0) + \nabla f(0,0) \begin{bmatrix} x-0 \\ y-0 \end{bmatrix} + \frac{1}{2}[x-0 \quad y-0] Hf(0,0) \begin{bmatrix} x-0 \\ y-0 \end{bmatrix} \\ &= 1 + [1 \quad 0] \begin{bmatrix} x \\ y \end{bmatrix} + \frac{1}{2}[x \quad y] \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} \\ &= 1 + x + \frac{1}{2}x^2 - \frac{1}{2}y^2 \end{aligned}$$

Example 10.7.

$$g(x,y) = \frac{\ln x}{1-y}$$

Find $p_2(x,y)$ at $(1,0)$.

Solution.

$$g(1,0) = 0$$

$$\nabla g = [g_x, g_y] = \left[\frac{1}{x(1-y)}, \frac{\ln x}{(1-y)^2} \right]$$

$$Hg = \begin{bmatrix} g_{xx} & g_{xy} \\ g_{yx} & g_{yy} \end{bmatrix} = \begin{bmatrix} -\frac{1}{x^2(1-y)} & \frac{1}{x(1-y)^2} \\ \frac{1}{x(1-y)^2} & \frac{2\ln x}{(1-y)^3} \end{bmatrix}$$

$$\nabla g(1,0) = [1 \quad 0] \quad Hg(1,0) = \begin{bmatrix} -1 & 1 \\ 1 & 0 \end{bmatrix}$$

$$\begin{aligned} p_2(x,y) &= g(0,0) + \nabla g(0,0) \begin{bmatrix} x-1 \\ y \end{bmatrix} + \frac{1}{2}[x-1 \quad y] Hg(1,0) \begin{bmatrix} x-1 \\ y \end{bmatrix} \\ &= 0 + [1 \quad 0] \begin{bmatrix} x-1 \\ y \end{bmatrix} + \frac{1}{2}[x-1 \quad y] \begin{bmatrix} -1 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} x-1 \\ y \end{bmatrix} \\ &= (x-1) - \frac{1}{2}(x-1)^2 + (x-1)y \end{aligned}$$

10.1.2 Application to local maximum / minimum

Suppose $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ is C^2, and a is a critical point of f.

Then, $\nabla f(a) = \vec{0}$. For x near a ,

$$\begin{aligned} f(x) &\approx p_2(x) \\ &= f(a) + \nabla f(a)(x-a) + \frac{1}{2}(x-a)^\top Hf(a)(x-a) \\ &= \quad f(a) + \underbrace{\frac{1}{2}(x-a)^\top Hf(a)(x-a)}_{\text{This term determines whether } f(x)>f(a) \text{ or } f(x)<f(a)} \end{aligned}$$

For $n = 1$, i.e. f is 1-variable.

$$\frac{1}{2}(x-a)^\top Hf(a)(x-a) = \frac{1}{2}f''(a)(x-a)^2$$

Recall: Second Derivative Test

This may be viewed as a consequence of Taylor's Theorem. That is, if $f'(a) = 0$, then near $x = a$, we have:

$$f(x) \approx f(a) + \underbrace{f'(a)(x-a)}_{=0} + \frac{1}{2}f''(a)(x-a)$$

IFRAME

The sign of the second derivative at $x = a$ essentially tells us whether locally the graph of the function looks like an upward or downward parabola.

For $n = 2$, the 2nd order term is:

$$\frac{1}{2}[x - x_0 \quad y - y_0] \underbrace{\begin{bmatrix} f_{xx}(x_0,y_0) & f_{xy}(x_0,y_0) \\ {}_{yx}(x_0,y_0) & f_{yy}(x_0,y_0) \end{bmatrix}}_{f \text{ is } C^2 \Rightarrow \text{Symmetric}} \begin{bmatrix} x - x_0 \\ y - y_0 \end{bmatrix}$$

To understand the nature of critical points, we study **quadratic forms** of 2 variables.

$$\begin{aligned} q(x,y) &= [x \quad y] \begin{bmatrix} A & B \\ B & C \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} \\ &= Ax^2 + 2Bxy + Cy^2 \end{aligned}$$

Does $q(x,y)$ have a definite sign (always positive or always negative) for $(x,y) \neq (0,0)$?

We can determine it by completing square.

Example 10.8.

$$q(x,y) = 2xy \left(= [x \;\; y] \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} \right)$$

IFRAME

Note $q(x,y) = \frac{1}{2}(x+y)^2 - \frac{1}{2}(x-y)^2$ Along $x+y=0$, i.e. $y=-x$,

$$q(x,-x) = -2x^2 < 0 \text{ for } x \neq 0$$

Along $x - y = 0$, i.e. $y = x$

$$q(x,x) = 2x^2 > 0 \text{ for } x \neq 0$$

Hence, q has no definite sign, i.e. indefinite.

Clearly $(0,0)$ is a critical point of $q(x,y)$ but neither local maximum nor minimum.

Such a critical point is called a **saddle point** .

Example 10.9.

$$q(x,y) = 17x^2 - 12xy + 8y^2 \left(= [x \;\; y] \begin{bmatrix} 17 & -6 \\ -6 & 8 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} \right)$$

IFRAME

Does $q(x,y)$ have a definite sign?

Solution.

$$\begin{aligned} q(x,y) &= 17[x^2 - \frac{2 \cdot 6}{17}xy + (\frac{6}{17})^2 y^2] + (8 - \frac{36}{17})y^2 \\ &= 17(x - \frac{6}{17}y)^2 + 10y^2 \quad \circledast \end{aligned}$$

Hence, $q(x,y) > 0 = q(0,0)$ for $(x,y) \neq (0,0)$ Hence, The critical point $(0,0)$ is a local minimum. Also global minimum of $q(x,y)$.

Remark. Expression like $\circledast$ is called diagonalization of quadratic form. It is not unique!

For example $q(x,y) = \underset{\uparrow}{5(\frac{x+2y}{\sqrt{5}})^2} + \underset{\uparrow}{20(\frac{2x-y}{\sqrt{5}})^2}$ is another diagonalization.

"Orthogonal" change of coordinates

10.1.3 Higher dimension example

Example 10.10.

$$q(x, y, z) = xy + yz + zx$$

Definite sign for $(x, y, z) \neq (0, 0, 0)$?

Solution.

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

Let $u = \frac{x+y}{2} \quad v = \frac{x-y}{2}$. Then

$$\begin{aligned}
q &= u^2 - v^2 + 2uz \\
&= (u^2 + 2uz + z^2) - v^2 - z^2 \\
&= (u + z)^2 - v^2 - z^2 \\
&= (\frac{x + y}{2} + z)^2 - (\frac{x - y}{2})^2 - z^2 \\
&= \underset{\substack{\uparrow \\ \text{positive}}}{\frac{1}{4}} \; (x + y + 2z)^2 - \frac{1}{4}(x \underset{\uparrow}{-} y)^2 \underset{\substack{\uparrow \\ \text{negative}}}{-} z^2
\end{aligned}$$

On the plane $x + y + 2z = 0$, i.e. $z = -\frac{x+y}{2}$

$$\begin{aligned}
q &= q(x, y, -\frac{x + y}{2}) \\
&= -\frac{1}{4}(x - y)^2 - \frac{1}{4}(x + y)^2 < 0 \text{ for } (x, y, z) \neq (0, 0, 0)
\end{aligned}$$

Along the line $x - y = z = 0$, i.e. $y = x, z = 0$

$$\begin{aligned}
q(x, y, z) &= q(x, x, 0) \\
&= x^2 > 0 \text{ for } x \neq 0
\end{aligned}$$

Hence, the critical point $(0, 0, 0)$ is a saddle point. For general theory, need linear algebra:

Diagonalization of quadratic form, eigenvalues $\cdots$

Definition 10.11. Let A be a $n \times n$ symmetric matrix.

Then A is said to be

- **positive definite** if $x^\top Ax > 0$ for all column vectors $x \in \mathbb{R}^n \backslash \{\vec{0}\}$

- **negative definite** if $x^\top Ax < 0$ for all column vectors $x \in \mathbb{R}^n \backslash \{\vec{0}\}$
- **indefinite** if $\exists$ column vectors $x, y \in \mathbb{R}^n \backslash \{\vec{0}\}$ such that $x^\top Ax > 0$ and $y^\top Ay < 0$

Remark. These are not all the possible cases:

There are symmetric matrix which is not positive definite, negative definite nor indefinite.

Example 10.12.

$$[x \quad y] \begin{bmatrix} 1 & 0 \\ 0 & 4 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = x^2 + 4y^2 > 0 \quad \text{for all } \begin{bmatrix} x \\ y \end{bmatrix} \in \mathbb{R}^2 \backslash \{\vec{0}\}$$

Hence, $\begin{bmatrix} 1 & 0 \\ 0 & 4 \end{bmatrix}$ is positive definite.

Example 10.13.

$$[x \quad y] \begin{bmatrix} -1 & 0 \\ 0 & -4 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = -x^2 - 4y^2 < 0 \quad \text{for all } \begin{bmatrix} x \\ y \end{bmatrix} \in \mathbb{R}^2 \backslash \{\vec{0}\}$$

Hence, $\begin{bmatrix} -1 & 0 \\ 0 & -4 \end{bmatrix}$ is negative definite.

Example 10.14.

$$[x \quad y] \begin{bmatrix} -1 & 0 \\ 0 & 4 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = -x^2 + 4y^2$$

$$[1 \quad 0] \begin{bmatrix} -1 & 0 \\ 0 & 4 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = -1 < 0$$

$$[0 \quad 1] \begin{bmatrix} -1 & 0 \\ 0 & 4 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \end{bmatrix} = 4 > 0$$

Hence, $\begin{bmatrix} -1 & 0 \\ 0 & 4 \end{bmatrix}$ is indefinite.

Example 10.15.

$$[x \quad y] \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = x^2 \geq 0 \quad \text{for all } \begin{bmatrix} x \\ y \end{bmatrix} \in \mathbb{R}^2$$

$$[0 \quad 1] \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \end{bmatrix} = 0 \Rightarrow \text{ not positive definite}$$

Hence, $\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$ is neither positive/negative definite nor indefinite.

Example 10.16.

$$\begin{aligned}
&\begin{bmatrix} x & y \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 2 & 5 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} \\
=& x^2 + 4xy + 5y^2 \\
=& (x^2 + 4xy + 4y^2) + y^2 \\
=& (x + 2y)^2 + y^2 > 0 \quad \text{for all } \begin{bmatrix} x \\ y \end{bmatrix} \in \mathbb{R}^2 \backslash \{\vec{0}\}
\end{aligned}$$

Hence, $\begin{bmatrix} 1 & 2 \\ 2 & 5 \end{bmatrix}$ is positive definite.

MATH 2010 Chapter 11

11.1 Second Derivative Test

Last time: Definiteness of symmetric matrix

Theorem 11.1. *Suppose $\Omega \subseteq \mathbb{R}^n$ is open, $f : \Omega \to \mathbb{R}$ is C^2, and $a \in \Omega$ is a critical point (i.e. $\nabla f(a) = 0$).*

If $Hf(a)$ is:

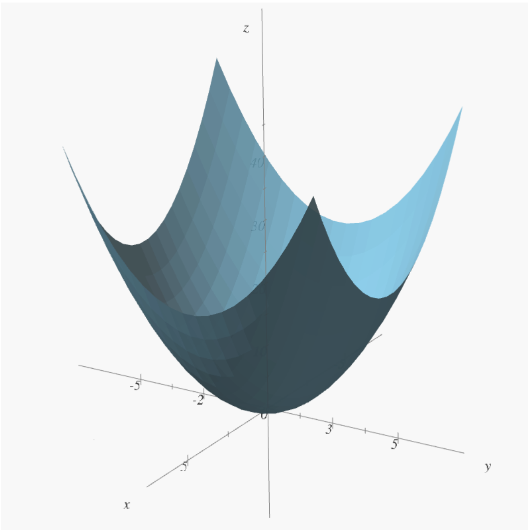
- **positive definite**, *then a corresponds to a local minimum.*
- **negative definite**, *then a corresponds to a local maximum.*
- **indefinite**, *then a is a saddle point.*

Idea of proof:
Use Taylor's Theorem.
$\nabla f(a) = 0 \Rightarrow$ For x near a,

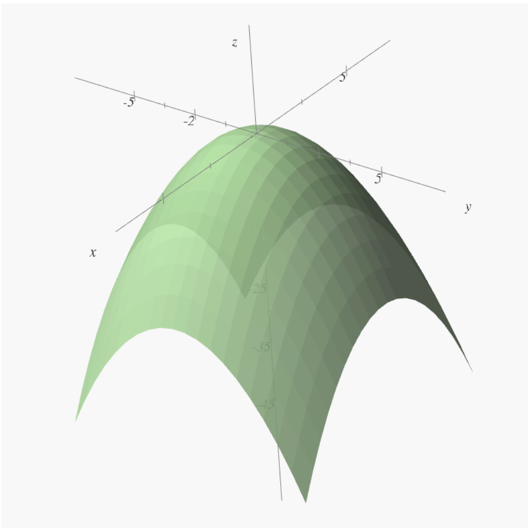
$$f(x) - f(a) \approx \frac{1}{2}(x-a)^T Hf(a)(x-a)$$

If $Hf(a)$ is positive definite
$R.H.S. > 0$ for all $x \neq a$
$\Rightarrow f(x) - f(a) > 0$ for all $x \neq a$ and near a.
$\Rightarrow f$ has a local minimum at a.
"Proof" is similar for the other two cases.
Geometrically,

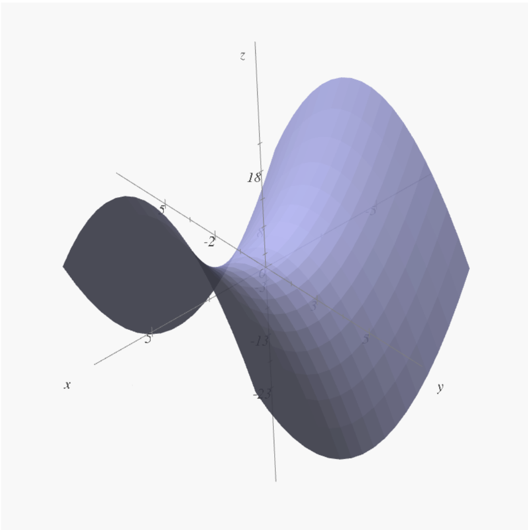
1. $Hf(a)$ is positive definite (e.g. $f = x^2 + y^2$ at $(0,0)$)



2. $Hf(a)$ is negative definite (e.g. $f = -x^2 - y^2$ at $(0,0)$)



3. $Hf(a)$ is indefinite (e.g. $f = x^2 - y^2$ at $(0,0)$)



How do we determine the definiteness of $Hf(a)$?
For the simple case $n = 2$, it can be done easily by completing square.

Theorem 11.2. *Let* $M = \begin{bmatrix} A & B \\ B & C \end{bmatrix}$ *be a symmetric* 2×2 *matrix with real coefficients. Then:*

- M *is positive definite* $\Leftrightarrow \det(M) > 0, A > 0$
- M *is negative definite* $\Leftrightarrow \det(M) > 0, A < 0$
- M *is indefinite* $\Leftrightarrow \det(M) < 0$

Remark. $\det(M) = AC - B^2$.

Proof of Theorem 11.2. Let $q(x, y) = [x \quad y]M \begin{bmatrix} x \\ y \end{bmatrix} = Ax^2 + 2Bxy + Cy^2$

Case I ($A \neq 0$)

$$\begin{aligned} Aq(x, y) &= A^2x^2 + 2ABxy + ACy^2 \\ &= (Ax + By)^2 + (AC - B^2)y^2 \end{aligned}$$

Clearly,

$$\begin{gathered} q(x, y) > 0 \quad \forall (x, y) \neq (0, 0) \Leftrightarrow AC - B^2 > 0, A > 0 \\ q(x, y) < 0 \quad \forall (x, y) \neq (0, 0) \Leftrightarrow AC - B^2 > 0, A < 0 \\ q(x, y) \text{ change signs} \Leftrightarrow AC - B^2 < 0 \end{gathered}$$

Case II ($A = 0$) $AC - B^2 = -B^2 \leq 0$

$$q(x, y) = 2Bxy + Cy^2 = y(2Bx + Cy)$$

Clearly q is neither positive or negative definite and is indefinite $\Leftrightarrow B \neq 0$
$\Leftrightarrow AC - B^2 < 0$ $\square$

Theorem 11.3 (Second Derivative Test for functions of two variables)**.** *If* $\Omega \subseteq \mathbb{R}^2$ *is open,* $f : \Omega \to \mathbb{R}$ *is* C^2*,* $a \in \Omega$*,* $\nabla f(a) = 0$*. Then,*

1. $f_{xx}f_{yy} - f_{xy}^2 > 0, f_{xx} > 0$ *at* $a \Rightarrow a$ *is a local minimum*
2. $f_{xx}f_{yy} - f_{xy}^2 > 0, f_{xx} < 0$ *at* $a \Rightarrow a$ *is a local maximum*
3. $f_{xx}f_{yy} - f_{xy}^2 < 0$ *at* $a \Rightarrow a$ *is a saddle point*
4. $f_{xx}f_{yy} - f_{xy}^2 = 0$ *at* $a \Rightarrow$ *inconclusive*

Remark.

- $f_{xx}f_{yy} - f_{xy}^2 = \det(Hf)$
- In Item 4, the point a can correspond to a local maximum/minimum or saddle point.

Example 11.4.

$$f(x, y) = 3x^2 - 10xy + 3y^2 + 2x + 2y + 3$$

Find and classify critical points of f.

Solution. f is polynomial, so is differentiable on $\mathbb{R}^2$

$$\begin{aligned} \nabla f &= [f_x \quad f_y] \\ &= [6x - 10y + 2 \quad -10x + 6y + 2] \end{aligned}$$

$$\begin{aligned} \nabla f = \vec{0} &\Leftrightarrow \begin{cases} 6x - 10y + 2 = 0 \\ -10x + 6y + 2 = 0 \end{cases} \\ &\Leftrightarrow (x, y) = (\frac{1}{2}, \frac{1}{2}) \end{aligned}$$

$\therefore (\frac{1}{2}, \frac{1}{2})$ is the only critical point.

$$Hf = \begin{bmatrix} f_{xx} & f_{xy} \\ f_{yx} & f_{yy} \end{bmatrix} = \begin{bmatrix} 6 & -10 \\ -10 & 6 \end{bmatrix}$$

$f_{xx}f_{yy} - f_{xy}^2 = (-6)^2 - (-10)^2 = -64 < 0.$
By 2^{nd} derivative test, $(\frac{1}{2}, \frac{1}{2})$ is a saddle point.

Example 11.5.

$$f(x,y) = 3x - x^3 - 3xy^2$$

Find and classify critical points of f.

Solution. f is a polynomial, so is differentiable on $\mathbb{R}^2$.

$$\begin{aligned}\nabla f &= [f_x \quad f_y] \\ &= [3 - 3x^2 - 3y^2 \quad - 6xy]\end{aligned}$$

$$\nabla f = 0$$
$$\Leftrightarrow \begin{cases} 3 - 3x^2 - 3y^2 = 0 \cdots (1) \\ -6xy = 0 \cdots (2) \end{cases}$$

$(2) \Rightarrow x = 0$ or $y = 0$
If $x = 0$, $(1) \Rightarrow 3 - 3y^2 = 0 \Rightarrow y = \pm 1$
If $y = 0$, $(1) \Rightarrow 3 - 3x^2 = 0 \Rightarrow x = \pm 1$
Hence, there are 4 critical points: $(0, \pm 1), (\pm 1, 0)$.

$$Hf = \begin{bmatrix} f_{xx} & f_{xy} \\ f_{yx} & f_{yy} \end{bmatrix} = \begin{bmatrix} -6x & -6y \\ -6y & -6x \end{bmatrix}$$

- $a = (0, 1)$

  Then, $Hf(a) = \begin{bmatrix} 0 & -6 \\ -6 & 0 \end{bmatrix}$.

  $\det Hf(a) = -36 < 0$.

  Hence, the point $(0, 1)$ corresponds to a saddle point.

- $a = (0, -1)$

  Then, $Hf(a) = \begin{bmatrix} 0 & 6 \\ 6 & 0 \end{bmatrix}$.

  $\det Hf(a) = -36 < 0$.

  Hence, the point $(0, -1)$ corresponds to a saddle point.

- $a = (1, 0)$

  Then, $Hf(a) = \begin{bmatrix} -6 & 0 \\ 0 & -6 \end{bmatrix}$.

  $\det Hf(a) = 36 > 0$.

  $f_{xx}(a) = -6 < 0$.

  Hence $(1, 0)$ corresponds to a local maximum.

- $a = (-1, 0)$

  Then, $Hf(a) = \begin{bmatrix} 6 & 0 \\ 0 & 6 \end{bmatrix}$.

  $\det Hf(a) = 36 > 0$.

  $f_{xx}(a) = 6 > 0$.

  Hence $(-1, 0)$ corresponds to a local minimum.

IFRAME

Example 11.6. Inconclusive from 2^{nd} derivative test

$$f(x,y) = x^2 + y^4 \qquad g(x,y) = x^2 - y^4 \qquad h(x,y) = -x^2 - y^4$$
$$\nabla f = [2x \quad 4y^3] \qquad \nabla g = [2x \quad -4y^3] \qquad \nabla h = [-2x \quad -4y^3]$$

$\Rightarrow$ $(0,0)$ is a critical point of f, g, h.

$$Hf = \begin{bmatrix} 2 & 0 \\ 0 & 12y^2 \end{bmatrix} \quad Hg = \begin{bmatrix} 2 & 0 \\ 0 & -12y^2 \end{bmatrix} \quad Hh = \begin{bmatrix} -2 & 0 \\ 0 & -12y^2 \end{bmatrix}$$

$$Hf(0,0) = Hg(0,0) = \begin{bmatrix} 2 & 0 \\ 0 & 0 \end{bmatrix} \quad Hh(0,0) = \begin{bmatrix} -2 & 0 \\ 0 & 0 \end{bmatrix}$$

$\Rightarrow$ Each Hessian matrix has zero determinant at $(0,0)$, so the 2^{nd} derivative test is inconclusive.

Remark. Clearly, f, g, h has local minimum, saddle point and local maximum at $(0,0)$ respectively.

11.1.1 Second Derivative Test for Functions of n Variables

Let $f : \Omega \subseteq \mathbb{R}^n \to \mathbb{R}$ be C^2, $a \in \Omega$, $\nabla f(a) = 0$.

$$Hf(a) = \begin{bmatrix} f_{x_1x_1} & f_{x_1x_2} & \cdots & f_{x_1x_n} \\ f_{x_2x_1} & f_{x_2x_2} & \cdots & f_{x_2x_n} \\ \vdots & \vdots & \ddots & \vdots \\ f_{x_nx_1} & f_{x_nx_2} & \cdots & f_{x_nx_n} \end{bmatrix}$$

f is $C^2 \Rightarrow Hf(a)$ is symmetric. From linear algebra, there exists an orthogonal $n \times n$ matrix P (i.e. $P^\top P = I_n$) such that:

$$P^T Hf(a) P = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \lambda_n \end{bmatrix}$$

where λ_i are eigenvalues of $Hf(a)$. Hence,

$$\boxed{Hf(a) \text{ is } \begin{cases} \text{positive definite} & \Leftrightarrow \quad \text{All } \lambda_i > 0 \\ \text{negative definite} & \Leftrightarrow \quad \text{All } \lambda_i < 0 \\ \text{indefinite} & \Leftrightarrow \quad \text{Some } \lambda_i > 0, \text{some } \lambda_j < 0 \end{cases}}$$

11.1.2 Another way to check definiteness of $Hf(a)$

Let H_k be the k by k submatrix

$$H_k = \begin{bmatrix} f_{x_1x_1} & f_{x_1x_2} & \cdots & f_{x_1x_k} \\ f_{x_2x_1} & f_{x_2x_2} & \cdots & f_{x_2x_k} \\ \vdots & \vdots & \ddots & \vdots \\ f_{x_kx_1} & f_{x_kx_2} & \cdots & f_{x_kx_k} \end{bmatrix}$$

1. $Hf(a)$ is positive definite $\Leftrightarrow$ $\det H_k > 0$ for $k = 1, 2, \cdots, n$
2. $Hf(a)$ is negative definite $\Leftrightarrow$ $\det H_k \begin{cases} < 0 \text{ if } k \text{ is odd} \\ > 0 \text{ if } k \text{ is even} \end{cases}$

For $n = 2$,

$$\det H_1 = \det[f_{xx}] = f_{xx}$$
$$\det H_2 = \det \begin{bmatrix} f_{xx} & f_{xy} \\ f_{yx} & f_{yy} \end{bmatrix} = f_{xx}f_{yy} - f_{xy}^2$$

Same result as before.

11.2 Lagrange Multiplier

Finding extrema under constraints.

Example 11.7. Find the point on the parabola $x^2 = 4y$ closest to $(1, 2)$.

Find minimum of $f(x, y) = (x - 1)^2 + (y - 2)^2$ under constraint $g(x, y) = x^2 - 4y = 0$.

(Constraint: expressed as a level set $g = 0$)

IFRAME

https://www.math3d.org/Z2YmkbAD

IFRAME

Theorem 11.8 (Lagrange Multipliers). *Let f, g be C^1 functions on $\Omega \subseteq \mathbb{R}^n$*

$$S = g^{-1}(c) = \{x \in \Omega : g(x) = c\}$$

Suppose

1. *a corresponds to a local extremum of f on S*

2. $\nabla g(a) \neq 0$

Then $\begin{cases} \nabla f(a) &= \lambda \nabla g(a) \text{ for some } \lambda \in \mathbb{R} \\ g(a) &= c \end{cases}$

Remark. 1. λ is called a **Lagrange Multiplier**.

2. Let $F(x, \lambda) = f(x) - \lambda(g(x) - c)$
   Then $\nabla F(x, \lambda) = (\underbrace{\nabla(f(x) - \lambda g(x))}_{n \text{ components}}, g(x) - c)$

Find critical points point of f under constraint $g = c$

$$\Updownarrow$$

Find critical point of F without constraint

Back to Example 11.7,

$$\text{Minimize } f(x, y) = (x-1)^2 + (y-2)^2$$

$$\text{Constraint } g(x, y) = x^2 - 4y = 0$$

Solution. f, g are C^1 on $\mathbb{R}^2$.

$$\nabla f = [2(x-1) \quad 2(y-2)]$$

$$\nabla g = [2x \quad -4] \neq \vec{0} \text{ on } \mathbb{R}^2$$

Suppose (x, y) is a local extremum of $f(x, y)$ on $g(x, y) = 0$.
Then, by Lagrange multipliers,

$$\begin{cases} \nabla f(x, y) &= \lambda \nabla g(x, y) \text{ for some } \lambda \in \mathbb{R} \\ g(x, y) &= 0 \end{cases}$$

$$\Rightarrow \quad \begin{cases} 2(x-1) &= 2\lambda x & \cdots (1) \\ 2(y-2) &= -4\lambda & \cdots (2) \\ x^2 - 4y &= 0 & \cdots (3) \end{cases}$$

$(1) \Rightarrow x - 1 = \lambda x \Rightarrow x(1-\lambda) = 1$

$(2) \Rightarrow y - 2 = -2\lambda \Rightarrow y = 2(1-\lambda) = \frac{2}{x}$

$(3) \Rightarrow x^2 - \frac{8}{x} = 0, x^3 - 8 = 0 \Rightarrow x = 2$

$\therefore y = \frac{2}{2} = 1$, and now it is easy to check $(x, y) = (2, 1)$ is a solution.

Geometrically, f must have a minimum on $g = 0$.

By the Lagrange Multipliers Theorem, only one point can be that minimum point.

$\Rightarrow f$ has minimum at $(2, 1)$ on $g = 0$.

To summarize, to find the minimum of $f(x, y) = (x-1)^2 + (y-2)^2$ under the constraint $g(x, y) = x^2 - 4y$, we solve:

$$\begin{cases} \nabla f &= \lambda \nabla g \\ g &= 0 \end{cases}$$

Exercise 11.9. Find the point on the parabola $x^2 = 4y$ closest to $(2, 5)$.

$f(x, y) = (x-2)^2 + (y-5)^2$

$g(x, y) = x^2 - 4y$

Remark. The system:

$$\begin{cases} \nabla f &= \lambda \nabla g \\ g &= 0 \end{cases}$$

has solutions. Global minimum on $g = 0 : (4, 4)$.

Not local extremum on $g = 0 : (-2, 1)$.

Example 11.10. Maximize xy^2 on the ellipse

$$x^2 + 4y^2 = 4$$

Solution.

$$\text{Let } f(x, y) = xy^2$$
$$g(x, y) = x^2 + 4y^2$$

Note f is continuous and the ellipse $g = 4$ is closed and bounded.

By EVT, f has global maximum and minimum on $g = 4$.

$$\nabla f = [y^2 \quad 2xy]$$

$$\nabla g = [2x \quad 8y]$$

Note: $\nabla g \neq 0$ on $x^2 + 4y^2 = 4$.

Lagrange multipliers:

$$\left\{\begin{array}{rcl}\nabla f & = & \lambda\nabla g \\ g & = & 4\end{array}\right. \Leftrightarrow \left\{\begin{array}{rcll} y^2 & = & 2\lambda x & \cdots(1) \\ 2xy & = & 8\lambda y & \cdots(2) \\ x^2+4y^2 & = & 4 & \cdots(3)\end{array}\right.$$

Case 1: If $y=0$, then

$$(3) \Rightarrow x^2 = 4 \Rightarrow x = \pm 2,$$
$$\lambda = 0 \text{ by } (1)$$

$\therefore (x,y) = (\pm 2, 0)$.

Case 2: If $y \neq 0$, then:

$$\frac{(2)}{(1)} \Rightarrow \frac{2xy}{y^2} = \frac{8\lambda y}{2\lambda x} \Rightarrow \frac{2x}{y} = \frac{4y}{x} \Rightarrow x^2 = 2y^2$$

By (3), $6y^2 = 4 \Rightarrow y = \pm\sqrt{\frac{2}{3}}$

$\therefore x^2 = 2y^2 = \frac{4}{3} \Rightarrow x = \pm\sqrt{\frac{4}{3}}$

$\therefore (x,y) = (\pm\sqrt{\frac{4}{3}}, \pm\sqrt{\frac{2}{3}})$.

Compare values of f at the 6 points found using Lagrange Multipliers:

$f(x,y) = xy^2$

$f(\pm 2, 0) = 0$

$f(\sqrt{\frac{4}{3}}, \pm\sqrt{\frac{2}{3}}) = \sqrt{\frac{4}{3}} \cdot \frac{2}{3} = \frac{4}{3\sqrt{3}}$ (maximum)

$f(-\sqrt{\frac{4}{3}}, \pm\sqrt{\frac{2}{3}}) = -\sqrt{\frac{4}{3}} \cdot \frac{2}{3} = -\frac{4}{3\sqrt{3}}$ (minimum)

Hence, for $f(x,y)$ on $g=4$,

Global maximum value $= \frac{4}{3\sqrt{3}}$ at $(\sqrt{\frac{4}{3}}, \pm\sqrt{\frac{2}{3}})$

Global minimum value $= -\frac{4}{3\sqrt{3}}$ at $(-\sqrt{\frac{4}{3}}, \pm\sqrt{\frac{2}{3}})$

Remark. We may use another form of Lagrange Multiplier.

Let $F(x,y,\lambda) = f(x,y) - \lambda(g(x,y)-4) = xy^2 - \lambda(x^2+4y^2-4)$.

Then, $\nabla F = (y^2-2\lambda x, 2xy-8\lambda y, x^2+4y^2-4)$, $\nabla F = 0 \Leftrightarrow \left\{\begin{array}{rcl} y^2-2\lambda x & = & 0 \\ 2xy-8\lambda y & = & 0 \\ x^2+4y^2-4 & = & 0\end{array}\right.$

Same system as before.

For problems of finding maximum/minimum of $f : A \to \mathbb{R}$,

Lagrange Multipliers can be used to study f on ∂A. Consider a previous example:

Example 11.11. Find global maximum/minimum of:

$$f(x,y) = x^2 + 2y^2 - x + 3 \text{ for } x^2 + y^2 \leq 1$$

Solution. Domain $= A = \{(x,y) \in \mathbb{R}^2 : x^2 + y^2 \leq 1\}$

As found before, f has only one critical point $(\frac{1}{2}, 0)$ in $\mathrm{Int}(A)$, with $f(\frac{1}{2}, 0) = \frac{11}{4}$.

To study f on $\partial A = \{(x,y) \in \mathbb{R}^2 : x^2 + y^2 = 1\}$ by Lagrange Multipliers:

Let $g(x,y) = x^2 + y^2$

$$\nabla g = (2x, 2y) \neq \vec{0} \text{ on } \partial A (g = 1)$$

$$\left\{ \begin{array}{rcl} \nabla f & = & \lambda \nabla g \\ g & = & 1 \end{array} \right. \Leftrightarrow \left\{ \begin{array}{rcll} 2x - 1 & = & 2\lambda x & \cdots (1) \\ 4y & = & 2\lambda y & \cdots (2) \\ x^2 + y^2 & = & 1 & \cdots (3) \end{array} \right.$$

$$\begin{aligned} (2) &\Rightarrow (4 - 2\lambda)y = 0 \\ &\Rightarrow \lambda = 2 \text{ or } y = 0 \end{aligned}$$

For $\lambda = 2$:

By (1),

$$\begin{aligned} 2x - 1 &= 4x \\ x &= -\frac{1}{2} \end{aligned}$$

By (3),

$$y = \pm \frac{\sqrt{3}}{2}$$

For $y = 0$,

By (3),

$$x = \pm 1$$

Comparing values of f at five points:

$$f(\frac{1}{2}, 0) = \frac{11}{4}$$
$$f(-\frac{1}{2}, \frac{\sqrt{3}}{2}) = f(-\frac{1}{2}, -\frac{\sqrt{3}}{2}) = \frac{21}{4}$$
$$f(1, 0) = 3$$
$$f(-1, 0) = 5$$

Hence, maximum value $= \frac{21}{4}$ at $(-\frac{1}{2}, \pm\frac{\sqrt{3}}{2})$ and minimum value = $\frac{11}{4}$ at $(\frac{1}{2}, 0)$.

MATH 2010 Chapter 12

Question:

When does f have global extrema subject to constraint $g = c$?

A sufficient condition:

- The level set $S = \{g = c\}$ is closed and bounded
- f is continuous on S

By EVT, f has global extrema on S.

12.1 Quadratic Constraint on 2 Variables (Conic Section)

$$g(x, y) = Ax^2 + 2Bxy + Cy^2 + Dx + Ey + F$$

Some typical examples of $g = c$:

1. $\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1, a, b > 0$ (Ellipse. Circle if $a = b$)

   IFRAME

2. $\frac{x^2}{a^2} - \frac{y^2}{b^2} = 1, a, b > 0$ (Hyperbola)

   IFRAME

   Remark. $xy = c, c \neq 0$ also a hyperbola.

   IFRAME

3. $y = ax^2, a \neq 0$ (Parabola) (only 1 quadratic term)

   IFRAME

4. Degenerate Cases

- $\frac{x^2}{a^2} + \frac{y^2}{b^2} = 0 \rightsquigarrow$ a point $(0, 0)$
- $\frac{x^2}{a^2} + \frac{y^2}{b^2} = -1 \rightsquigarrow$ empty set
- $\frac{x^2}{a^2} - \frac{y^2}{b^2} = 0 \rightsquigarrow \frac{x}{a} = \pm\frac{y}{b}$
  (a pair of intersecting lines)
- $x^2 = c \rightsquigarrow x = \pm\sqrt{c}$
  (a pair of parallel lines (double line if $c = 0$))

By a change of coordinates, any quadratic constraint $g(x, y) = c$ can be transformed to one of the forms above:

$\Rightarrow$ Ellipse, Hyperbola, Parabola, Degenerate

Each quadratic constraint corresponds to the intersection of a plane with a cone:

IFRAME

Example 12.1.

$$17x^2 - 12xy + 8y^2 + 16\sqrt{5}x - 8\sqrt{5}y = 0$$

$$\Leftrightarrow \frac{(u+1)^2}{1^2} + \frac{v^2}{2^2} = 1, \text{ where } u = \frac{2x - y}{\sqrt{5}} \quad v = \frac{x + 2y}{\sqrt{5}}$$

Remark. In the last example, u and v are chosen so that u-axis $\perp$ v-axis.

Such u and v can be found using theory of symmetric matrices in linear algebra. Among the non-degenerate quadratic constraints above, only ellipse is closed and bounded.

Any continuous $f(x, y)$ restricted to an ellipse has both global maximum and global minimum.

It is not true for hyperbola and parabola:

A continuous $f(x, y)$ restricted to a hyperbola or parabola may not have global maximum or minimum.

12.2 Quadratic Constraint for 3-variable

$$g(x, y, z) = Ax^2 + By^2 + Cz^2 + 2Pxy + 2Qyz + 2Rzx + Dx + Ey + Fz + G$$

12.2.1 Some typical examples of $g = c$

Graph $\frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} = 1, a, b, c > 0$
How to graph it?
Start with the unit sphere:

$$x^2 + y^2 + z^2 = 1$$

IFRAME
Then, stretch in x, y, z directions according to the values of a, b, c:
IFRAME
Graph $\frac{x^2}{a^2} + \frac{y^2}{b^2} - \frac{z^2}{c^2} = 1$
Up to rescaling, can assume $a = b = c = 1$

$$\rightsquigarrow x^2 + y^2 - z^2 = 1$$

Let $r = \sqrt{x^2 + y^2}$ = distance from (x, y, z) to z-axis $r^2 - z^2 = 1$ Hyperbola
IFRAME
$x^2 + y^2 - z^2 = 1 \cdots (2)$
Hyperboloid of 1 sheet
IFRAME
Graph $\frac{x^2}{a^2} + \frac{y^2}{b^2} - \frac{z^2}{c^2} = -1$

$r^2 - z^2 = -1$ Hyperbola
IFRAME
$x^2 + y^2 - z^2 = -1$ Hyperboloid of 2 sheets
IFRAME

Exercise 12.2. Graph

- $x^2 + y^2 - z^2 = 0$ (Elliptical cone)
- $z = x^2 + y^2$ (Elliptical Paraboloid)
- $z = x^2 - y^2$ (Hyperbolic Paraboloid)

12.2.2 Graph of standard quadratic surfaces

Example 12.3.

$x^2 + y^2 = 1$ Cylinder of Ellipse

IFRAME

$z = x^2$ Cylinder of parabola

IFRAME

Similar to the case of 2-variable:

Any quadratic constraint $g(x, y, z) = c$ can be transformed to one of the standard forms by a change of coordinates. Among the cases above, only ellipsoid is closed and bounded.

Any continuous $f(x, y)$ restricted to an ellipsoid has both global maximum and global minimum.

This is not the case for other quadratic surfaces.

Back to finding maximum/minimum under constraint.

Example 12.4. Find the point on the ellipse:

$$x^2 + xy + y^2 = 9$$

(**Exercise.** Show that this is indeed an ellipse.) with maximum x-coordinate.

Solution. Let $f(x, y) = x$

$$g(x, y) = x^2 + xy + y^2$$

Maximize f under constraint $g = 9$

The ellipse $g = 9$ is closed and bounded.

f is continuous. By EVT, maximum exists.

$$\nabla f = [1 \quad 0]$$

$$\nabla g = [2x + y \quad x + 2y]$$

Note $\nabla g = [0 \quad 0] \Leftrightarrow (x, y) = (0, 0)$

$(0, 0)$ is not on the ellipse. Use Lagrange Multipliers,

$$\left\{ \begin{array}{rcl} \nabla f & = & \lambda \nabla g \\ g & = & 9 \end{array} \right. \Rightarrow \left\{ \begin{array}{rcll} 1 & = & \lambda(2x + y) & \cdots (i) \\ 0 & = & \lambda(x + 2y) & \cdots (ii) \\ x^2 + xy + y^2 & = & 9 & \cdots (iii) \end{array} \right.$$

$(i) \Rightarrow \lambda \neq 0$

$\therefore (ii) \Rightarrow x + 2y = 0 \Rightarrow x = -2y \cdots (iv)$

Put (iv) into (iii) ,

$$(-2y)^2 + (-2y)y + y^2 = 9 \Rightarrow 3y^2 = 9, y = \pm\sqrt{3}$$

By (iv) , $(x, y) = (-2\sqrt{3}, \sqrt{3})$ or $(2\sqrt{3}, -\sqrt{3})$

Comparing x-coordinates, answer is $(2\sqrt{3}, -\sqrt{3})$.

Example 12.5. Find the point(s) on the hyperboloid $xy - yz - zx = 3$ closest to the origin.

Remark. It may be shown such closest point(s) exist.

For example, after a suitable change of coordinates, the surface is equivalent to the "two-piece" hyperboloid:

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

The distance between the origin and any point (x, y, z) on the hyperboloid above is simply:

$$\sqrt{x^2 + y^2 + z^2} = \sqrt{2x^2 + 2y^2 + 1} \geq 1,$$

which clearly has an absolute minimum.

However, the hyperboloid is not bounded $\Rightarrow$ farthest point does not exist.

Solution. Let $f(x, y, z) = \|(x, y, z) - (0, 0, 0)\|^2 = x^2 + y^2 + z^2$

Minimize f under constraint

$$g(x, y, z) = xy - yz - zx = 3$$

$$\nabla f = [2x \quad 2y \quad 2z] \quad \nabla g = [y - z \quad x - z \quad -x - y]$$

Note $\nabla g \neq [0, 0, 0]$ on $g = 3$

Use Lagrange Multipliers,

$$\left\{ \begin{array}{rcl} \nabla f &=& \lambda \nabla g \\ g &=& 3 \end{array} \right. \underset{\text{(Ex)}}{\Rightarrow} \left\{ \begin{array}{rcl} (x, y, z) &=& \pm(1, 1, -1) \\ \lambda &=& 1 \end{array} \right.$$

Note $f(1, 1, -1) = f(-1, -1, 1) = 3$

$\therefore$ Closest points are $\pm(1, 1, -1)$

Corresponding distance $= \sqrt{3}$

12.3 Lagrange Multipliers - Multiple Constraints

Theorem 12.6. *Lagrange Multipliers with multiple constraints*

Let $f, g_1, g_2, \cdots, g_k$ be C^1 functions on $\Omega \subseteq \mathbb{R}^n$

$$S = \{x \in \Omega : g_i(x) = c_i, i = 1, \cdots, k\}$$

Suppose

1. *a is a local extremum of f on S*

2. *$\nabla g_1(a), \cdots, \nabla g_k(a)$ are linearly independent*

Then

$$\begin{cases} \nabla f(a) &= \sum_{i=1}^{k} \lambda_i \nabla g_i(a) \quad \textit{for some } \lambda_1, \cdots, \lambda_k \in \mathbb{R} \\ g_i(a) &= c_i \quad\quad\quad\quad\quad\;\; \textit{for } i = 1, \cdots, k \end{cases}$$

Example 12.7. Maximize $f(x, y, z) = x^2+2y-z^2$ on the line $L = \begin{cases} 2x - y &= 0 \\ y + z &= 0 \end{cases}$ in $\mathbb{R}^3$

It is given that f has maximum on L

Solution. Let $g_1(x, y, z) = (2x - y)$ and $g_2(x, y, z) = y + z$

$$\nabla f = [2x \quad 2 \quad -2z]$$

$$\left.\begin{matrix} \nabla g_1 = [2 \quad -1 \quad 0] \\ \nabla g_2 = [0 \quad 1 \quad 1] \end{matrix}\right\} \text{ linearly independent}$$

Use Lagrange Multipliers,

$$\begin{cases} \nabla f &= \lambda_1 \nabla g_1 + \lambda_2 \nabla g_2 \\ g_1 &= 0 \\ g_2 &= 0 \end{cases}$$

$$\text{Hence } \begin{cases} 2x &= 2\lambda_1 + 0\lambda_2 & \cdots (1) \\ 2 &= -\lambda_1 + \lambda_2 & \cdots (2) \\ -2z &= 0\lambda_1 + \lambda_2 & \cdots (3) \\ 2x - y &= 0 & \cdots (4) \\ y + z &= 0 & \cdots (5) \end{cases}$$

$(4), (5) \Rightarrow 2x = y = -z$

$(1), (3) \Rightarrow \lambda_1 = x \quad \lambda_2 = -2z$

$(2) \Rightarrow -x - 2z = 2 \Rightarrow -x + 4x = 2 \Rightarrow x = \frac{2}{3}$

$\Rightarrow y = \frac{4}{3}, z = -\frac{4}{3}$ Since solution is unique and maximum exists, it must occur at $(x, y, z) = (\frac{2}{3}, \frac{4}{3}, -\frac{4}{3})$ with maximum value $f(\frac{2}{3}, \frac{4}{3}, -\frac{4}{3}) = \frac{4}{3}$

Example 12.8. Find the minimum distance (provided that it exists) between

$$C : xy = 1 \text{ and } L : x + 4y = \frac{15}{8}$$

IFRAME

Solution. Let $f(x, y, u, v) = \|(x, y) - (u, v)\|^2 = (x-u)^2 + (y-v)^2$

To find distance:

Minimize $f(x, y, u, v)$ under constraints

$$g_1(x, y, u, v) = xy = 1$$

$$g_2(x, y, u, v) = u + 4v = \frac{15}{8}$$

$$\begin{aligned}\nabla f &= [2(x-u) \quad 2(y-v) \quad -2(x-u) \quad -2(y-u)] \\ \nabla g_1 &= [y \quad x \quad 0 \quad 0] \\ \nabla g_2 &= [0 \quad 0 \quad 1 \quad 4]\end{aligned}$$

$\nabla g_1, \nabla g_2$ are linearly independent $\Leftrightarrow (x, y) \neq (0, 0)$

But $xy = 1 \Rightarrow \nabla g_1, \nabla g_2$ are linearly independent on $g_1 = 1$ and $g_2 = \frac{15}{8}$

Use Lagrange Multipliers,

$$\left\{\begin{array}{rcl} \nabla f &=& \lambda_1 \nabla g_1 + \lambda_2 \nabla g_2 \\ g_1 &=& 1 \\ g_2 &=& \frac{15}{8} \end{array}\right. \Rightarrow \left\{\begin{array}{rcll} 2(x-u) &=& \lambda_1 y & \cdots \ (1) \\ 2(y-v) &=& \lambda_1 x & \cdots \ (2) \\ -2(x-u) &=& \lambda_2 & \cdots \ (3) \\ -2(y-v) &=& 4\lambda_2 & \cdots \ (4) \\ xy &=& 1 & \cdots \ (5) \\ u+4v &=& \frac{15}{8} & \cdots \ (6) \end{array}\right.$$

Case 1:

If $\lambda_1 = 0$ or $\lambda_2 = 0$, then

$$x = u, y = v$$

$(6) \Rightarrow x = \frac{15}{8} - 4y$

$(5) \Rightarrow (\frac{15}{8} - 4y)y = 1 \Rightarrow -4y^2 + \frac{15}{8}y - 1 = 0$

No real solution

Case 2:

If $\lambda_1, \lambda_2 \neq 0$, then

$$\frac{1}{4} = \frac{x-u}{y-v} = \frac{y}{x} \Rightarrow x = 4y$$

$(5) \Rightarrow 4y^2 = 1 \Rightarrow y = \pm\frac{1}{2}$

$\therefore (x, y) = (2, \frac{1}{2})$ or $(-2, -\frac{1}{2})$

If $(x,y) = (2, \frac{1}{2})$,

$$\frac{2-u}{\frac{1}{2}-v} = \frac{1}{4} \Rightarrow 8 - 4u = \frac{1}{2} - v$$

$$\Rightarrow \begin{cases} -4u + v &= -\frac{15}{2} \\ u + 4v &= \frac{15}{8} \end{cases}$$

$$\Rightarrow (u, v) = (\frac{15}{8}, 0)$$

If $(x,y) = (-2, \frac{1}{2})$,

By similar calculation, $(u,v) = (-\frac{225}{136}, \frac{15}{17})$ Comparing the two solutions,

f attains minimum at $(x,y,u,v) = (2, \frac{1}{2}, \frac{15}{8}, 0)$

Distance between C and $L = \sqrt{f(2, \frac{1}{2}, \frac{15}{8}, 0)} = \frac{\sqrt{17}}{8}$

12.3.1 Where the Lagrange Multipliers Method "Fails"

Example 12.9. Provided that it exists, find the minimum of $f(x,y) = x$ on:

$$g(x,y) = x^3 - y^2 = 0.$$

IFRAME

It is easy to see that The absolute minimum of f occurs at $(x,y) = (0,0)$.

But:

$$\nabla f = [1 \quad 0],$$

while:

$$\nabla g = [3x^2 \quad -2y]$$

Hence,

$$\nabla f = \lambda \nabla g$$

has no solutions.

So, a naive (i.e. fail to check all conditions) application of Lagrange Multipliers would "miss" the point $(0,0)$ where the minimum occurs.

In general, when ∇g is not necessarily nonzero, one has to separately consider the points where ∇g is zero, after solving for the Lagrange multipliers.

Example 12.10. Provided that it exists, find the maximum of $f(x,y,z) = -y$ on:

$$g_1(x,y,z) = x^2 - y^2 - y^3 - z = 0$$
$$g_2(x,y,z) = y^2 + z = 0$$

The function f in fact attains its absolute maximum at $(0,0,0)$ (why?).

IFRAME

But:

$$\nabla f = [0 \quad -1 \quad 0],$$

while:

$$\nabla g_1 = [2x \quad -2y - 3y^2 \quad -1]$$
$$\nabla g_2 = [0 \quad 2y \quad 1]$$

Hence, there are no λ_1, λ_2 such that:

$$\nabla f(0,0,0) = \lambda_1 \nabla g_1(0,0,0) + \lambda_2 \nabla g_2(0,0,0),$$

which is a direct consequence of the linear dependence of the vectors:

$$\nabla g_1(0,0,0) = [0 \quad 0 \quad -1]$$

$$\nabla g_2(0,0,0) = [0 \quad 0 \quad 1].$$

(Their span is not "large enough" to accomodate $\nabla f(0,0,0)$.)

12.4 Implicit Function Theorem

Question When can we ”solve” a constraint?

For example, if $g(x,y) = c$, can we find $y = h(x)$ such that $g(x, h(x)) = c$?

Example 12.11. Consider level set $g(x,y) = x^2 - y^2 = 0$

IFRAME

Near $(0,0), y = x$? $y = -x$? or $\pm|x|$?

y is not uniquely determined by x

Example 12.12. $S : x^2 + y^2 + z^2 = 2$ in $\mathbb{R}^3$

IFRAME

Question: 3 variables, 1 equation $\Rightarrow$ S is 2-dimensional surface?

Solve for $z = h(x,y)$?

$x = k(y,z)$?

We focus locally near $(0,1,1)$

If we can solve for z as a differentiable function $z=z(x,y)$ near $(0,1,1)$,by implicit differentiation on $x^2+y^2+z^2=2$

$$\frac{\partial}{\partial x}: 2x+2z\frac{\partial z}{\partial x}=0$$

$$\frac{\partial}{\partial y}: 2y+2z\frac{\partial z}{\partial y}=0$$

At $(x,y,z)=(0,1,1) \Rightarrow \left\{\begin{array}{rcl} 0+2\frac{\partial z}{\partial x} & = & 0 \\ 2+2\frac{\partial z}{\partial y} & = & 0 \end{array}\right.$

$\Rightarrow [\frac{\partial z}{\partial x}, \frac{\partial z}{\partial y}] = [0 \quad -1]$ at $(0,1,1)$

How about x as a differentiable function $x=x(y,z)$ near $(0,1,1)$?

If so, by implicit differentiation,

$$\frac{\partial}{\partial y}: 2x\frac{\partial x}{\partial y}+2y=0$$

$$\frac{\partial}{\partial z}: 2x\frac{\partial x}{\partial y}+2y=0$$

Put $(x,y,z)=(0,1,1) \Rightarrow \left\{\begin{array}{rcl} 0+2 & = & 0 \\ 0+2 & = & 0 \end{array}\right.$ (coefficient of $\frac{\partial x}{\partial y}$ is $\frac{\partial g}{\partial x}=0$)

Contradiction!

$\therefore x$ is not a differentiable function of y,z near $(0,1,1)$

Reason:

For $x^2+y^2+z^2=2$,

If $y,z>1$ a little bit, no solution for x.

If $y,z<1$ a little bit, 2 solution for x.

Let $g(x,y,z)=x^2+y^2+z^2$.

Difference in the two cases:

At $(0,1,1)$,

$$\frac{\partial g}{\partial z}=2z\neq 0$$

$$\frac{\partial g}{\partial x}=2x=0$$

In general, given constraint $F(x,y,z)=c$

If $z = z(x,y)$, then by implicit differentiation,

$$\left.\begin{array}{lrcl} \frac{\partial}{\partial x}: & \frac{\partial F}{\partial x} + \frac{\partial F}{\partial z}\frac{\partial z}{\partial x} & = & 0 \\ \frac{\partial}{\partial y}: & \frac{\partial F}{\partial y} + \frac{\partial F}{\partial z}\frac{\partial z}{\partial y} & = & 0 \end{array}\right\} \circledast$$

If $F(\vec{a}) = c, \frac{\partial F}{\partial z}(\vec{a}) \neq 0$, then $\circledast$ has solution (No contradiction)

$\therefore z = z(x,y)$ may exist and

$$\begin{bmatrix} \frac{\partial z}{\partial x} & \frac{\partial z}{\partial y} \end{bmatrix} = -\frac{1}{\frac{\partial F}{\partial z}(\vec{a})} \begin{bmatrix} \frac{\partial F}{\partial x}(\vec{a}) & \frac{\partial F}{\partial y}(\vec{a}) \end{bmatrix}$$

Example 12.13. Multiple Constraints

$$C \left\{ \begin{array}{rcll} x^2 + y^2 + z^2 & = & 2 & \text{3 variables} \\ x + z & = & 1 & \text{2 equations} \end{array} \right.$$

IFRAME

Question: Is C a 1-dimensional curve? $y = y(x)$? $z = z(x)$?

If we can solve for y, z as differentiable functions $y(x), z(x)$

$$\text{Implicit Differentiation} \Rightarrow \left\{ \begin{array}{rcl} 2x + 2y\frac{dy}{dx} + 2z\frac{dz}{dx} & = & 0 \\ 1 + \frac{dz}{dx} & = & 0 \end{array} \right.$$

$$\begin{bmatrix} 2y & 2z \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \frac{dy}{dx} \\ \frac{dz}{dx} \end{bmatrix} = \begin{bmatrix} -2x \\ -1 \end{bmatrix}$$

If this linear system has a solution, then $y = y(x), z = z(x)$ may exist.

For instance, if $(x,y,z) = (0,1,1)$,

$$\begin{bmatrix} 2 & 2 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \frac{dy}{dx} \\ \frac{dz}{dx} \end{bmatrix} = \begin{bmatrix} 0 \\ -1 \end{bmatrix}$$

$$\Rightarrow \begin{bmatrix} \frac{dy}{dx} \\ \frac{dz}{dx} \end{bmatrix} = \begin{bmatrix} 2 & 2 \\ 0 & 1 \end{bmatrix}^{-1} \begin{bmatrix} 0 \\ -1 \end{bmatrix} = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$$

In general, given $F_1(x,y,z) = c_1$ and $F_2(x,y,z) = c_2$

Suppose $F_i(a,b,c) = c_i, i = 1,2$.

Do there exist differentiable functions $y = y(x), z = z(x)$ near (a,b,c) such that

$$\left\{ \begin{array}{rcl} F_1(x, y(x), z(x)) & = & c_1 \\ F_2(x, y(x), z(x)) & = & c_2 \end{array} \right. ?$$

If so, by implicit differentiation,

$$\begin{cases} \frac{\partial F_1}{\partial x} + \frac{\partial F_1}{\partial y}\frac{dy}{dx} + \frac{\partial F_1}{\partial z}\frac{dz}{dx} &= 0 \\ \frac{\partial F_2}{\partial x} + \frac{\partial F_2}{\partial y}\frac{dy}{dx} + \frac{\partial F_2}{\partial z}\frac{dz}{dx} &= 0 \end{cases}$$

$$\begin{bmatrix} \frac{\partial F_1}{\partial y} & \frac{\partial F_1}{\partial z} \\ \frac{\partial F_2}{\partial y} & \frac{\partial F_2}{\partial z} \end{bmatrix} \begin{bmatrix} \frac{dy}{dx} \\ \frac{dz}{dx} \end{bmatrix} = \begin{bmatrix} -\frac{\partial F_1}{\partial x} \\ -\frac{\partial F_2}{\partial x} \end{bmatrix}$$

If $\begin{bmatrix} \frac{\partial F_1}{\partial y}(\vec{a}) & \frac{\partial F_1}{\partial z}(\vec{a}) \\ \frac{\partial F_2}{\partial y}(\vec{a}) & \frac{\partial F_2}{\partial z}(\vec{a}) \end{bmatrix}^{-1}$ exists at $\vec{a} = (a, b, c)$,

then $\begin{bmatrix} \frac{dy}{dx}(a) \\ \frac{dz}{dx}(a) \end{bmatrix} = \begin{bmatrix} \frac{\partial F_1}{\partial y}(\vec{a}) & \frac{\partial F_1}{\partial z}(\vec{a}) \\ \frac{\partial F_2}{\partial y}(\vec{a}) & \frac{\partial F_2}{\partial z}(\vec{a}) \end{bmatrix}^{-1} \begin{bmatrix} -\frac{\partial F_1}{\partial x}(\vec{a}) \\ -\frac{\partial F_2}{\partial x}(\vec{a}) \end{bmatrix}$

Generally,
given $n + k$ variables
k equations

$$\begin{cases} F_1(x_1, \cdots, x_n, y_1, \cdots, y_k) &= c_1 \\ &\vdots \\ F_k(x_1, \cdots, x_n, y_1, \cdots, y_k) &= c_k \end{cases}$$

When can $y_1, \cdots, y_k$ be expressed as functions of $x_1, \cdots, x_n$ locally?

MATH 2010 Chapter 13

13.1 Implicit Function Theorem

Theorem 13.1 (Implicit Function Theorem)**.** *Let $\Omega \subseteq \mathbb{R}^{n+k}$ be open, $F : \Omega \to \mathbb{R}^k$ be C^1*

Denote $x = (x_1, \cdots, x_n), y = (y_1, \cdots, y_k)$

$$\begin{aligned} F(\vec{x}, \vec{y}) &= \begin{bmatrix} F_1(\vec{x}, \vec{y}) \\ \vdots \\ F_k(\vec{x}, \vec{y}) \end{bmatrix} \\ &= \begin{bmatrix} F_1(x_1, \cdots, x_n, y_1, \cdots, y_k) \\ \vdots \\ F_k(x_1, \cdots, x_n, y_1, \cdots, y_k) \end{bmatrix} \end{aligned}$$

Suppose $(a, b) \in \Omega$, where $a = \vec{a} \in \mathbb{R}^n$, $b = \vec{b} \in \mathbb{R}^k$, such that

$$F(a, b) = \vec{c} = \begin{bmatrix} c_1 \\ \vdots \\ c_k \end{bmatrix} \in \mathbb{R}^k$$

and the $k \times k$ matrix

$$(D_{\vec{y}}F)\,|_{(a,b)} := \left[\frac{\partial F_i}{\partial y_j}(a, b)\right]_{1 \leqslant i,j \leqslant k} = \begin{bmatrix} \frac{\partial F_1}{\partial y_1}(a, b) & \cdots & \frac{\partial F_1}{\partial y_k}(a, b) \\ \vdots & & \vdots \\ \frac{\partial F_k}{\partial y_1}(a, b) & \cdots & \frac{\partial F_k}{\partial y_k}(a, b) \end{bmatrix} \text{ is invertible.}$$

Then, there exist open sets $U \subseteq \mathbb{R}^n$ containing a , $V \subseteq \mathbb{R}^k$ containing b and a C^1 function

$$\begin{gathered} \varphi : U \to V \\ x = (x_1, \cdots, x_n) \in U \to y = (y_1, \cdots, y_k) \in V \end{gathered}$$

such that

1. $\varphi(a) = b$

2. $F(x, \varphi(x)) = c$

3. For any $(x, y) \in U \times V$ *such that* $F(x, y) = c$*, we have* $y = \varphi(x)$*.*

4. For $1 \le i \le k$*,* $1 \le j \le n$*,*

$$\left[\frac{\partial y_i}{\partial x_j}(a)\right] = \left[\frac{\partial \varphi_i}{\partial x_j}(a)\right] = -\left((D_{\vec{y}}F)^{-1}\, D_{\vec{x}}F\right)_{ij}$$

Remark. Provided that we know $\vec{y} = \vec{y}(\vec{x})$ is a differentiable function of $\vec{x}$, the last item follows from the Chain Rule.

If $\vec{y}$ is a differentiable function of $\vec{x}$, then $(\vec{x}, \vec{y})$ may be viewed as a vector-valued function:

$$\vec{\phi} : \mathbb{R}^n \longrightarrow \mathbb{R}^{n+k}$$

where:

$$\vec{\phi}(\vec{x}) = \begin{bmatrix} \vec{x} \\ \vec{y}(\vec{x}) \end{bmatrix}$$

Applying the chain rule for differentiation with respect to $\vec{x}$ to both sides of:

$$\underbrace{F(\vec{x}, \vec{y})}_{F(\vec{\phi}(\vec{x}))} = \vec{c},$$

we have:

$$\underbrace{\left[\begin{array}{ccc|ccc} \frac{\partial F_1}{\partial x_1} & \cdots & \frac{\partial F_1}{\partial x_n} & \frac{\partial F_1}{\partial y_1} & \cdots & \frac{\partial F_1}{\partial y_k} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \frac{\partial F_1}{\partial x_1} & \cdots & \frac{\partial F_k}{\partial x_n} & \frac{\partial F_k}{\partial y_1} & \cdots & \frac{\partial F_k}{\partial y_k} \end{array}\right]}_{DF = \left[D_{\vec{x}}F \mid D_{\vec{y}}F\right]} \underbrace{\left[\begin{array}{ccc} \frac{\partial \phi_1}{\partial x_1} & \cdots & \frac{\partial \phi_1}{\partial x_n} \\ \vdots & \vdots & \vdots \\ \frac{\partial \phi_n}{\partial x_1} & \cdots & \frac{\partial \phi_n}{\partial x_n} \\ \hline \frac{\partial \phi_{n+1}}{\partial x_1} & \cdots & \frac{\partial \phi_{n+1}}{\partial x_n} \\ \vdots & \vdots & \vdots \\ \frac{\partial \phi_{n+k}}{\partial x_1} & \cdots & \frac{\partial \phi_{n+k}}{\partial x_n} \end{array}\right]}_{D\vec{\phi}} = \vec{0}$$

$$\begin{bmatrix} \frac{\partial F_1}{\partial x_1} & \cdots & \frac{\partial F_1}{\partial x_n} \\ \vdots & \vdots & \vdots \\ \frac{\partial F_1}{\partial x_1} & \cdots & \frac{\partial F_k}{\partial x_n} \end{bmatrix} I_n + \begin{bmatrix} \frac{\partial F_1}{\partial y_1} & \cdots & \frac{\partial F_1}{\partial y_k} \\ \vdots & \vdots & \vdots \\ \frac{\partial F_1}{\partial y_1} & \cdots & \frac{\partial F_k}{\partial y_k} \end{bmatrix} \begin{bmatrix} \frac{\partial y_1}{\partial x_1} & \cdots & \frac{\partial y_1}{\partial x_n} \\ \vdots & \vdots & \vdots \\ \frac{\partial y_k}{\partial x_1} & \cdots & \frac{\partial y_k}{\partial x_n} \end{bmatrix} = \vec{0}$$

Hence,

$$\begin{bmatrix} \frac{\partial y_1}{\partial x_1} & \cdots & \frac{\partial y_1}{\partial x_n} \\ \vdots & \vdots & \vdots \\ \frac{\partial y_k}{\partial x_1} & \cdots & \frac{\partial y_k}{\partial x_n} \end{bmatrix} = - \begin{bmatrix} \frac{\partial F_1}{\partial y_1} & \cdots & \frac{\partial F_1}{\partial y_k} \\ \vdots & \vdots & \vdots \\ \frac{\partial F_1}{\partial y_1} & \cdots & \frac{\partial F_k}{\partial y_k} \end{bmatrix}^{-1} \begin{bmatrix} \frac{\partial F_1}{\partial x_1} & \cdots & \frac{\partial F_1}{\partial x_n} \\ \vdots & \vdots & \vdots \\ \frac{\partial F_1}{\partial x_1} & \cdots & \frac{\partial F_k}{\partial x_n} \end{bmatrix}$$

Remark. Applying the theorem to the special case where we have a real-valued differentiable function $F : \mathbb{R}^2 \longrightarrow \mathbb{R}$ satisfying:

$$F(x, y) = c,$$

we have:

$$\frac{\partial y}{\partial x} = -\frac{F_x}{F_y}$$

Example 13.2. $n = 2, k = 1$.

Consider the surface (a sphere to be precise) desribed by the following equation:

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

Is z implicitly a function of (x, y) near the point $(0, 1, 1)$ on the surface?

If so, what is $\dfrac{\partial z}{\partial x}, \dfrac{\partial z}{\partial y}$ at $(x, y, z) = (0, 1, 1)$?

Solution. The equation describing the surface is equivalent to $F(x, y, z) = 2$, where:

$$F(x, y, z) = x^2 + y^2 + z^2 : \mathbb{R}^3 \longrightarrow \mathbb{R}^1$$

Observe that F is C^1 on $\mathbb{R}^3$. In the context of the IFT, we have $k = 1, n = 3-1 = 2$.

So, it is possible that any one ($k = 1$) of (x, y, z) is implicitly a function of the other two ($n = 2$).

We have:

$$D_{(z)}F = [F_z] = [2z].$$

At $(0, 1, 1)$, we have $[F_z]|_{(0,1,1)} = [2 \cdot 1] = [2]$, which is an invertible 1×1 matrix, with inverse:

$$\left(D_{(z)}F|_{(0,1,1)}\right)^{-1} = [1/2]$$

Hence, the conditions of IFT are satisfied. The variable z is implicitly a C^1 function of (x, y) near $(0, 1, 1)$ on the surface.

Moreover, we have:

$$\begin{aligned}
\begin{bmatrix} \frac{\partial z}{\partial x} & \frac{\partial z}{\partial y} \end{bmatrix}\Big|_{(0,1,1)} &= -\left(D_{(z)}F|_{(0,1,1)}\right)^{-1} \begin{bmatrix} F_x & F_y \end{bmatrix} |_{(0,1,1)} \\
&= -[1/2] \begin{bmatrix} F_x & F_y \end{bmatrix} |_{(0,1,1)} \\
&= -[1/2] \begin{bmatrix} 2x & 2y \end{bmatrix} |_{(0,1,1)} \\
&= -[1/2] \begin{bmatrix} 0 & 2 \end{bmatrix}
\end{aligned}$$

Hence:
$$\frac{\partial z}{\partial x}|_{(0,1,1)} = -\frac{1}{2} \cdot 0 = 0, \quad \frac{\partial z}{\partial y}|_{(0,1,1)} = -\frac{1}{2} \cdot 2 = -1$$

Example 13.3. $n = 1, k = 2$ Consider the curve (a circle) which is the intersection the following two surfaces:

$$\begin{cases} x^2 + y^2 + z^2 & = 2 \\ x + z & = 1 \end{cases}$$

IFRAME
What does IFT say in this case?
Define $F : \mathbb{R}^{2+1} \longrightarrow \mathbb{R}^2$ as follows:

$$F(x, y, z) = \begin{bmatrix} F_1 \\ F_2 \end{bmatrix} = \begin{bmatrix} x^2 + y^2 + z^2 \\ x + z \end{bmatrix}$$

Then, the curve in question corresponds to the constraint:

$$F(x, y, z) = \vec{c} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$$

Since the codomain of F is $\mathbb{R}^2$ ($k = 2$), and F has 3 variables, by IFT, any two ($k = 2$) of the variables of F could implicitly be a function of the remaining one variable ($n = 3 - k = 1$).

We have:
$$D_{(y,z)}F = \begin{bmatrix} F_{1,y} & F_{1,z} \\ F_{2,y} & F_{2,z} \end{bmatrix} = \begin{bmatrix} 2y & 2z \\ 0 & 1 \end{bmatrix}$$

Hence, at, for example, the point $(0, 1, 1)$, we have:

$$D_{(y,z)}F\big|_{(0,1,1)} = \begin{bmatrix} 2 & 2 \\ 0 & 1 \end{bmatrix},$$

which is invertible.

Hence, by IFT, on the curve the variables y, z are implicitly differentiable functions of x near the point $(0, 1, 1)$, with:

$$\begin{bmatrix} \frac{\partial y}{\partial x} \\ \frac{\partial z}{\partial x} \end{bmatrix} = - \; D_{(y,z)}F\big|_{(0,1,1)}^{-1} \begin{bmatrix} F_{1,x} \\ F_{2,x} \end{bmatrix}\Bigg|_{(0,1,1)} = -\frac{1}{2} \begin{bmatrix} 1 & -2 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} 2 \cdot 0 \\ 1 \end{bmatrix}$$
$$= \begin{bmatrix} 1 \\ -1 \end{bmatrix}$$

Remark. There is no intrinsic ordering to the variables x, y, z, so we might equally ask whether x, y are implicitly functions of z on the curve.

Since

$$D_{(x,y)}F\Big|_{(0,1,1)} = \begin{bmatrix} F_{1,x} & F_{1,y} \\ F_{2,x} & F_{2,y} \end{bmatrix}\Bigg|_{(0,1,1)} = \begin{bmatrix} 2x & 2y \\ 1 & 0 \end{bmatrix}\Bigg|_{(0,1,1)} = \begin{bmatrix} 0 & 2 \\ 1 & 0 \end{bmatrix}$$

is invertible, the IFT says that on the curve the variable x, y are implicity functions of z near the point $(0, 1, 1)$, with:

$$\begin{aligned} \begin{bmatrix} \frac{\partial x}{\partial z} \\ \frac{\partial y}{\partial z} \end{bmatrix} &= -\, D_{(x,y)}F\Big|_{(0,1,1)}^{-1} \begin{bmatrix} F_{1,z} \\ F_{2,z} \end{bmatrix}\Bigg|_{(0,1,1)} = \frac{1}{2}\begin{bmatrix} 0 & -2 \\ -1 & 0 \end{bmatrix}\begin{bmatrix} 2\cdot 1 \\ 1 \end{bmatrix} \\ &= \begin{bmatrix} -1 \\ -1 \end{bmatrix} \end{aligned}$$

Example 13.4. Consider the constraints

$$\begin{cases} xz + \sin(yz - x^2) & = & 8 \\ x + 4y + 3z & = & 18 \end{cases}$$

Near $(2, 1, 4)$, can we express 2 of the variables as functions of the remaining variable?

Solution. Let $F(x, y, z) = \begin{bmatrix} F_1(x,y,z) \\ F_2(x,y,z) \end{bmatrix} = \begin{bmatrix} xz + \sin(yz - x^2) \\ x + 4y + 3z \end{bmatrix}$.

Then:

$$DF = \begin{bmatrix} z - 2x\cos(yz - x^2) & z\cos(yz - x^2) & x + y\cos(yz - x^2) \\ 1 & 4 & 3 \end{bmatrix}$$

$$DF(2,1,4) = \begin{bmatrix} 0 & 4 & 3 \\ 1 & 4 & 3 \end{bmatrix}$$

$D_{(x,y)}F|_{(2,1,4)}$:

$$\begin{vmatrix} 0 & 4 \\ 1 & 4 \end{vmatrix} = -4 \neq 0 \overset{IFT}{\Rightarrow} x, y \text{ can be expressed as functions of } z \text{ near } (2,1,4)$$

$D_{(x,z)}F|_{(2,1,4)}$:

$$\begin{vmatrix} 0 & 3 \\ 1 & 3 \end{vmatrix} = -3 \neq 0 \overset{IFT}{\Rightarrow} x, z \text{ can be expressed as functions of } y \text{ near } (2,1,4)$$

$D_{(y,z)}F|_{(2,1,4)}$:

$$\begin{vmatrix} 4 & 3 \\ 4 & 3 \end{vmatrix} = 0$$

Hence, $D_{(y,z)}F|_{(2,1,4)}$ is not invertible. We may not conclude from the IFT whether y, z are locally functions of x near $(2, 1, 4)$.

Remark. The variables y, z are in fact not differentiable functions of x near $(2, 1, 4)$.

Otherwise, by implicit differentiation, we have:

$$\begin{aligned} D_{(y,z)}F|_{(2,1,4)} \left.\begin{bmatrix} \frac{dy}{dx} \\ \frac{dz}{dx} \end{bmatrix}\right|_{(2,1,4)} &= -D_xF|_{(2,1,4)} = -\begin{bmatrix} 0 \\ 1 \end{bmatrix} \\ \begin{bmatrix} 4 & 3 \\ 4 & 3 \end{bmatrix} \left.\begin{bmatrix} \frac{dy}{dx} \\ \frac{dz}{dx} \end{bmatrix}\right|_{(2,1,4)} &= -\begin{bmatrix} 0 \\ 1 \end{bmatrix}, \end{aligned}$$

which has no solutions, i.e. not satisfied for any values of $\frac{\partial y}{\partial x}|_{(2,1,4)}, \frac{\partial z}{\partial x}|_{(2,1,4)}$.

Hence, y, z cannot be differentiable functions of x near $(2, 1, 4)$.

Remark. Implicit Function Theorem has many important applications, such as rigorous proofs of

1. Implicit differentiation
2. Tangent plane of surface $F(x, y, z) = c$
3. Lagrange Multipliers

13.2 Inverse Function Theorem

Theorem 13.5 (Inverse Function Theorem)**.** *Let $\Omega \subseteq \mathbb{R}^n$ be open.*

$f : \Omega \to \mathbb{R}^n$ be $C^1, f(a) = b$.

Suppose $Df(a)$ is an invertible $n \times n$ matrix

Then, there exist open sets $U_1, U_2 \subseteq \mathbb{R}^n, a \in U_1, b \in U_2$ and a C^1 function $g : U_2 \to U_1$ such that:

1. *$g(b) = a$*
2. *$g(f(x)) = x$ for all $x \in U_1$*
   $f(g(y)) = y$ for all $y \in U_2$
   (g is a local inverse of f : $g = (f|_{U_1})^{-1}$)

3. $Dg(b) = Df(a)^{-1}$

Remark. The Inverse Function Theorem is equivalent to Implicit Function Theorem.

Idea:

$\Rightarrow$: Given $F(x, y) = c$, where $x \in \mathbb{R}^n$ and $y, c \in \mathbb{R}^k$, apply the Inverse Function Theorem to: $H(x, y) = (x, F(x, y)) : \mathbb{R}^{n+k} \longrightarrow \mathbb{R}^{n+k}$.

$\Leftarrow$: Given $H : \mathbb{R}^n \longrightarrow \mathbb{R}^n$, apply the Implicit Function Theorem to $F(x, y) = 0$, where $F(x, y) = y - H(x) : \mathbb{R}^{2n} \longrightarrow \mathbb{R}^n$.

Example 13.6. $f : \mathbb{R}^2 \to \mathbb{R}^2$

$f(x, y) = (x^2 - y^2, 2xy)$

Clearly $f(-x, -y) = f(x, y)$

$\Rightarrow f$ is not injective and has no global inverse.

How about local inverse?

Solution.

$$f(x, y) = \begin{bmatrix} x^2 - y^2 \\ 2xy \end{bmatrix}$$

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

$$\det Df(x, y) = 4(x^2 + y^2) > 0 \Leftrightarrow (x, y) \neq (0, 0)$$

By the Inverse Function Theorem, f is locally invertible with differentiable local inverse.

For instance, let $g(u, v)$ be a local inverse of $f(x, y)$ near $(x, y) = (1, -1)$.

Then $f(1, -1) = (0, -2) \Rightarrow g(0, -2) = (1, -1)$, and:

$$Dg(0, -2) = Df(1, -1)^{-1} = \begin{bmatrix} 2 & 2 \\ -2 & 2 \end{bmatrix}^{-1} = \frac{1}{4} \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}$$

In fact, we can find $g(u, v)$ explicitly.

Let:

$$\begin{cases} u &= x^2 - y^2 \\ v &= 2xy \end{cases}$$

Near $(x, y) = (1, -1), x \neq 0 \Rightarrow y = \frac{v}{2x}$

$\therefore u = x^2 - (\frac{v}{2x})^2$

$$4x^4 - 4ux^2 - v^2 = 0$$

$$\begin{aligned} x^2 &= \frac{4u \pm \sqrt{(-4u)^2 - 4(4)(-v^2)}}{8} \\ &= \frac{u \pm \sqrt{u^2 + v^2}}{2} \end{aligned}$$

Let $(x, y) = (1, -1)$, then $(u, v) = (0, -2)$.

$$\Rightarrow 1^2 = \frac{0 \pm \sqrt{4}}{2}$$

So, we may reject the negative sign, and it follows that:

$$x^2 = \frac{u + \sqrt{u^2 + v^2}}{2}$$

$$\Rightarrow x = \sqrt{\frac{u + \sqrt{u^2 + v^2}}{2}}, y = \frac{2v}{x} = \frac{2\sqrt{2}v}{\sqrt{u + \sqrt{u^2 + v^2}}}$$

Hence,

$$g(u, v) = \left(\sqrt{\frac{u + \sqrt{u^2 + v^2}}{2}}, \frac{2\sqrt{2}v}{\sqrt{u + \sqrt{u^2 + v^2}}} \right)$$

In both Implicit and Inverse Function Theorems,
we assume a Jacobi matrix to be invertible.

Without this assumption , the theorems are **inconclusive** on the existence of local implicit or inverse function. See the examples below:

Implicit Function Theorem

- $F(x, y) = x^2 - y^2 = 0$
  $\frac{\partial F}{\partial y}|_{(0,0)} = 0$
- $F(x, y) = x^3 - y^3 = 0$
  $\frac{\partial F}{\partial y}|_{(0,0)} = 0$
  $y = x$ locally (and in fact globally).

Inverse Function Theorem

- $f(x) = x^2$

  $f'(0) = 0$

  Not injective near $x = 0$.

  Hence, no local inverse near $x = 0$.

- $f(x) = x^3$

  $f'(0) = 0$

  The function f has a global inverse: $g(y) = \sqrt[3]{y}$